An Additional “Depth” of Reverberation Helps Content Stand Out: Media Content Emphasis Using Audio Reverberation Effect

Ren Gang, Samarth H. Shivaswamy, Stephen Roessner, Mark F. Bocko, Dave Headlam

University of Rochester, Rochester, NY, USA

***Abstract*—The audio reverberation effect adds in a perceptual depth dimension to media content and can be intelligently applied to emphasis the target media segments. A quantitative model of this emphasis effects is derived from subjective evaluation experiments.**

I. INTRODUCTION

In this paper we propose a media content emphasis scheme that changes the attention pattern of the audience by applying audio reverberation effect. Specifically we apply artificial reverberation to part of a media sequence using signal processing methods and utilize the contrast of reverberation effects to underscore important media segments.

The proposed media content emphasis scheme utilizes human capability for perceiving an aural space. This perception capability is extremely sharp at the transition points of reverberation effects and thus the perceptual/cognitive response can be employed to emphasize target media segments. The proposed media emphasis model is based on subjective evaluation experiments to ensure its effectiveness.

II. SYSTEM ARCHITECTURE

The proposed system architecture is illustrated in Fig. 1. The audio reverberation effect is applied to audio content selectively based on the media emphasis model. Audio content analysis algorithm based on [1] detects the reverberation effect in the original audio signal. The media content with strong reverberation is identified and refrained from applying additional reverberation effect because too much accumulated reverberation degrades media intelligibility [2]. The media emphasis model then translates the media emphasis instructions in programming script to reverberation pattern assignment.

Optional dereverberation algorithms [2,3] (marked with an asterisk in Fig. 1) can be further applied to provide additional flexibility of audio effect controls. However, these algorithms produce audible artifacts that “tag” the media segments and compromise the accuracy of media emphasis model. Based on these considerations, effect of dereverberation is not included in media emphasis model.

III. COMPARISON WITH CONVENTIONAL METHODS

A. Comparison with Loudness-Based Methods

Conventionally television stations boost the audio loudness during advertisement session to attract audience attention. These loudness-based media content emphasis methods suffer from the following three drawbacks:

- Boosting volume has negative psychological effects.
- The audio speakers in consumer electronics devices have limited loudness range, which limits the boosting range.
- Automatic volumes control limits the loudness dynamics.

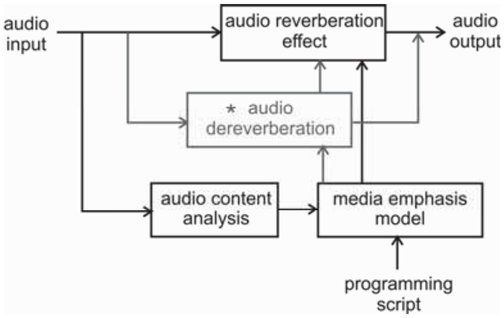


Fig. 1. System architecture. The media emphasis model decides the audio reverberation pattern based on the media emphasis instruction in the programming scripts and input audio characteristics.

Even with these drawbacks of loudness-based methods boosting audio volume was still a standard practice due to the importance of content emphasis to media programming.

The CALM (Commercial Advertisement Loudness Mitigation) Act [4] further necessitates a media content emphasis method not based on loudness. The CALM act was passed in 2010 and effective since 2011. By specifying the acceptable volume range during an advertisement session, CALM essentially eliminates the possibility of loudness-based methods. Our proposed media content emphasis method achieves similar functionalities while conforms to the CALM specifications and thus serves as an important substitution of pre-CALM methods.

B. Comparison with Content-Based Methods

Media contents naturally attract different levels of user attention. For example, a more interesting story line helps an advertisement segment stand out. Carefully designing of background music or audio effect can dramatically change audience attention patterns. Compared with these content based methods, the merit of our proposed method includes:

- Audio reverberation effect based method is orthogonal to media content and is straightforward to apply.
- A quantitative model is provided and thus we can easily form a price hierarchy. For other content-based methods, the emphasis extend is difficult to quantify.
- The persuasive effect is more subconscious: or perceive-but-not-articulate, which is amenable for media emphasis.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

IV. DESIGN MEDIA EMPHASIS PATTERNS

In this paper we limit the media emphasis patterns to (1) three consecutive speech-only advertisement segments and (2) three consecutive music segments. The detailed test and reference patterns for speech ("SI/SII") are illustrated in Fig. 2(a), (b). Type I test patterns ("SI/MI") have one reverberant test segment. Type II test patterns ("SI/MI") have two reverberant test segments. The reverberation time is 0.8 second for speech test segments ("S") and 1.2 second for music test segments ("M"). Test media segments in a same pattern are intentionally designed to be similar: the test speech segments are designed with similar content and length, and are announced by the same person; the music segments are of similar composition and performance style.

The auditory loudness of test media segments is equalized to exclude the emphasis effect of loudness. The reference patterns ("SI-R", "SII-R") have uniform reverberation configurations that concur with the majority reverberation settings of each test types. The test patterns for music are not plotted here but can be obtained by change all "S" segments in Fig. 2(a),(b) to "M" segments. We also add in two background segments ("B1" and "B2") in the beginning and ending of each test pattern to avoid abrupt program "cutting-in".

V. SUBJECTIVE EVALUATION-BASED MODELS

Test subjects (human listeners) rate the perceptual significance of each test segment using 1(least significant)-6 (most significant) scale. 5 replications of 16 patterns (test/reference) are assigned randomly to 20 test subjects. Every test subject rates all 4 types of test patterns as illustrated in Fig. 2(c),(d). Using this test assignment method we ensure that the test segments do not repeat for the same test subject. This feature is designed to mitigate the memory effect: If the same test segment is repeated to a test subject, she tends to give the same rating and thus couples the two experiments.

Suppose a test subject rated three consecutive segments as r_1, r_2, and r_3, we normalize these ratings as $3r_1/(r_1+r_2+r_3)$, $3r_2/(r_1+r_2+r_3)$, and $3r_3/(r_1+r_2+r_3)$. This method split 3 tokens of significance "currency" among three segments. The normalized rating in each segment is then compared to the segment at the same position in its reference pattern using two-sample t-test [5]. For example, the t-test of "S1; SI-1" is obtained by comparing the 5 normalized ratings of replicated "S1; SI-1" to 5 normalized ratings of "S1: SI-R".

From the test statistics we conclude (1) a change of rating patterns due to reverberation at significance level of 0.05 [5]; (2) media segments near the transition points of reverberation effect have larger rating changes. We also calculate the mean values (of 5 replications) of normalized ratings for every media segment in each pattern. The emphasis model in Fig. 3 is calculated as the difference of this mean value between a test pattern and its reference pattern, which is an unbiased estimator of the difference between two samples (5 test vs. reference pattern replications) [5]. For example, the emphasis effect of "S1; SI-1" is calculated as the difference between the mean normalized rating of "S1; SI-1" and "S1; SI-R".

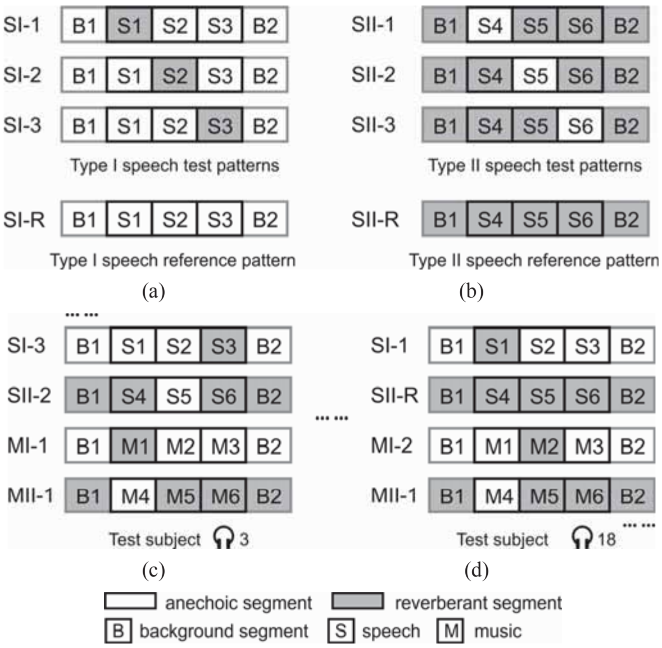


Fig. 2. Test media emphasis patterns: (a) and (b) are test patterns for speech. (c) and (d) are examples of test sessions assigned to two test subjects.

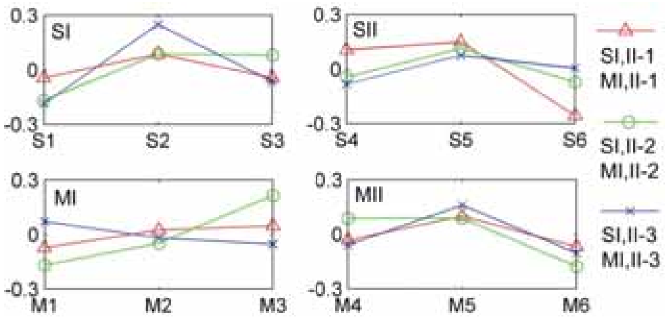


Fig. 3. Emphasis models derived from subjective evaluation experiments.

VI. SUMMARY

We apply reverberation patterns to add an additional "depth" dimension to audio content and achieve effective media emphasis functionalities. Quantitative media emphasis model is obtained from subjective evaluation experiments using statistical analysis. These effect patterns and quantitative models find important applications in media programming, advertisement pricing and interactive media distribution.

REFERENCES

[1] Ren Gang, Bocko, M.F., Headlam, D., "Reverberation features identification from music recordings using the discrete wavelet transform," in *Proceedings of 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP),* 14-19 March 2010, pp. 161-164.

[2] Naylor, P.A.; Gaubitch, N.D.(Eds.), *Speech Dereverberation,* Springer: New York, NY, 2010, pp. 24-28.

[3] Yasuraoka, N., Yoshioka, T.; Nakatani, T.; Nakamura, A.; Okuno, H.G.,"Music dereverberation using harmonic structure source model and Wiener filter," in *Proceedings of 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP),* 14-19 March 2010, pp. 53-56.

[4] CALM: http://www.fcc.gov/encyclopedia/loud-commercials/

[5] D.C. Montgomery, *Design and Analysis of Experiments,* 7th ed., John Wiley & Sons: Hoboken, NJ, 2009, pp. 34-43.

SoftOC: Real-time Projector-wall-camera Communication System

Chengcheng Pei, *Student Member, IEEE,* Zaichen Zhang, *Member, IEEE,* Shujian Zhang
ccpei87@gmail.com, zczhang@seu.edu.cn, sjz@seu.edu.cn

***Abstract*— Camera-equipped consumer electronics tend to become cheaper and cheaper currently. These cameras can provide high-speed communication links to combat the shortage of radio spectrum. This paper introduces SoftOC, a novel real-time projector-wall-camera communication system. We name our system software optical communication (SOC) system. In order to improve the reliability of these systems and to speed up the computing process, we propose some schemes. We verify the effectiveness of our proposed schemes on real-time communication prototype.**

I. INTRODUCTION

More and more consumer electronics are camera-equipped today. In [1, 2], cameras and liquid crystal displays (LCDs) are proposed to act as secure high-speed communication links. In these links, data are modulated to the spatial frequency domain of a series of coded images, which are emitted by LCDs and then captured by cameras. This kind of modulation is called spatial discrete multiple tone (SDMT).

However, line-of-sights are essential for these LCD-camera links. More and more consumer electronics will be projector-equipped in the near future. Projector-wall-camera links were proposed and studied as the extension of LCD-camera links to the non-line-of-light (NLOS) ones [3] in order to increase link robustness and ease of use.

We present a real-time projector-wall-camera communication system, SoftOC, for the first time. To improve their reliability, we propose nonlinear quantization, iterated positioning scheme, and over-sampling based synchronization. To speed up the computing process, we propose parallel architecture of convolutional coding and decoding. We verify the effectiveness of our proposed schemes on real-time communication system demo. In addition, our proposed two-channel transmission scheme has higher information rate than existing SDMT schemes.

The core part of our SoftOC system is a software package, which can run on general computers connected to projectors and cameras. We name this kind of systems software optical communication (SOC) systems. Software packages can be downloaded to consumer electronics with pico-projectors or cameras to provide communication services without aid of extra hardware, which can make our life more colorful.

II. PROJECTOR-WALL-CAMERA COMMUNICATION SYSTEMS

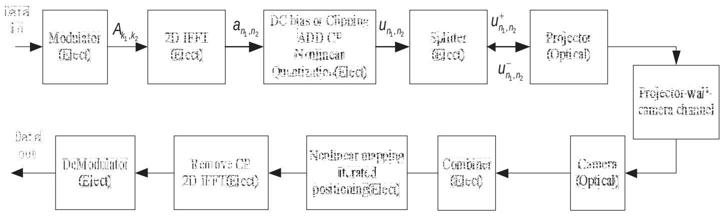


Fig.1 Projector-wall-camera Communication System

The general block diagram of our system is illustrated in Fig.1. At the transmitter, the input data, such as movies or songs, are separated into several symbols. Each symbol is mapped onto a $N_1 \times N_2$ complex matrix A using QPSK constellation. The modulator applies 2D inverse fast Fourier transform (IFFT) on A .Let a denote IFFT of A . A must have Hermitian symmetry to keep a real.

$$A = \begin{bmatrix} A_{0,0} & L & A_{0,N_2-1} \\ L & L & L \\ A_{N_1-1,0} & L & A_{N_1-1,N_2-1} \end{bmatrix} \tag{1}$$

where N_1 and N_2 must be even. The entries of A are given by

$$A_{k_1,k_2} = A^*_{N_1-k_1,N_2-k_2} \tag{2}$$

where k_1 and k_2 ($k_1 = 0,1, L , N_1 - 1$ and $k_2 = 0,1, L , N_2 - 1$)are the row and column indices respectively, $(\bullet)^*$ denotes the complex conjugate. The entries of a are given by

$$a_{n_1,n_2} = \frac{1}{\sqrt{N_1 N_2}} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} A_{k_1,k_2} e^{j(\frac{2\pi k_1 n_1}{N_1} + \frac{2\pi k_2 n_2}{N_2})} \tag{3}$$

where $n_1 = 0,1,..., N_1 - 1$ and $n_2 = 0,1,..., N_2 - 1$.

All elements of a must be discretized to modulate the optical intensity of respective projector pixels.

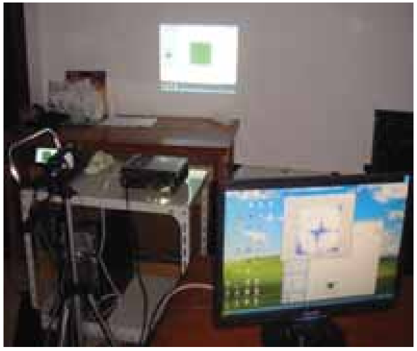


Fig.2 Real-time prototype

Let u denote discrete version of a . u is displayed by the projector. All elements of u are given by

This work is supported by NSFC project (60902010), the Research Fund of NCRL (No. 2012A03), National Science & Technology Major projects of China (2009ZX03006-008-02 and 2010ZX03006-009), and the Program for New Century Excellent Talents in University (NCET-09-0299).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$u_{n_1,n_2} = L(a_{n_1,n_2}) \tag{4}$$

where $L(\bullet)$ is the nonlinear quantization function given by.

$$L(x) = \begin{cases} \left\lfloor 255\left[0.5(1-e^{-gx})+0.5\right]\right\rfloor, x \geq 0 \\ \left\lfloor 255\left[-0.5(1-e^{gx})+0.5\right]\right\rfloor, x < 0 \end{cases} \tag{5}$$

where $\lfloor \bullet \rfloor$ is the floor function, and g should be properly selected to make $L(x)$ range from 0 to 255.

The positive and negative parts of u_{n_1,n_2} are extracted and transmitted in two independent separable channels, respectively. For example, the positive part of u_{n_1,n_2} , $u^+_{n_1,n_2}$ is transmitted in the red channel; the negative part of u_{n_1,n_2} , $u^-_{n_1,n_2}$ is transmitted in the blue channel. Both $u^+_{n_1,n_2}$ and $u^-_{n_1,n_2}$ are non-negative.

$$\begin{aligned} u^+_{n_1,n_2} &= u_{n_1,n_2}\varepsilon(u_{n_1,n_2}) \\ u^-_{n_1,n_2} &= -u_{n_1,n_2}\varepsilon(-u_{n_1,n_2}) \end{aligned} \tag{6}$$

where $\varepsilon(\bullet)$ is the step function.

At the receiving end, we get the estimation of transmitted signal by Equation (7).

$$z_{n_1,n_2} = z^+_{n_1,n_2}\varepsilon(z^+_{n_1,n_2} - z^-_{n_1,n_2}) - z^-_{n_1,n_2}\varepsilon(z^-_{n_1,n_2} - z^+_{n_1,n_2}) \tag{7}$$

where $z^+_{n_1,n_2}$ and $z^-_{n_1,n_2}$ denote the electronic signals received in two color channels through cameras, respectively.

A cyclic prefix (CP) should be added at the transmitter and removed at the receiver. The length of the CP depends on the positioning accuracy and inter-symbol interference (ISI).

III. OUR SCHEMES

A. Nonlinear quantization

Nonlinear quantization can make full use of quantization space. Our experiments on the real-time demo system prove that nonlinear quantization can effectively improve the system performance.

B. Iterated positioning scheme

Our proposed scheme is shown in Fig.3. Usually, about 5 iterations (including corner detection and perspective correction [2]) can improve the system performance greatly, especially when corner detection scheme does not work very well.

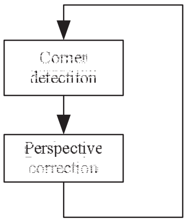


Fig.3 Iterated positioning scheme

C. Over-sampling based synchronization

Over-sampling based synchronization is to combine all the frames received by over-sampling to estimate certain frames.

D. Parallel architecture of convolutional coding and decoding

Parallel architecture of convolutional coding and decoding are introduced to speed up the signal processing.

E. Two-channel transmission

The positive and negative parts of u_{n_1,n_2} are extracted and transmitted in two independent channels, respectively.

IV. CONCLUSION

Physical layer configuration is given in Table.1, respectively. Our system can transmit real-time video with data rate of 3.84 Mbps.

Table.1 Physical layer configuration

N_1 (N_2)	81
Symbol number per frame	96
Frame number per second	50
Resolution ratio of LCD	1920*1080

We model Equation (7) with a nonlinear system. We also demonstrate that the two-channel transmission scheme has higher information rate than existing SDMT schemes.

Due to shortage of radio spectrum current communication data rate cannot meet people’s needs. Our proposed system can provide extra communication service with the aid of cheaper and cheaper cameras. Beside lots of advantages compared with radio communications, we have proved that our system has extremely potential capacity [4].

REFERENCE

[1] Hranilovic, S.; Kschischang, F.R.; , "A pixelated MIMO wireless optical communication system," *Selected Topics in Quantum Electronics, IEEE Journal of* , vol.12, no.4, pp.859-874, July-Aug. 2006

[2] S. D. Perli, N. Ahmed, and D. Katabi. Pixnet:designing interference-free wireless links using lcd-camera pairs. Pro-ceedings of ACM International Conference on Mobile Com-puting and Networking, 2010.

[3] Pei, C.C.; Zhang, Z.C.; Fang, W.X.; Zhang, S.J.; , "2D-DPSK for quasi-diffuse pixelated wireless optical channels," *Communication Technology (ICCT), 2011 IEEE 13th International Conference on* , vol., no., pp.556-559, 25-28 Sept. 2011

[1] Chengcheng Pei; Zaichen Zhang;, " On the information rate of different SDMT modulations," *Wireless Communications and Signal Processing (WCSP), 2012 International Conference on* , to appear.

Implementation of a Practical Query-by-Singing/Humming (QbSH) System and Its Commercial Applications

Chai-Jong Song, Hochong Park, Chang-Mo Yang, Sei-Jin Jang, and Seok-Phil Lee

Abstract—**This paper proposes a practical query-by-singing/humming (QbSH) system that retrieves polyphonic music such as an MP3 and its commercial applications. The performance of music retrieval system is mainly affected by the server. This paper discusses developing the state-of-the-art server side software stack which has several managers and plug-in engines. It also describes implementation of digital signal processor (DSP) module for stand-alone embedded platforms. The paper shows three different models for its commercial applications like smart phone, laptop and karaoke. We evaluate the performance of the proposed system with polyphonic music datasets using users' humming datasets as the input query.**

I. INTRODUCTION

The consumption of digital music has already exceeded that of analog music. The number of content services in the mobile device is also increasing rapidly. It causes an increase in the consumers' requests for convenient and efficient content retrieval services. From this perspective, the QbSH system using the user's singing/humming, not text inputs, attracts much attention. This is very useful in various devices with the limitation in the user interface.

The basic function of a QbSH system is searching for the most similar music to the query from the DB. There is clearly the main criterion for the QbSH system: how can it find the music precisely and rapidly? Until now, most of the QbSH systems have been using the monophonic music database (DB) like musical instrument digital interface (MIDI) files. While QbSH systems using monophonic DB have advantages in the search time and accuracy, there is a critical issue for users who do not use monophonic music at all. Using polyphonic music files in DB is required for the practical QbSH system which is able to be served in various commercial fields. Therefore, we need to simplify the polyphonic music signals for the rapid searching because they are too complicated to be compared directly.

The feature extraction procedure for the polyphonic music is necessary to establish a QbSH system. The melody sequence is known as the best feature among many things through preview researches [1][2]. The problem is the accuracy of the melody extraction because the errors occurred from this stage are propagated to the matching stage. Therefore, we have to pay attention to design the algorithms for the melody extraction and matching engine. Those are tightly coupled with the performance of the system accuracy. The matching engine also affects the latency of the system as much as indexing feature DB does.

II. DEVELOPMENT OF THE PROPOSED QBSH

We propose the practical QbSH system which consists of four main functional modules. The first one suppresses the background noise of user's query signal and then estimates the pitch sequences from that. The second one extracts the features from the polyphonic music using the harmonic structure which is a key characteristic of human vocal and musical instruments. The third one is indexing the feature DB that is built by the second one. The last one is the matching engine finding the music that has the most similar features to an input sequence from the first one. We utilize the advanced algorithms for the proposed QbSH system [3][4].

A. The State-of-the-art software stack for the server

A QbSH system is a kind of server-client model. The server is required to be maintained efficiently because it is in charge of the most work of the system like managing the network connections, searching for the music, and updating the system status. Keeping this in mind, we design and implement the state-of-the-art software stack for the server. Fig. 1 depicts its structure consisted of six managers and four DBs. It provides application programming interface (API) on the top level for programmers. It also has the plug-in interfaces for the engine module developers. This structure makes it possible to develop the modules independently. Every engine module is able to be easily replaced to alternative one on this structure.

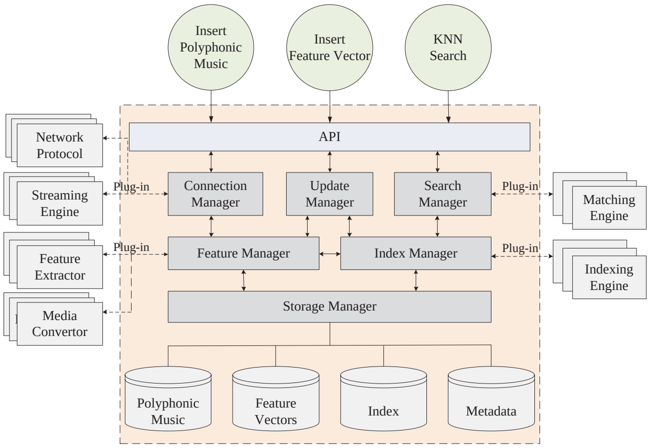


Fig. 1. Block diagram of software stack for the server

The server initializes and updates DBs when checking new music files added. It creates a new DB thread for adding new music files while serving other requests from the clients. It also

[1]This work was supported by the Ministry of Knowledge Economy (MKE) grant funded by Korea government (No.10037244).

C. Song and H. Park are with Electronics Engineering Department, Kwangwoon University, Seoul, Korea (e-mail: jcsong@keti.re.kr).

C. Song, C. Yang, S. Jang and S. Lee are with Digital Media Research Center, Korea Electronics Technology Institute, Seoul, Korea.

makes a new connection thread for every client connection request respectively in order to support multi connection. We introduce alarm method to notify new music file addition to the update manager. It can reduce the system overhead from the polling method.

B. DSP module for embedded clients

Embedded platform has the performance limitation because it is designed for the specific purpose using system on chip (SoC). The margin of the performance is not enough for a new service, which requires revising its hardware. However, it is not easy to revise the hardware due to the financial and timing issues especially. For this reason, DSP module is developed in this study. It can provide the QbSH service to the embedded platform without any hardware revision. It has the universal serial bus (USB) interface to communicate with host platform like set-top box (STB) and karaoke. The feature extraction from user’s query is ported into this module. We optimize this function through three steps. At the initial step, we get over 390 million clocks per frame. It goes down to around 870 thousand clocks at middle stage. Eventually we reach to 347 thousand clocks. In addition, the following optimization techniques are included:

- ▶ utilizing the look-up tables for fixed functions like *cos.*
- ▶ using the optimized functions given by the compiler.
- ▶ making the pipelines using the parallel operations.
- ▶ moving the variables to the internal memory.
- ▶ reducing the count of load and store instructions.

III. EVALUATION AND IMPLEMENTATION

For the system evaluation, we build 3 polyphonic music datasets named Audio Feature Analysis (AFA100), AFA450, and AFA2000 which contains 100, 450 and 2,000 files respectively. The dataset consists of 8 genres like dance, ballad, trot, children, rock, R&B, pops and carol. Length of each play varies between 4 and 6 minutes except for children and carol.

For the input query of the system, the query dataset is recorded from 3 women and 18 men by singing/humming a part of the song randomly for 12 seconds. We set up the real world recording environment as much as possible like class room, office, and living room. Every music track of AFA100 has 12 query clips respectively. So, the query dataset contains 1,200 query clips. It is clustered into 3 parts that are the intro, the climax, and the rest of the music tracks. Interestingly, we figure out that the intro part is over 60%, the climax part is about 30% and the rest part is under 10%. It was generally expected that the climax part would be more than the others.

A. Evaluation

We use the mean reciprocal rank (MRR) of top 20 matches to evaluate the system performance. The MRR is widely used for measuring the accuracy of the query system in many researches [5]. We also measure the percentage of the correct answer included in top 1, top 5, top 10, and top 20 matches. The length of input query is 8, 10, and 12 seconds. Fig. 2 shows the result of the system performance. We figure out that the input query over 10 seconds yields good performance and the most matches are in the top 5. It shows the similar result for AFA450 and AFA2000. The search per a query with 12 seconds length takes about 2.9 and 8.6 seconds in average for AFA450 and AFA2000 respectively.

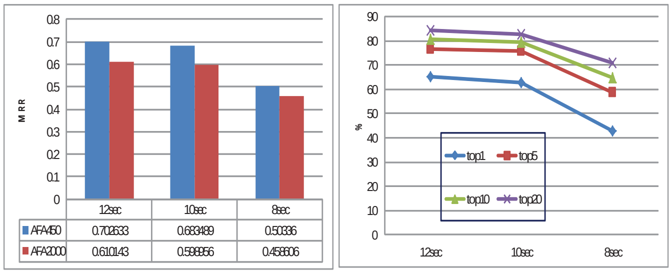


Fig. 2. The performance of the system; Left) MRR, Right) Ratio of the ranks

B. Various types of clients and the applications

Using the developed QbSH system, we implemented the music retrieval service for the various types of platforms. Fig. 3 shows the implementation result and three different models of the music retrieval service. The first one is for the mobile devices like smart phone/pad as model 1. The second one is for the web service as model 2. The last one is for the embedded platforms like STB as model 3.

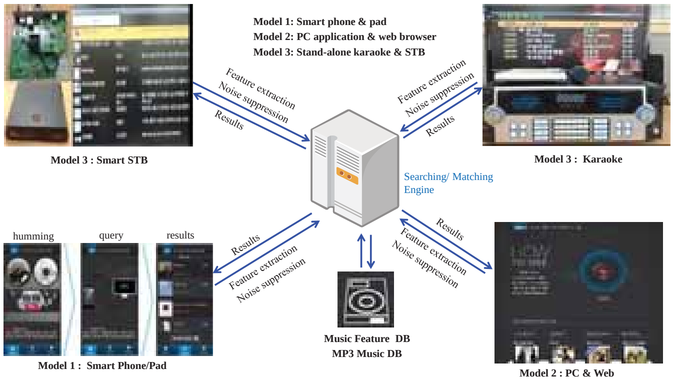


Fig. 3. Implementation result with three application models

REFERENCES

[1] J. S. R. Jang and M. Y. Gao, "A query-by-singing system based on dynamic programming," *in Proc. Int. Workshop on Intelligent Systems Resolution*, pp. 85-89, 2000.

[2] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith, "Query by Humming: Musical Information Retrieval in an Audio Database," in *Proc. of the third ACM Int. Conf. on Multimedia*, pp. 231-236, 1995

[3] K. Kim, K.R Park, S.J Park, S.P Lee and M.Y Kim. "Robust Query-by-Singing/Humming System against Background Noise Environments," *IEEE Trans. on Consumer Electronics*, vol.57, no.2, pp. 720-725, May 2011

[4] D. Jang, C. Song, S. Shin, S. Park, S. Jang, and S. Lee, "Implementation of a matching engine for a practical query-by-singing/humming system," *Proc. ISSPIT*, Dec. 2011

[5] J. S. Downie, “The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research”, *Acoust. Sci. & tech*, 29, 4, 2008

Automatic Recognition of Major End-Uses in Disaggregation of Home Energy Display Data

Michael ZEIFMAN, *Senior Member*, *IEEE*, Kurt Roth, and Johannes STEFAN
Fraunhofer Center for Sustainable Energy Systems, Cambridge, USA

***Abstract*—Disaggregation can make actionable the information provided by home energy displays or smart meters. However, the known disaggregation methods require training to match disaggregated data to actual appliances. We propose a statistical approach for automatic matching between the disaggregated data and major end-uses.**

I. INTRODUCTION

Home energy displays (HEDs) usually provide whole house, near real-time electricity consumption information [1]. The primary goal of HEDs is to help save energy for the homeowner. Since the whole-house information is not actionable, data disaggregation, or nonintrusive appliance load monitoring (NIALM), can be implemented to provide appliance-specific information [2]. However, available NIALM methods suitable to the HED data (i.e., real power sampled at 1 Hz) require significant occupant efforts to train the algorithms to recognize the actual appliances [2], [3]. In this paper, we propose an approach for automatic recognition of the major end-uses in the disaggregated data. We implement power- and time features for statistical characterization of the end-uses and apply a simple Bayes classifier to matching between the disaggregated patterns and appliance classes.

II. METHODOLOGY

A. NIALM Method

To disaggregate the low-frequency HED data, NIALM algorithms usually use stepwise power changes as the main feature. These stepwise power changes occur when, e.g., the on-off appliances are turned on or off. Since the distributions of the power draw of appliances often overlap, the accuracy of traditional NIALM methods is usually low [2]. We have recently developed a new NIALM method that implements time features, duration of time on and duration of time off, to better separate the overlapped in power draw appliances [3]. The method is capable of finding on-off appliances or their combinations and tracking them in time, but it yet cannot match the found appliances to the actual household appliances. Figure 1 illustrates this problem.

B. Major End-Uses

Among various appliances and appliance groups, the following end-uses account for more than 80% of average electricity household consumption [4]: (1) space cooling systems, (2) space heating systems, (3) domestic hot water, (4) lighting, (5) refrigerators, (6) electric clothes dryers, and (7) consumer electronics. All of these end-uses, but dimming lights that are not considered in this work, can be approximated as on-off appliances. In this work, we concentrate on these seven end-uses. For the consumer electronics, we consider televisions and desktop computers with monitors as these devices account for more than 85% of consumer electronics energy consumption [4].

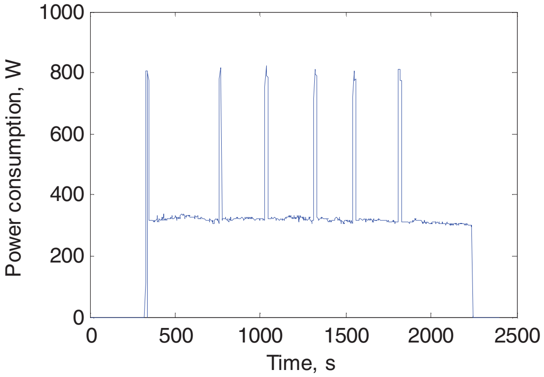


Figure 1. Power consumption of an appliance disaggregated by our NIALM algorithm. Which actual appliance does this pattern correspond to? (clothes dryer). Note that this particular appliance is a combination of two on-off devices, motor and heating element.

C. Statistical Characterization by Prior Knowledge

Prior data exist for the power draw information of the seven major end-uses we selected. For example, Ref. [5] lists average lighting power draw, number of bulbs, number of switches, and average duration of time on for different room types (e.g., living room, kitchen), building establishments (e.g., multifamily) and bulb types (e.g., incandescent, CFL or halogen). These data can be used in designing statistical distribution models.

In this work, we use the maximum entropy concept [6] to select a statistical model that is most suitable to the available prior knowledge. This concept yields a Beta distribution for the case of the known range and mean value [7]. However, the underlying random variables, the duration of time on and the change of power, are not statistically independent for the lighting loads. This dependence is based on the room type. For each room type, nonetheless, these variables can be assumed to be independent so that the joint probability density function (PDF) for each room is a product of the marginal PDFs. For the entire household lighting, the joint distribution function will be a mixture of the joint PDFs corresponding to the room types.

Advantage of the discrete wattage values of the bulbs on the market can be taken to better characterize the distribution of the power draw. Assuming that the fluctuations of the power draw around the face value can be characterized by a normal

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

distribution, we arrive at a convolution of normal and beta distributions for the power draw of lighting.

Figure 2 shows a marginal PDF of power draw for lighting loads in a living room of a single detached U.S. home. The characteristic peaks on the PDF are due to the discrete wattage of the bulbs.

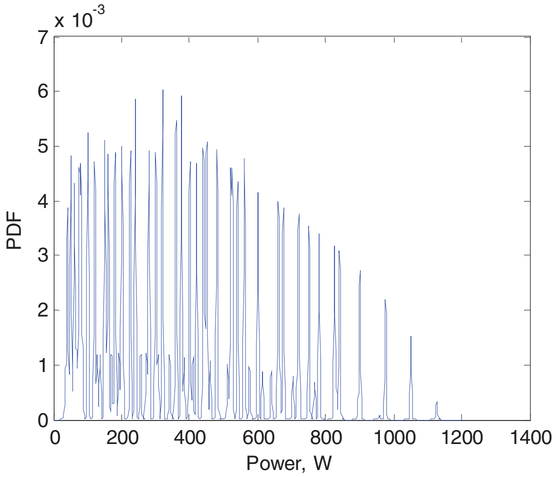


Figure 2. Marginal PDF of lighting loads for a living room.

The other six end-uses are easier to characterize, because the underlying prior knowledge is not as granulated as the knowledge on the lighting loads. We get the joint Beta distributions for the power and durations of time on/off using the average values from the literature. Note that the power off distribution of, e.g., the clothes dryers is a mixture of two Beta distributions, with one component corresponding to the cycling load of the heating element (see Figure 1) and the other component corresponding to the time between drying cycles.

D. Statistical Characterization by Fine Features

To further improve the statistical characterization, we performed signature collection for the seven end-uses. Typical signatures for lighting and televisions are shown in Figures 3 and 4.

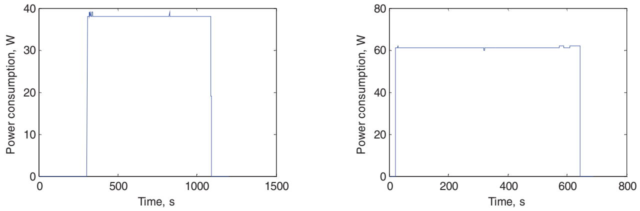


Figure 3. Time dependent power consumption of lamps with incandescent bulbs. Left panel: 40-W bulb, right panel: 60-W bulb. Note that the actual power draw deviates from the face value.

It is seen in the Figures that the lighting load is very steady, and that the televisions are characterized by either significant fluctuations of the power draw due to the variability of brightness and sound (CRT) or by typical power drops at channel changes (LCD). These fine features can be characterized mathematically and be used in conjunction with the general prior knowledge (see Section II.C).

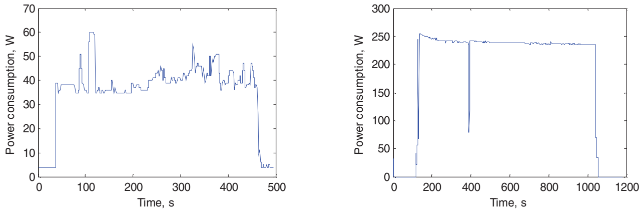


Figure 4. Time dependent power consumption of televisions. Left panel: CRT, right panel: LCD. CRT started from a standby mode. Characteristic power drop for LCD corresponds to channel change.

III. TESTING

We have designed a naïve Bayes classifier [6] on the basis of the statistical models for the seven end-uses considered in Section II. For all other possible loads, we use a single uniform joint distribution of power/time. We then conducted testing in residences using the TED home energy displays [8] for both aggregated power and submetering. The testing was performed during a cooling season so that only six end-uses were present at the test homes.

The disaggregation was performed by our set of algorithms [3]. The obtained disaggregated data were matched to the six end-uses or to the class “else.” For the most challenging end-uses, lighting and consumer electronics, the classification accuracy in terms of the F-measure [3] was 65% and 70% respectively if we used only the prior-knowledge characterization. The use of the fine features dramatically increased the accuracy, to 92% for the lighting and to 90% for the consumer electronics.

IV. CONCLUSION

A statistical classification scheme, capable of automatic recognition of disaggregated HED data is proposed and preliminary tested. The scheme implements prior knowledge on the major electric residential end-uses along with their fine features to get reasonable classification accuracy. More comprehensive tests of the scheme and development of more complex appliance models are currently underway.

REFERENCES

[1] Erhardt-Martinez, K. (2010), et. al., “Advanced Metering Initiatives and Residential Feedback Programs: a Meta-Review for Household Electricity-Saving Opportunities,” ACEE Report E105.

[2] Zeifman, M. and K. Roth (2011), “Nonintrusive Appliance Load Monitoring: Review and Outlook,” *IEEE Transactions on Consumer Electronics* 57, p. 76-84.

[3] Zeifman, M., “Disaggregation of home energy display data using probabilistic approach,” *IEEE Transactions on Consumer Electronics* 58, pp. 23-31, 2012.

[4] Building Energy Data Book 2009. US Department of Energy.

[5] Ashe, M., et al., “2010 U.S. Lighting Market Characterization,” Report to U.S. Department of Energy, 2012.

[6] Tou, J. T. & Gonzalez, R. C. (1974) Pattern Recognition Principles (Addison-Wesley).

[7] Lee, R.C. and W.E. Wright (1994), “Development of human exposure-factor distributions using maximum-entropy inference,” *Journal of Exposure Analysis and Environmental Epidemiology*, pp. 329-341.

[8] TED – The Energy Detective, home energy display manufactured by Energy, Inc.

BluePot: An Ambient Persuasive Approach to Domestic Energy Saving

Qi LIU, *Member, IEEE*

***Abstract*--Ambient persuasive technologies positively influence user behaviors by motivational strategies. In this paper, an ambient persuasive approach, called BluePot is presented to save energy and can achieve up to 13% of energy usage at homes.**

I. INTRODUCTION

Recent research on domestic energy management has shown that cost, installation complexity and lack of interactive strategies become a burden to deploy smart grid at home [1]. Presenting energy consumption in a clear way and using computerized tools can be critical to achieve energy saving and efficiency [2].

In this paper, an ambient persuasive solution, called BluePot is presented in a domestic energy management system. A Bluetooth enabled photo frame is used playing persuasive pictures helping households to be aware of energy consumption and hence achieve energy saving.

This paper focuses on the design and implementation of the ambient approach. According to the experiment results, over 10% of energy saving has been achieved.

II. RELATED WORK

Ambient persuasive technologies aim to positively influence user attitudes and/or behaviors [3, 4], and have shown greater effects than regular persuasive methods [4]. An Ambient Intelligent can be formed with massive embedded devices being connected and networked, sharing and distributing information and intelligence in the users' environment [5].

Research efforts have been offered in order to take advantages of ambient persuasive technologies to the domestic energy monitoring and saving systems. The WattBot project [6] monitored energy consumption and offered feedback via smart mobile phones. The ténéré tree was virtualized in [7] that visualized the famous tree shedding its leaves when energy was being overused using a LCD display. In [8] glowing power cords were used to indicate power consumption. These applications have shown great power on energy saving using ambient persuasive technologies to fulfill local and single user's requirements.

III. SYSTEM DESIGN OF BLUEPOT

A. Display hardware

A display is used in BluePot to interact with our end users by showing them persuasive pictures containing information of their current energy consumption. As the only output media, the display needs to be stable, robust, and simple to use. A Bluetooth enabled photo frame was finally chosen for Bluetooth and photo frame together meet all our display hardware requirements. Once paired, persuasive picture can be transferred to the photo frame without any additional dougles or cables, or manual operations. Plus a regular photo frame nowadays provides sufficient built-in memory for the storage of photographs with extension card slots for extra space.

B. Persuasive images

Persuasive images are a set of pictures that implies the status of current energy consumption of a user's house, as shown in Fig. 1. The first three pictures, Fig.1 (a) – (c) illustrating the growth of a sunflower are used during the pilot stage, when the baseline is established. After that, when the system steps into the regular monitoring stage, the energy usage a household consumes will be evaluated and shown as one of the pictures in Fig. 1 (d) – (i) from a shine glory flower to a lonely pot with the flower wilted.

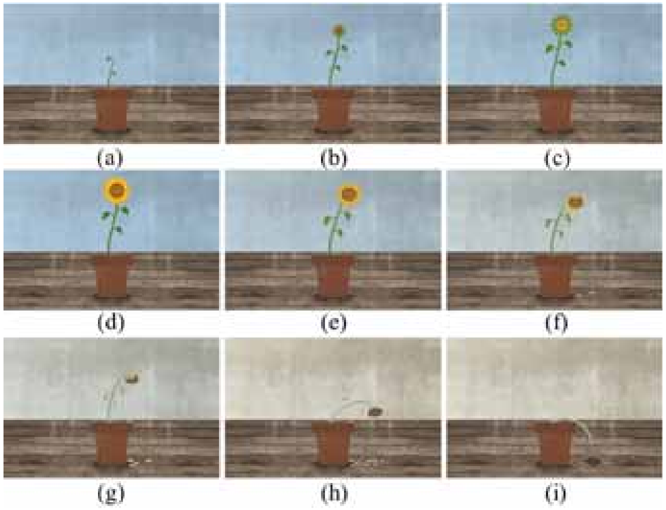


Fig. 1. Persuasive images to indicate how good/bad the energy is used in a user's house. (a) - (c) depict the pilot stage, when the base line is established. (d) - (i) show the present usage of a household's energy usage.

DEHEMS is funded from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement No.224609. This work is also supported by A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

At present, the energy usage is evaluated according to the current power consumption; however, this method suffers from the difficulty to define practical thresholds between the persuasive pictures. A household profile is being proposed in order to position a house in the right usage category with consideration of the number of bedrooms and family occupants, and the property type. Corresponding persuasive pictures can then be generated according to the current energy usage and the category the house fits.

C. System development: integration with DEHEMS

The energy usage in each user's house is retrieved by a domestic energy management system, DEHEMS [9], where the total power consumption is retrieved via JSON formatted web request APIs. The DEHEMS receives energy data from a number of different sensing technologies, including electrical mains circuit sensing, individual appliance-level sensing and gas mains sensing. Implementation of the ambient interface helps to infer and reason the energy behavior of the households and to test the effectiveness of the innovative persuasive strategies.

D. Enabling community contribution and competition

The concept of community is also defined in the BluePot supported by DEHEMS, where five living labs are defined in five European cities. A set of community pictures is also designed using a garden to indicate the current energy consumption of the community, as shown in Fig. 2. Every member in the living lab contributes their energy usage to make the community garden either look sunny and cheerful (Fig. 2 (a) – (c)), or cloudy and gloomy (Fig. 2 (d) – (f)).

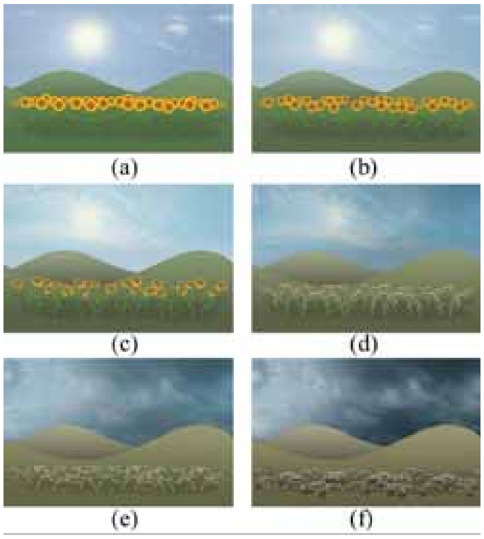


Fig. 2. Persuasive images to indicate how good/bad the energy is used in a community.

The evaluation is based on the competition of the five living labs. Instead of focusing at the current power consumption, the status of the community garden is determined according to its rank in all five living labs.

IV. EXPERIMENTS AND RESULTS

The BluePot has been installed in 10 houses for testing and evaluation purposes. The baseline has been established that records energy consumption for two weeks. In each house, a Bluetooth gateway is implemented in order to gather energy information from the DEHEMS server, pair with the Bluetooth photo frame, and determine and forward appropriate persuasive pictures to the device. As shown in Table 1, the results indicate that 9% - 13% of energy saving is achieved.

TABLE I
ENERGY SAVING USING BLUEPOT

Week	Avg. Energy Usage (Kwh)	Energy Saving Compared to the Baseline[a]
1	73.742	9.1%
2	71.497	11.9%
3	70.709	12.9%
4	70.592	13%

[a] The average energy consumption in the baseline is 81.155kwh.

V. MATH

A new ambient persuasive approach, BluePot has been presented in this paper, in order to achieve domestic energy saving. A Bluetooth photo frame is used to indicate a user's current energy usage via a set of pictures. The difference of the pictures implies the severity of the depletion. According to the experiment results, an average of 9 – 13% of energy can be saved via this approach.

EXAMPLES OF REFERENCE STYLES

[1] M. Zeifman, "Disaggregation of Home Energy Display Data Using Probabilistic Approach", *IEEE transactions on Consumer Electronics*, vol. 58, no. 1, pp. 23-31, 2012.

[2] C. Fischer, "Feedback on household electricity consumption: a tool for saving energy?", *Energy Efficiency*, vol. 1, no. 1, pp. 79-104, 2008.

[3] B. J. Fogg, *Persuasive technology: using computers to change what we think and do*, Morgan Kaufmann, Menlo Park, 2003.

[4] E. H. L. Aarts, P. Markopoulos, and B. E. R de Ruyter, "The persuasiveness of ambient intelligence", In *Security, privacy and trust in modern data management*, Petkovic M, Jonker W, Eds. Springer, Berlin, 2007.

[5] E.H. L. Aarts and B. E. R. Ruyter, "New research perspectives on ambient intelligence", *J Ambient Intell Smart Environ*, vol. 1 pp. 5–14, 2009.

[6] D. Petersen, J. Steele, and J. Wilkerson, "WattBot: a residential electricity monitoring and feedback system", In *Proceedings of the 27th international Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '09*, pp. 2847-2852, April 2009.

[7] J. Kim, Y. Kim, and T. Nam, "The ténéré: design for supporting energy conservation behaviors", In *Proceedings of the 27th international Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '09*, pp. 2643-2646, April 2009.

[8] A. Gustafsson, and M. Gyllenswärd, "The power-aware cord: energy awareness through ambient information display", In *Extended Abstracts on Human Factors in Computing Systems, CHI '05*, pp.1423-1426, April 2005.

[9] V. Sundramoorthy, G. Cooper, N. Linge, and Q. Liu, "Domesticating Energy-Monitoring Systems: Challenges and Design Concerns", *IEEE Pervasive Computing*, vol. 10, no.1, pp. 20-27, 2011.

Developments of the In-Home Display Systems for Residential Energy Monitoring

Dong Sik Kim*, Beom Jin Chung**, Sung-Yong Son**, and Jeongjoon. Lee***
*Hankuk University of Foreign Studies, Automan Co. Ltd **Gachon University ***LS Industrial Systems Co., Ltd

Abstract—**In order to efficiently reduce the amount of the electricity usage in the residential area, the demand response (DR) of the consumers is of importance. The in-home display (IHD) system provides energy monitoring information for the consumer DR. Recently, we have developed several types of IHD systems, which are based on 2.4GHz ZigBee, the power line communication technique, and the sub-GHz narrow-bandwidth radios. In this paper, different types of IHDs are introduced and their technologies including network architectures are compared.**

I. Introduction

The residential sector, unlike the commercial or industrial sector, is composed of multiple small energy consumers and wastes significant amount of supplied electric power. This fact implies the potential of the energy saving in the residential side [1]. In order to efficiently reduce the amount of electricity waste in the residential side, the consumer needs to change the usage behavior. Research in real-time energy monitoring has been conducted recently. From the monitoring, the energy consumption patterns are acquired and can be analyzed for the reduction of the energy consumption. Through extensive research, it is revealed that the real-time direct display, which is provided by the *in-home display* (IHD) system, can significantly decrease the energy waste, especially for the inclining block rate case. The price rates are changed for consumptions over several tiers of certain amounts per billing period in the inclining block rate system. For example, Korea Electric Power Corp. (KEPCO) has inclining block rates with 6 tiers, which consist of 5 tiers of up to every 100kWh and a tier of above 500kWh. Based on the automatic meter reading (AMR) infrastructure, the area AMR server gathers the metering data from the *smart meter* [2] and transmits electricity usage information including the metering data to the IHD system. The IHD system is composed of the *area IHD server*, which receives the electricity usage information from the AMR server, and the *IHD device* of the residential side for consumer's monitoring the energy usage.

We have conducted several projects to develop and implement the IHD systems for different environments and areas in Korea. The developed IHD systems can be classified into the following two approaches. The first approach provides a low-cost system with minimal additional hardware under an existing infrastructure. In the second approach, each smart meter site individually transmits the electricity usage information to the corresponding IHD device. In this paper, the developed IHD systems, which are especially based on the wireless links within the categories of the two approaches, are introduced and compared. The developed IHD systems have been practically tested and used by more than 30,000 households recently in Korea.

II. Low-Cost In-Home Display System Based on the Star Connection

The development of the low-cost IHD system, which belongs to the first approach, has the following properties:

- Star connection for direct transmission.
- Long-range narrow-bandwidth (NB) radios.

This system directly transmits the electricity usage information from the area IHD server site to each IHD device site via repeaters in a form of the star connection. In order to provide a long-range wireless link, NB radios are employed. Even though the data rate is quite low, the minimal additional hardware can guarantee an affordable construction of the IHD system. Hence, the approach is appropriate for constructing a low-cost IHD system without modifying the existing metering systems at a quite low price.

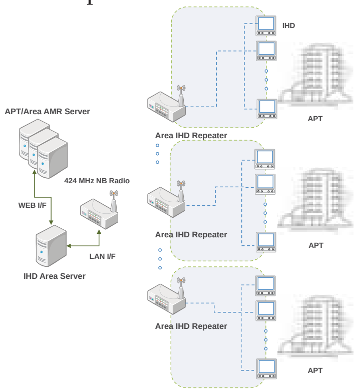


Fig. 1. Low-cost in-home display system based on the narrow-bandwidth radios of the 424MHz band in the star connection.

The configuration of the low-cost IHD system is illustrated in Fig. 1. In order to simplify the wireless data transmission, the area IHD server directly transmits data to each IHD device via the area IHD repeaters. The wireless link, which is conducted by the NB radios, covers up to 2-3km for an apartment complex, and has 12.5kHz channel spacing with the 8.5kHz bandwidth. Here, the narrow bandwidth ensures high receiver sensitivities and thus provides long-range wireless links. The 424MHz or 447MHz band is used for the NB radios in compliance with Korea Communications Commission (KCC). The battery-operated IHD device has a NB transceiver and the information display part, which is a customized segment type LCD to minimize the battery consumption as shown in Fig. 2.

In this low-cost IHD system, the residential consumer activates the data request by pressing a button of the IHD device or the IHD device automatically requests the data periodically. The area IHD server then responses to each IHD

This research was partly supported by the MKE (Ministry of Knowledge Economy), Korea under the support program supervised by KETEP (Korea institute of Energy Technology Evaluation and Planning), (2010101040003B).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

device. The data, which includes the electric usage and the current electric charge or rate, is composed of the 200Byte data. However, since the transmission rate of the NB radio is as low as 2.4kbps, we have a problem that transmitting these data occupies the wireless link channel for a long period. To solve this problem, we transmit the data by classifying into three types of packets, of which contents possess similar properties, respectively. The *complex common packet*, which includes the date, temperature, and weather data, is common for all home. Hence, this packet is broadcasted to the all IHD devices at a specified time, when the IHD devices awake. Other two types of packets, which are different depending on the home, are the *home daily* and *home real-time packets*. The former is updated ones a day, but the later is updated frequently as the consumer's request.

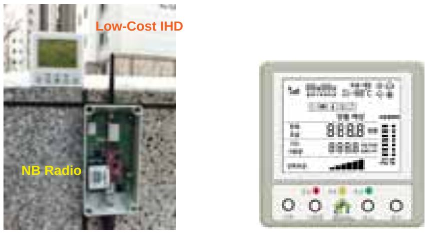


Fig. 2. 424 GHz NB radio in the area IHD server site and the IHD device for the low-cost IHD system.

By conducting an experiment for an apartment complex with 1,000 households in a north area of Seoul, Korea, we find that 23.4% of data amounts are saved and 31.5% of battery consumption in miliampere-hour is reduced, by classifying the packets.

III. In-Home Display System Based on the Point-to-Point Wireless Connection

The developed IHD systems, which are based on the second approach, have the following properties:

- Point-to-point fast transmission
- Short-range ZigBee or NB radio

Each smart meter is connected with the AMR server through the corresponding *interface module*, which is adjacent to the smart meter, based on the wired serial communication channels, such as ISO485. Hence, attaching or adding a wireless radio to the interface module can efficiently manage data between the AMR server and the IHD device in a form of the point-to-point wireless connection if we can change or modify the interface module. Since the smart meter is usually located outside but relatively close to the consumer home, a short-range radio is enough to transmit the data to the IHD device. Here, we consider ZigBee as well as the NB radios depending on the residential environments. The interface module can also be connected with the heating and warm-water meters since the meters including the smart meters are usually gathered within a place. Hence, this structure is appropriate for the unified metering systems including the gas meter and water meter reading. Since each smart meter is equipped with the interface module and a radio, the implementation cost is relatively high compared to the former approach case.

The interface module, which is a media convertor from ISO485 to the NB radio RF signal, is shown in Fig. 3. The NB radio is included within the interface module with a built-in antenna and can obtain a good wireless link margin even for a complicate structure of the apartment complex. The wireless communication burden of this system is much smaller than the case of the low-cost IHD system since an NB radio is only responsible for one IHD device with 30meter coverage. The IHD device has a color TFT display with a touch panel and is powered by an AC adaptor as shown in Fig. 3.

Fig. 4 shows another IHD system that employs the ZigBee smart energy profile (SEP) 1.0. Compared to the case of Fig. 3, in which the interface module is a simple media convertor, the interface module of Fig. 4 has the ZigBee radio and stack, and plays a role of the ZigBee coordinator. The ZigBee radio should be very close to the IHD device due to the short link range below 10meters of 2.4GHz, and hence the ZigBee radio part is separated from the interface module as shown in Fig. 4.

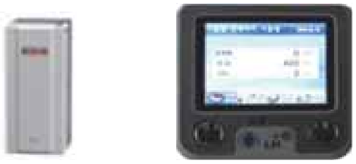


Fig. 3. 424MHz NB radio-based interface module and the color IHD device..

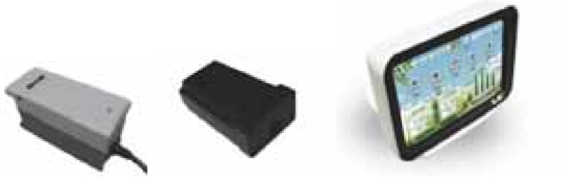


Fig. 4. 2.4GHz ZigBee-based interface module and the ZigBee radio, and the color IHD device.

IV. Conclusions

Since the dependency on the physical layer for home networks will no longer be a problem in SEP 2.0, we might expect that various energy saving devices with radio modules, such as WiFi, power line communication (PLC), ZigBee, etc. can coexist. Hence, careful selection of the communication media as well as the network architecture is of importance to guarantee stable deliveries of the effective information to the customers. Through the developments of the IHD systems in the home area, we can notice that the NB radios, which are based on the sub-GHz smart utility network (SUN), are more appropriate for an enough wireless link margin than the 2.4GHz-based ZigBee case.

References

[1] M. A. Alahmad, P. G. Wheeler, A. Schwer, J. Eiden, and A. Brumbaugh, "A comparative study of three feedback devices for residential real-time energy monitoring," *IEEE Trans. Industrial Electronics*, vol. 59, pp. 2002-2013, Apr. 2012.

[2] S. Ahmad, "Smart metering and home automation solutions for the next decade," in *Proc. IEEE Int. Conf. Emerging Trends in Networks and Communications*, Apr. 2011, pp. 200-204.

Smart Heating and Air Conditioning Scheduling with Customer Convenience in a Home Energy Management System

Hyung-Chul Jo, *Student Member*, IEEE, Sangwon Kim, and Sung-Kwan Joo, *Member*, IEEE
The School of Electrical Engineering, Korea University, Seoul, Korea

***Abstract*-- A home energy management system (HEMS) is expected to play an important role in saving energy costs under time-varying electricity price in a smart home environment. The development of HEMS requires studies on various energy resources in a home, such as energy storage system, and fuel cell. However, in HEMS, very limited research about heating and air conditioning scheduling with customer convenience has been conducted. This study presents a smart heating and air conditioning scheduling method that considers customer convenience and characteristics of the thermal device and an optimization-based approach to minimize the cost in HEMS.**

I. INTRODUCTION

Time-varying retail pricing schemes are being designed by utility companies in order to reduce the increasing energy demand. A home energy management system (HEMS), represented in Fig. 1, can play an important role in saving energy costs under time-varying electricity price in a smart home environment. Wireless communication networks that control and monitor the appliances in HEMS have been studied. However, very limited research has been conducted in the area of scheduling algorithms that control the thermal devices in a home, such as heating, ventilating, and air-conditioning (HVAC) systems, which have large electrical energy requirements.

In [1], [2], a model and algorithm for HEMS to coordinate the operating schedule of some devices in micro-grid and a building are proposed; however, HVAC has not been considered. The scheduling result from the formulation proposed in [1], [2] is ideal because the cooling or heating demand used for HVAC control is considered as a parameter. In this study, a smart heating and air conditioning scheduling algorithm based on mixed integer non-linear programming (MINLP) considering the mathematical model of HVAC in a home is proposed.

II. MODELING OF HVAC IN SMART HOME

An algorithm for the integrated scheduling of all the energy resources in a home is required for customers desiring to reduce the overall energy cost by the introduction of smart home and time-varying price. The energy resources to be scheduled by the algorithm are restricted to power grid, energy storage system (ESS), electric vehicle (EV), fuel cell (FC) and HVAC because it is impossible to control all the equipment in a home without causing inconvenience to the customer. In this research, several profiles and parameters are assumed to be known from forecasts based on historical data and information obtained from the utility companies.

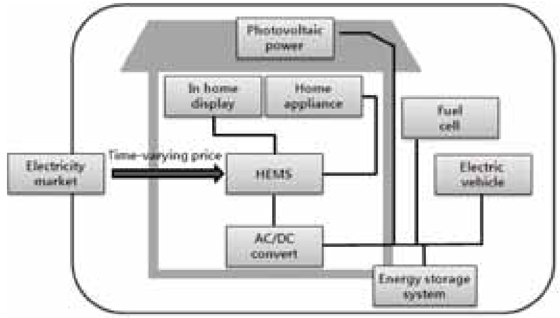


Fig. 1 Concept of the home energy management system (HEMS).

The basic specification for the scheduling algorithm in which the overall energy cost and characteristics of household devices and appliances are considered as the objective function and constraints, respectively, was described in [2]. This study discusses the HVAC model and the convenience of the customer.

A. HVAC model

It is difficult for a customer to frequently modify the HVAC setting in order to reduce the energy cost under a time-varying price. Hence, HVAC is assumed to have an automatic control system based on the control signal obtained from the scheduling algorithm for HEMS. An HVAC model that schedules the operation mode and output by considering the temperature and time-varying price is proposed.

Utility companies provide the information about the coefficient of performance (COP) that represents the ratio of the change in heat to the supplied electrical demand. Hence, the HVAC model is typically specified as follows:

$$q_{HVAC}(t) = COP \times p_{HVAC}(t) \tag{1}$$

where $q_{HVAC}(t)$ is the cooling or heating demand generated by the HVAC system at time t and $p_{HVAC}(t)$ is the electrical demand of the HVAC system at time t.

In order to estimate the power demand for an HVAC system, the cooling and heating demand should be calculated by considering the COP and the temperature in the discrete time domain. For this model, it is assumed that the composition and the volume in indoor air are constant.

The indoor cooling or heating demand based on the first law of thermodynamics is specified as follows:

$$\begin{aligned} q_{load}(t) &= \alpha \cdot (T_{in}(t+1) - T_{in}(t)) + ... \\ &\{Z_{HVAC}(t) \cdot (\beta_{on} - \beta_{off}) + \beta_{off}\} \cdot (T_{out}(t) - T_{in}(t)) \end{aligned} \tag{2}$$

where $q_{load}(t)$ is the cooling or heating demand at time t; $T_{in}(t)$ is the indoor temperature at time t; $T_{out}(t)$ is the outdoor

This work was supported by the Human Resources Development of the Korean Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Knowledge Economy (No. 20114010203010).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

temperature at time t; $Z_{HVAC}(t)$ is a variable representing the operation of HVAC at time t; α is the indoor heat capacity; and β_{on} and β_{off}, estimated from historical data, are parameters reflecting the outdoor temperature.

B. *Convenience of the Customer*

In order to minimize the overall energy cost, the HVAC need not be operated when a home is unoccupied. An estimation of the occupancy of a home by customer is needed to determine the HVAC operating time. For the estimation of the customer occupancy in a home, an algorithm to learn the manual selection of the customer has been proposed in [3]. Nonintrusive load monitoring (NILM) described in [4] could also be an alternative method to estimate the unoccupied period in a house. However, NILM requires sensors, and hence, the algorithm that learns the manual selection of customer has a greater possibility of being commercialized than the algorithm based on NILM. In this study, the learning algorithm [3] is adopted for estimation of customer occupancy.

The indoor temperature and the HVAC operating time should be limited to the preferred temperature established by the customer within the HVAC operating time *S*. This requirement can be specified as follows:

$$T_{\min} \le T_{in}(t) \le T_{\max}, \ \forall t \in S \qquad (3)$$

where T_{min} and T_{max} are the minimum and maximum values, respectively, of the indoor temperature.

III. HEATING AND AIR CONDITIONING SCHEDULING METHOD WITH CUSTOMER CONVENIENCE

The solution to the problem formulated as MINLP by the above model could be global or local optimization depending on the convexity. If the problem is not convex, it is necessary to verify that the solution obtained is the global optimal solution. Owing to the complexity of this process, the problem must be converted to a convex form by convexification.

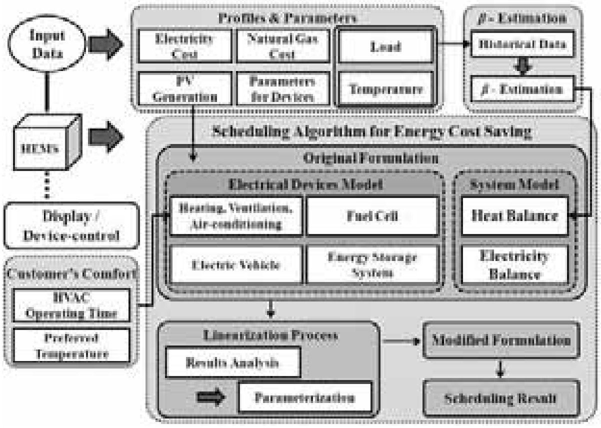


Fig. 2 Overview of the proposed scheduling algorithm for HEMS.

Constraint (2), proposed in section Ⅱ, is in the form of a product of the decision variables $Z_{HVAC}(t)$ and $T_{in}(t)$. This non-linear form causes a significant reduction in convergence ratio of the algorithm for HEMS. In order to improve the convergence ratio, the algorithm must be transformed to mixed integer programming (MIP) by a linearization process that transforms the decision variables to parameters. Fig. 2 represents an overview of the proposed algorithm for HEMS.

IV. DISCUSSION

In order to test the proposed method, a scenario based on data collected from several institutions during winter is created for the simulation. In Fig. 3, the energy cost is obtained from the simulation by considering different values in the T_{min} or T_{min} setting at home with ESS and FC. In Fig. 3, a case with a fixed indoor temperature represents the case in which T_{max} is equal to T_{min}. It can be seen from results in Fig. 3 that the energy cost can be reduced by setting the preferred indoor temperature to the specified range rather than a fixed indoor temperature while avoiding inconvenience to the customer. Further, the energy cost increases by approximately 2.5% when T_{max} is equal to T_{min} and fixed temperature increases by 1 degree.

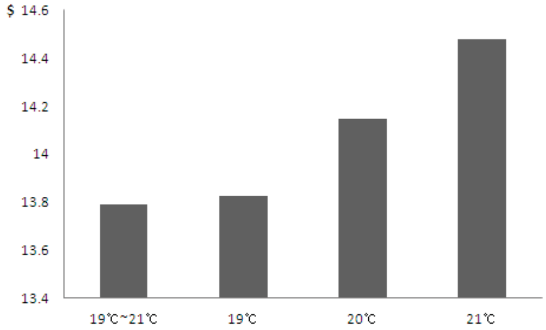


Fig. 3 Comparison of energy cost considering convenience of customer.

V. CONCLUSION AND FUTURE WORK

This paper described an HVAC model that considers the convenience of the customer and a method to solve the scheduling problem with a model for HEMS. The algorithm provides an optimal scheduling of energy resources and avoids inconvenience to the customer. This method is expected to reduce the cost for household owners because it can be applied to a variety of home environments.

This study focused on the temperature control in a home. However, it is not useful to operate HVAC systems in the unoccupied zones in a home. An algorithm that controls an HVAC system by using NILM to classify the zones in a home would be necessary.

REFERENCES

[1] C.Chen, S.Duan, T.Cai, B.Liu, and G.Hu, "Smart energy management system for optimal microgrid economic operation," *IET Renew. Power Gener.,* vol. 5, iss. 3, pp. 258-267, 2011.

[2] Xiaohong Guan, Zhanbo Xu, and Qing-Shan Jia, "Energy-Efficient Buildings Facilitated by Microgrid," *IEEE Trans. on Smart Grid*, vol. 1, no. 3, pp. 243-252, Dec. 2010.

[3] Nest Labs., "White Paper: Nest Learning Thermostat Efficiency Simulation: Update Using Data from First Three Month," Apr. 2012

[4] C.Laughman, L.Kwangduk, R.Cox, S.Shaw, S.Leeb, L.Norford, and P.Armstron, "Power signature analysis," *IEEE Power and Energy Mag.*, vol. 1, iss. 2, pp. 56-63, 2003.

Efficient One-Time Message Authentication Scheme for Metering Data in Smart Grid Systems

Young-Sam Kim and Joon Heo, *Member, IEEE*

***Abstract*—Metering data in smart grid systems constitute commercial information because they are related to charge directly. Thus, manipulation of the metering data has to be prevented, but the limitation of computation power of smart meters makes it difficult. This paper presents an efficient one-time message authentication scheme for metering data in resource-restrained smart meters. The proposed scheme can lower the amount of computation for cryptographic operation, and it can also reduce the size of the key and signature for transmission.**

I. INTRODUCTION

In smart grid systems, metering data are transmitted through communication networks. The metering data could be easily attacked for manipulation because they are directly related to charge [1], as shown in Figure 1. Actually, some recent studies have reported the vulnerabilities of smart grids, especially smart meters [2, 3]. To prevent these manipulation attacks, a message authentication method is needed. In particular, as the metering data are commercial data, a non-repudiation service could be also needed. Verifying the origin of data is very important in smart grid systems, but it is not easy. A smart meter has lower computation power than a management server, and hence, cryptographic functions that need high CPU resources are difficult to apply. Previous studies have shown that some asymmetric key cryptographic methods are not suitable for device or message authentication in smart grids [4, 5, 6]. Instead, they suggest lightweight asymmetric key methods such as Diffie-Hellman key agreement and one-time signature. In particular, as a previous study [6] has highlighted the importance of transmission time limit, the size of data should also be considered.

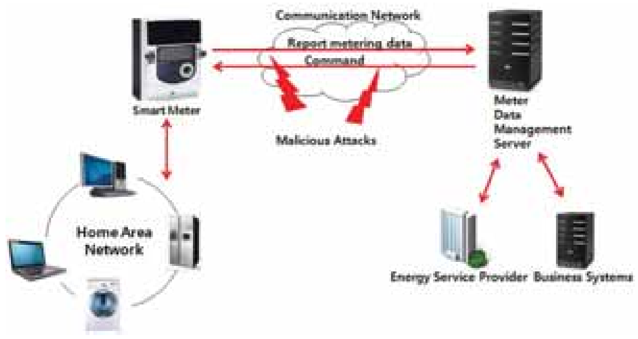


Fig 1. Malicious attack in smart grid systems

In this paper, we propose an efficient one-time signature scheme using homomorphic hash. An idea of generating signature in the proposed scheme is similar to the Hash to Obtain Random Subset (HORS) [7]. HORS is one of the fastest signature schemes and it has been investigated in previous papers [5, 6] for a comparative study. However, a key management mechanism highly affects to the performances of HORS. Time Valid (TV)-HORS [8] has proposed a hash-chain based key management mechanism. However, the time for key generation and signature verification is too long which is hard to apply in resource-constrained devices or time-critical systems.

Thus, we propose a new key management mechanism in order to lower the computation cost. The proposed scheme is probably faster than TV-HORS in terms of key generation and signature verification. In addition, our scheme has a smaller key pair size. We analyze these factors for performance analysis in section 3.

II. PROPOSED ONE-TIME AUTHENTICATION SCHEME

The metering data generated from the smart meter are sent to the Metering Data Management Server (MDMS) periodically. The homomorphic hash that will be used in the proposed scheme combines three properties, namely, homomorphism, one-wayness, and collision-resistance. A previous study [9] has employed homomorphic hash based on exponential operation, and we follow the same method. Note that any function satisfying the three properties stated above is permitted for homomorphic hash.

Table 1. Notation List.

Notation	Meaning
$sk_{<i,j>}$	*j*-th element of secret key at a sequence *i*
$pk_{<i,j>}$	*j*-th element of public key at a sequence *i*
$su_l^{(m)}$	*m*-th element of hash chain for updating *l*-th element of secret key
$pu_l^{(m)}$	*m*-th element of hash chain for updating *l*-th element of public key
t	the number of elements of a key pair
k	the number of elements that construct a signature
H	cryptographic hash function
HH	homomorphic hash function

The proposed message authentication scheme is constructed like the HORS scheme using homomorphic hash. HORS is a very fast signature scheme, but its efficiency mostly depends on a one-time key management mechanism. The outline of our key update mechanism is shown in Figure 2.

According to HORS, a signature is a part of the secret key elements, i.e., ($sk_{<1,1>}$, $sk_{<1,2>}$, $sk_{<1,3>}$) in Figure 2. This signature will be sent to the recipient and disclosed in an unauthenticated channel. Therefore, those elements must be updated immediately after use.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

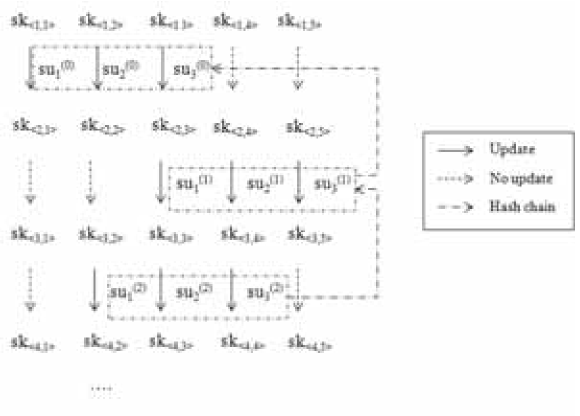

sk<1,1> sk<1,2> sk<1,3> sk<1,4> sk<1,5>
su1(0) su2(0) su3(0)
sk<2,1> sk<2,2> sk<2,3> sk<2,4> sk<2,5>
su1(1) su2(1) su3(1)
sk<3,1> sk<3,2> sk<3,3> sk<3,4> sk<3,5>
su1(2) su2(2) su3(2)
sk<4,1> sk<4,2> sk<4,3> sk<4,4> sk<4,5>
....
Update
No update
Hash chain

Fig 2. Proposed update mechanism for a secret key

TV-HORS, which proposes a key management mechanism for HORS, uses the *t* hash chain such that *t* is the number of the secret key elements. However, many elements are skipped when verifying a signature because only *k (*which is much smaller than *t*) elements are selected for a signature. This causes an increase in the number of hash operations for verification.

A public key is calculated from the secret key using a homomorphic hash as follows.

$$\begin{aligned} pk_{<i,j>} &= HH(sk_{<i,j>}) = HH(sk_{<i-1,j>} + su_l^{(i-2)}) \\ &= HH(sk_{<i-1,j>}) + HH(su_l^{(i-2)}) = pk_{<i-1,j>} + pu_l^{(i-2)} \end{aligned} \tag{1}$$

where $2 \le i \le N$, $1 \le j \le t$, $0 \le l \le k$, and N is the length of the hash chain. A recipient can calculate all public keys sequentially from the initial public key ($pk_{<1,1>}, \ldots, pk_{<1,t>}$) and the disclosed secret key using equation (1).

For example, suppose that the current signature is ($sk_{<2,3>}$, $sk_{<2,4>}$, $sk_{<2,5>}$), referring to Figure 2. To verify this signature, a recipient first calculates ($su_3^{(0)}$, -, -) and verifies these values along with the hash chain. Then, s/he calculates ($pu_3^{(0)}$, -, -) and ($pk_{<2,3>}$, $pk_{<2,4>}$, $pk_{<2,5>}$) using equation (1). If the value for updating public key, *pu*, is a blank '-' the current public key is the initial public key. Now, the signature ($sk_{<2,3>}$, $sk_{<2,4>}$, $sk_{<2,5>}$) can be verified because the homomorphic hash values for these values are ($pk_{<2,3>}$, $pk_{<2,4>}$, $pk_{<2,5>}$).

The security of the proposed scheme depends on the *k* hash chain for updating the secret key. Detailed security analysis will be described in the full paper.

III. PERFORMANCE ANALYSIS

The proposed one-time message authentication scheme is constructed with cryptographic hash and homomorphic hash operations. We asymptotically analyze our scheme and the TV-HORS to compare their performances. The results are listed in Table 2. The proposed scheme is more efficient in terms of key generation and signature verification. According to our simple implementation, the computation time of a homomorphic hash function is about three times greater than that of a cryptographic hash. Nevertheless, it is expected that general computation costs of the proposed scheme will be lower in terms of key generation and signature verification, according to the analysis.

For more detailed and empirical analysis, we implement TV-HORS and the proposed scheme using a resource-constrained device, i.e., a ZigBee module, instead of a smart meter, which makes it difficult to manipulate the internal processes. The performance evaluation results will be described in the full paper.

Table 2. Performance Analysis Results. R – random number generation, A – addition, n – the length of a key pair, τ – the number of traversal in hash chain, β – average number of hash computation in TV-HORS, h – hash size, r – a size of a random number.

Scheme	Computation Cost			Comm. Overhead
	Key Gen.	Signing	Verifying	
TV-HORS [9]	$tR + tnH$	$(1 + \tau k)H$	$(1 + \beta)H$	kh
Proposed scheme	$tR + knH + tHH$	$(1 + \tau k)H + kA$	$(1 + k)H + kHH$	kr

IV. CONCLUSION

Message authentication is necessary for smart grid systems, especially for metering data related to charge. To authenticate metering data, a smart meter has to compute cryptographic values and signatures, but it is a resource-intensive operation. A smart meter has relatively low resources compared to an MDMS; hence, an efficient message authentication scheme is needed. We proposed an efficient one-time signature scheme for metering data authentication. The performance evaluation results show that our scheme is more efficient than TV-HORS, although the computation time for signing is longer.

REFERENCES

[1] Stephen McLaughlin, Dmitry Podkuiko, and Patrick McDaniel, "Energy theft in the advanced metering infrastructure," LNCS 6027, pp. 176-187, 2010.

[2] Matthew Carpenter, Travis Goodspeed, Bradley Singletary, Ed Skousid, and Joshua Wright, "Advanced Metering Infrastructure Attack Methodology", InGaurdians White-paper, January 5, 2009.

[3] IOActive Press Release, "IOActive Verifies Critical Flaws in Next Generation Energy Infrastructure, March 23, 2009.

[4] Mostafa M. Fouda, Zubair Md. Fadlullah, Nei Kato, Rongxing Lu, and Xuemin Shen, "A lightweight message authentication scheme for smart grid communications," IEEE Transaction on Smart Grid, vol.2, issue 4. pp. 675-685, 2011.

[5] Monageng Kgwadi and Thomas Kunz, "Securing RDS broadcast messages for smart grid applications," Technical Report SCE-09-06, Department of Systems and Computer Engineering Carleton University, Ottawa, Canada, 2009.

[6] Xiang Lu, Wenye Wang, and Jianfeng Ma, "Authentication and Integrity in the Smart Grid: An Empirical Study in Substation Automation Systems," International Journal of Distributed Sensor Networks, April 2012.

[7] Leonid Reyzin and Natan Reyzin, "Better than BiBa: Short One-Time Signatures with Fast Signing and Verifying," Proceedings of 7th Australian Conference on Information Security and Privacy, pp. 144-153, 2002.

[8] Q. Wang, H. Khurana, Y. Huang, and K. Nahrstedt, "Time valid one-time signature for time-critical multicast data authentication," IEEE INFOCOM 2009, pp. 1233-1241, 2009.

[9] Decio Luiz Gazzoni Filho and Paulo Sergio Licciardi Messeder Barreto, "Demonstrating data possession and uncheatable data transfer," IACR Cryptology ePrint Archive, p. 150, 2006.

Block-Based Detection Systems for Visual Artifact Location

O. Eerenberg, J. Kettenis and P.H.N. de With

***Abstract*--The core of many video coding standards is formed by the Discrete Cosine Transform (DCT) for de-correlating spatial video data. When quantizing DCT sub-bands, artifacts may appear such as mosquito noise and ringing. Spatial artifact reduction requires artifact location information, to control the filter process, thereby avoiding unnecessary blurring of artifact-free regions. This location information can be derived, either in the time- or frequency domain. As coding artifacts are most annoying in flat or low-frequency regions, the objective of the detector is to localize these artifact-sensitive locations. The detection accuracy, coverage and sensitivity differ between the two possible detection domains. The time-domain solution has a 10-97% location detection performance, whereas the frequency domain results in 70-100% detection performance.**

I. INTRODUCTION

Image and video communication have been benefiting from advances in compression techniques achieved in the last decades. Many of the popular compression techniques deploy a 2D-DCT to decorrelate a block-based spatial region prior to quantization. Although this is an efficient method for removing irrelevant information from a video signal, there is also a strong drawback. In order to achieve sufficient compression ratio, also relevant information in the form of high-frequency information is removed by means of quantization. Removal of high-frequency information not only results in lack of sharpness, but also introduces coding artifacts. Examples of typical coding artifacts are blockiness, ringing and mosquito noise. Modern digital televisions perform Temporal Noise Reduction (TNR), which not only removes Gaussian noise, but to a certain extent also reduces coding artifacts, provided that these artifacts are not static [1]. However, for the situation that the artifact is static, spatial filtering has to be applied to attenuate the disturbance [2]. Visual artifact-location information is crucial for controlling the spatial filtering strength, in order to avoid loss of detail.

This paper describes two block-based visual artifact-location detection systems, suitable to detect regions in a DTV decoded image, which have a high probability to be contaminated with coding artifacts. Although the detection criteria are equal, two experimental detectors each operating in a different domain, are compared regarding the detection performance.

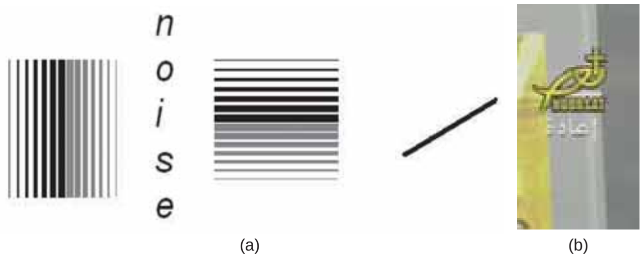


Fig. 1. Examples of mosquito and ringing artifacts due to MPEG-2 compression. (a) Image fragment with artifacts for Q=40. (b) Static image region containing logo with artifacts.

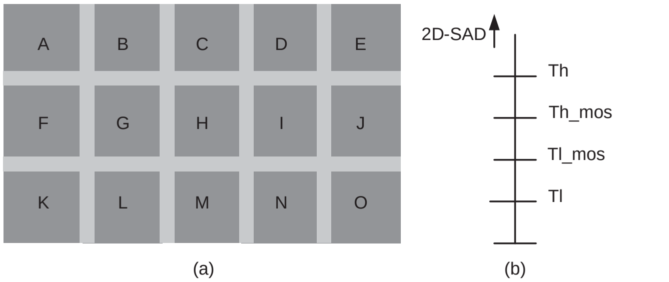


Fig. 2. Time-domain activity measurement. (a) Detection kernel deploying overlapped blocks. (b) Positioning of thresholds in SAD range.

II. DETECTION OF CODING ARTIFACT-PRONE LOCATIONS

Transform coding introduces artifacts, which are clearly noticeable around the transition between texture/edges and flat/low-frequency regions [3], see Fig. 1. This observation is a key feature for locating mosquito/ringing prone locations. The detection of such artifact-contaminated locations requires an *activity* measurement, revealing the presence of edges within a bounded region, followed by a spatial reasoning step, which results in a binary decision: *contaminated* versus *not-contaminated*. We investigate two different approaches.

A. Detection system in time domain

The activity metric in the time domain is based on a simple 2D high-pass filter, implemented as a 2D SAD according to

$$SAD = \sum_{y=j}^{M}\sum_{x=i}^{N}\left|P(x,y)-P(x+1,y)\right| + \sum_{y=j}^{M}\sum_{x=i}^{N}\left|P(x,y)-P(x,y+1)\right|. \quad (1)$$

For each pixel in the image, located at the centre of block H, the block-based metric is calculated using overlapped blocks, constructing a spatial kernel aperture, see Fig. 2(a). Basically, the spatial kernel aperture size depends on the size of block H, which contains the center pixel and therefore consists of an odd number of pixels, e.g. size 3x3 or 5x5 pixels. The surrounding blocks in vertical direction automatically obtain the same width, whereas the blocks in horizontal direction have equal height. The remaining blocks may have different sizes to create a 2D-filter, with a behavior other than a basic box-filter. The 2D-SAD value of each sub-region is compared, see Fig. 2(b), against a threshold *Th* and *Tl*, except sub-region *H*, which is compared against threshold *Th_mos* and *Tl_mos*. For the centre block *H*, the SAD classification is either `flat`, `texture` or `mosquito`, while for

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

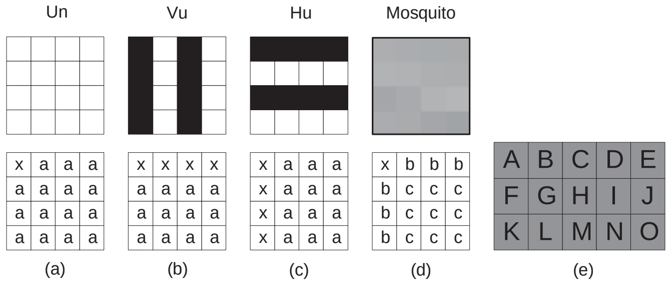


Fig. 3. Video features used for activity measurement in the transform domain. (a) Uniform area. (b) Vertical uniform. (c) Horizontal uniform. (d) Mosquito contaminated. (e) Transform-domain filter kernel.

the other blocks this is \`flat\` or \`texture\`. Using spatial reasoning, the results of the threshold operations are reduced to a binary signal, indicating if the center pixel of sub-block *H* is contaminated.

B. *Detection system in transform domain*

The activity detector in the transform domain is based on a 2D-DCT according to *Y=AXAT*. Matrix *X* has size 4x4 samples and *A* is a 4x4 transform matrix. After transformation, *Y* holds a set of 4x4 DCT coefficients describing the local video feature. The transform-domain activity filter-kernel consists of the same number of equally sized non-overlapped sub-blocks as the time-domain kernel, see Fig. 3(e).

In order to reduce the wide variety of energy distributions of each 4x4 region, the energy is matched with five video features. Figure 4 depicts the four video features, while the fifth video characteristic is \`texture\`, which applies if none of the four video features match. Prior to matching the supported video features, the 2D DCT sub-bands can be quantized to influence the video feature matching. The video matching process investigates the energy on locations indicated by a, see Fig. 3(a-c). For each video feature, the locations are squared, summed up and compared against a threshold *Tun* and *Tvh*, for the *uniform* and *horizontal/vertical* uniform video feature respectively. For the mosquito video feature, the sub-bands at location b and c are compared against a threshold *T*b and *Tc*. The final result is a spatial region which is classified by maximal five video features. The final classification is based on feature ranking, whereby the uniform feature has the highest position and remaining features follow the order of Fig. 3(a-d). On the basis of spatial reasoning, the transform-domain activity filter-kernel reduces this to a binary signal, indicating if the center block *H* is contaminated.

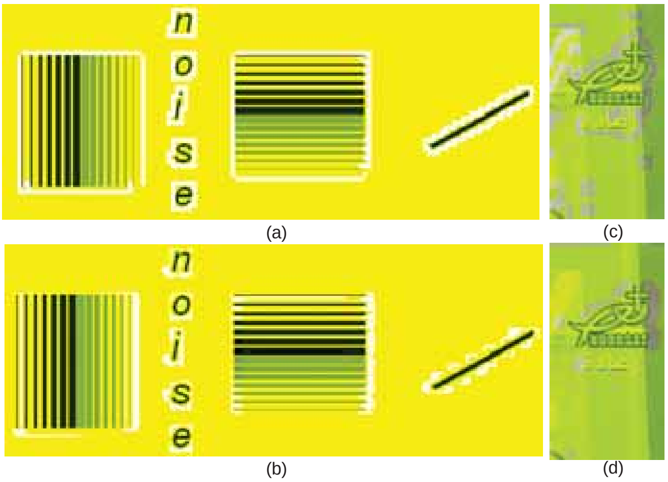


Fig. 4. Highlighted block-based artifact-location detection. (a) Transform-domain detection. (b) Time-domain detection. (c) Transform-domain detection. (d) Time-domain detection.

TABLE I

DETECTION PERFORMANCE OF VISUAL ARTIFACT LOCATIONS

	Detected Ringing 4x4 blocks (%)	Detected mosquito 4x4 blocks (%)	Detected Ringing pixels %	Detected mosquito pixels %
Vertical texture	100	70	87	19
Noise	0	99	0	50
Horizontal texture	78	70	10	97
Slanted edges	0	84	0	45

III. EXPERIMENTAL RESULTS

We have tested the detection systems for a broad range of TV images and specific test images for which the detection performance can be carefully evaluated. Due to space limitations, we present here some results on the basis of a few test images. For the test image in Fig. 1(a), the artifact locations are determined using a block grid of 4x4, using the definition that artifacts are most visible in the vicinity of the transition \`flat\` to \`texture\`, or visa versa. Hereby the pixels contained by the 4x4 blocks, form the reference pixels for validating the time domain-based detector. On the basis of this reference set, the performances of the two artifact-location systems are validated. The coverage results are depicted in Table 1 and visualized in Fig. 4. Figure 4(c-d) indicate the detection performance on a region containing a static logo.

IV. CONCLUSIONS

We have studied two block-based visual artifact-location detection systems, suitable to detect static regions in a DTV image. Visual artifact-location detection is successfully achieved, with a detection performance of 70-100% for the frequency domain and 10-97% for the time domain. The frequency-domain detector shows a higher selectivity and consistency, due to knowledge of the underlying video features. The time-domain detector locates on the average, the majority of the artifact-prone locations, but lacks selectivity.

REFERENCES

[1] C. Mantel, P. Ladret and T. Kunlin, "Temporal mosquito noise corrector", International Workshop Quality of Multimedia Experience, QoMEx 2009, pp. 29-31 July 2009.

[2] I. Kirenko, S. Ling and A. Nakonechny, "Quality Enhancement of Compressed Video Signals ", Proceedings of ICCE 2008, pp. 333-334 , 9-13 Jan. 2008.

[3] S.J.P. Westen, R.L. Lagendijk, and J. Biemond, “Adaptive spatial noise shaping for dct based image compression,” Acoustics, Speech, and Signal Processing, IEEE International Conference on, vol. 4, pp. 2124–2127, 7-10 May 1996.

LCU-Level Rate Control for Hierarchical Prediction Structure of HEVC

Dong-Il Park[1], Haechul Choi[2], Jin-soo Kim[2], Jin Soo Choi[3], and Jae-Gon Kim[1]
[1]School of Electronics, Telecom. & Computer Eng., Korea Aerospace University, Goyang, Gyeonggi-do, Korea
[2]Hanbat National University, Daejeon, Korea
[3]Electronics and Telecommunications Research Institute, Daejeon, Korea

Abstract-- **This paper presents a rate control scheme for the emerging HEVC which is currently being developed as a new video coding standard. The proposed scheme employs the largest coding unit of HEVC as a basic unit for the rate control. Moreover, bit allocation and quantization parameter decision schemes are introduced for the hierarchical prediction structure by taking into consideration the relative importance of each temporal layer and frame type.**

I. INTRODUCTION

The major purpose of the rate control is to regulate the bit stream according to the available bandwidth and a predefined buffer size. Recently, based on H.264/AVC [1], rate control methods have been intensively studied [2], [3]. On the other hand, a new video coding standard that is referred to as high efficiency video coding (HEVC) [4] has been developing. It is reported that HEVC can save about 33% in bit rate when compared with H.264/AVC High Profile [5]. Regarding the coding structure, H.264/AVC has a fixed-size basic block, 16×16 macroblock (MB), whereas HEVC defines the coding unit (CU) that supports various block sizes of 8×8 to 64×64 pixels. In conventional rate control methods of H.264/AVC, the MB is utilized as a basic unit (BU) for the rate control. This scheme implies that more accurate regulation of bits for HEVC may be possible if the CU structure is considered as the BU of the rate control.

This paper introduces a rate control scheme of HEVC that is realized on a frame basis and on the largest coding unit (LCU) basis. Moreover, the quadratic rate-distortion (R-D) model [3] is modified to exploit the characteristics of each temporal layer and frame type of the hierarchical prediction structure.

II. PROPOSED RATE CONTROL SCHEME

The proposed method consists of two main processing stages; the calculation of target bits and the quantization parameter (QP) decision according to the resulting target bits. Target bits are allocated to group of pictures (GOP), frame, and LCU in order.

In the frame-level rate control, target bits of a frame, $R_{frm,}$ is calculated by (1).

$$R_{frm} = (1-\alpha)\times\left\{\frac{W}{f}+\beta\times(B_t-B_c)\right\}+\alpha\times R_r\times\frac{w_k}{w_{sum}} \qquad (1)$$

where α and β are typically set to 0.75 and 0.5, respectively, as Liu *et al* [3], and W is the available channel bandwidth, and f represents the frame rate, and B_t is the target buffer occupancy, and B_c is the available buffer occupancy, and R_r is the remaining bits before encoding the current frame, and w_k denotes the weighting factor for frames with the k-th hierarchical level, and w_{sum} is the sum of weighting factors of the remaining frames. As in (1), by using the w_k, target bits are allocated to each frame while considering the relative importance of the hierarchical level. The next stage is to determine QP according to the resulting target bits, for which the quadratic R-D model is modified as follows:

$$\frac{R^{l,t}_{frm}}{MAD^{l,t}_{frm}} = \frac{X^{l,t}_1}{Q_{frm,l,t}} + \frac{X^{l,t}_2}{(Q_{frm,l,t})^2} \qquad (2)$$

where $MAD^{l,t}_{frm}$ is a mean absolute difference (MAD) of the frame of which the temporal level and the frame type are equal to l and t, respectively. It is linearly predicted by using the previously encoded frames with the same temporal level and frame type. $X^{l,t}_1$ and $X^{l,t}_2$ are the generated bits and distortion, respectively, of the previously encoded frames of which the temporal level and the frame type are equal to l and t, respectively. The $Q_{frm,l,t}$ derived by using (2) is used for the QP of the current frame.

In the LCU-level rate control, when given frame level remaining target bits, T_r, target bits for the p-th LCU, b_p, is calculated as follows:

$$b_p = T_r \times \frac{MAD^p_{LCU}}{MAD^p_{LCU} + N_{r,LCU} \times MAD^{mean}_{LCU}} \qquad (3)$$

where MAD^p_{LCU} is the MAD of the p-th LCU that is linearly predicted from the the $(p-1)$-th LCU, and $N_{r,LCU}$ is the number of the remaining LCUs within the current frame. MAD^{mean}_{LCU} is the average MAD of the 1st to the $(p-1)$-th LCUs and it is used as the predicted MAD for the remaining LCUs. As in (3), the number of target bits for the p-th LCU is specified as the ratio of the MAD of the current LCU to the total MAD of all the remaining LCUs. The QP for the p-th LCU, $Q_{LCU,p}$, is decided according to three conditions. First, for the first and the second LCUs within a frame, the frame level QP derived by (2) is directly used since the data are not accumulated sufficiently. Second, when the number of the remaining target

This research was supported by the MKE under the ITRC support program supervised by the NIPA (NIPA-2012-H0301-12-1006) and Basic Science Research Program through NRF funded by the MEST (No. 2012-0023109)

TABLE I
EXPERIMENTAL RESULTS OF FRAME-LEVEL RATE CONTROL

Sequence (Intra Period)	Target Bitrate [kbps]	Generated Bitrate [kbps]	Bitrate Diff. (m_E/σ_E)	PSNR(Y) [dB]
Kimono (24)	920	918.65	0.32/0.29	35.19
ParkScene (24)	2525	2521.48	0.47/0.58	34.26
BasketballDrive (48)	2970	2972.78	0.19/0.19	35.00
BQTerrace (64)	3000	2990.86	0.89/2.42	32.44
KBS2 (64)	5500	5501.09	0.41/1.76	35.46
MBC2 (64)	3000	3002.06	0.52/2.14	40.66
Average			0.47/1.23	35.50

TABLE II
EXPERIMENTAL RESULTS OF LCU-LEVEL RATE CONTROL

Sequence (Intra Period)	Target Bitrate [kbps]	Generated Bitrate [kbps]	Bitrate Diff. (m_E/σ_E)	PSNR(Y) [dB]
Kimono (24)	920	921.33	0.41/0.52	35.17
ParkScene (24)	2525	2527.72	0.40/0.48	34.16
BasketballDrive (48)	2970	2975.11	0.21/0.29	34.76
BQTerrace (64)	3000	3046.61	0.81/2.21	32.58
KBS2 (64)	5500	5500.65	0.23/0.31	35.43
MBC2 (64)	3000	2999.41	0.40/0.71	40.56
Average			0.41/0.75	35.44

bits becomes negative although there exist LCUs that are not encoded yet, the QP of the previous LCU plus 2 is assigned to the $Q_{LCU,p}$. In this case, to keep the smoothness of subjective quality, the adjusted QP is further bounded with the average QP for LCUs in the previous frame. When the first and the second conditions are not satisfied, most of LCUs generally belong to this case, the $Q_{LCU,p}$ is determined by using the quadratic R-D model in the similar way of the frame level QP decision scheme in (2), for which R_{frm} is replaced with b_p.

III. EXPERIMENTAL RESULTS

To evaluate the rate control performance for the hierarchical B picture structure, the proposed rate control scheme is implemented on the top of HM 4.0 [6], and the experimental conditions follows the random access (RA) configuration of the HEVC common test conditions [7].

Table I and Table II show the experimental results for the frame-level and the LCU-level rate control, respectively. In these tables, the bitrate difference between target bits and generated bits is represented as the mean, mE, and the standard deviation, σE. Fig. 1 is the generated bits every second to evaluate the bit fluctuation. As the experimental results, the generated bitrate is very close to the target bitrate. Note that the LCU-level rate control outperformed the frame-level rate control, which reveals that the proposed LCU-level rate control scheme is suitable for HEVC.

In addition, to evaluate the performance of the proposed method at a severe condition, we measured the occurrence percentage of the case where the number of target bits is forced to be zero. This case occurs when all of allocated bits for a GOP are exhausted before all frames of the GOP are not yet encoded due to an inaccurate decision of QP. As shown in Table III, the LCU-level rate control outperformed the frame-level rate control by on average 2 %.

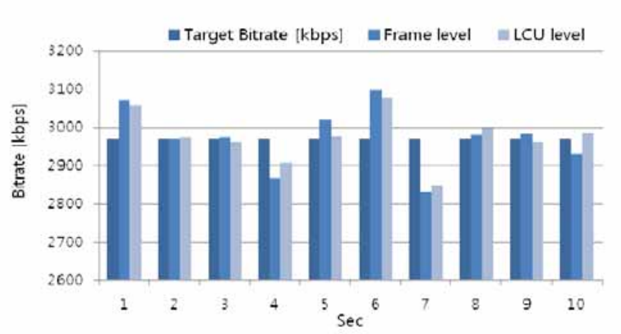


Fig. 1. Bitrate at each second for BasketballDrive (50 fps).

TABLE III
PERCENTAGE OF THE EXHAUSTED FRAME TARGET BITS

Sequence	Frame level[%]	LCU level[%]
Kimono	2.08	4.58
ParkScene	4.58	1.67
BasketballDrive	0	0
BQTerrace	18.50	11.83
KBS2	4.33	0
MBC2	4.17	3.83
Average	5.61	3.65

IV. CONCLUSIONS

In this paper, we present a rate control scheme for the emerging HEVC. In particular, to surely regulate bits for the hierarchical B picture structure, the temporal level and frame type are considered at both of the bit allocation and the QP decision processes. Moreover, the proposed method works at both of the frame-level and LCU-level, individually. The experimental results show that the generated bits are very close to the target bits in the RA configuration of the HEVC common test conditions.

In HEVC, the CU can be recursively partitioned to 4 CUs with smaller size, which is represented by the CU depth. As a further study, if bits are able to be controlled according to each depth of CU, the target bit rate would be achieved more efficiently.

REFERENCES

[1] ISO/IEC 14496-10: Information technology - Coding of audio-visual objects - Part 10: Advanced video coding, 2008.

[2] Z. Li, W. Gao, F. Pan and K. Pang, “Adaptive Basic Unit Layer Rate Control for JVT,” Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-G012, 7th meeting, Pattaya II, Thailand, Mar. 2003.

[3] Y. Liu, Z. G. Li and Y. C. Soh, “Rate Control of H.264/AVC Scalable Extension,” *Trans. Circuits Syst. Video Technol*., vol. 18, no. 1, pp. 116-121, Jan. 2008.

[4] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, and T. Wiegand, “High efficiency video coding (HEVC) text specification draft 8,” JCTVC-J1003, JCT-VC 10th meeting, Stockholm, July 2012.

[5] B. Li, G. J. Sullivan, and J. Xu, “Comparison of Compression Performance of HEVC Draft 6 with AVC High Profile,” JCTVC-I0409, JCT-VC 9th meeting, Geneva, Apr. 2012.

[6] K. McCann, B. Bross, S. Sekiguchi, and W.-J. Han, “HM4: high efficiency video coding (HEVC) test model 4 encoder description,” JCTVC-F802, JCT-VC 6th meeting, Torino, July 2011.

[7] F. Bossen, “Common test conditions and software reference configurations,” JCTVC-I1100, JCT-VC 9th meeting, Geneva, Apr. May 2012.

Energy-Delay Tradeoff Analysis and Enhancement in LTE Power-Saving Mechanisms

Wonjae Shin, *Member, IEEE,* Jung-Ryun Lee, *Member, IEEE*, Hyun-Ho Choi, *Member, IEEE*

***Abstract*— We propose a modified power-saving mechanism (PSM) that reversely applies the state transition of legacy LTE PSM by considering the attributes of network propagation delay. We analyze the PSM with respect to the energy consumption and the buffering delay, and characterize them as a simple energy-delay tradeoff (EDT) curve according to the operational parameters. The resulting EDT curves clearly show that the proposed PSM enhances the legacy PSM in various network environments and also guide to select an optimal parameter for maximizing the energy conservation while ensuring the quality of service.**

I. Introduction

Third-Generation Partnership Project (3GPP) Long-Term Evolution (LTE) wireless networks provide power-saving mechanism (PSM) called *discontinuous reception* (DRX) operation [1]. In viewpoint of energy consumption, the DRX operation can be divided into two operational states: *active* and *sleeping* states. The active-state user equipment (UE) keeps awake to receive packets with an activated transceiver. In contrast, the sleeping-state UE inactivates its transceiver and hence needs to periodically wake up to receive an indication message. Prior to the beginning of sleeping state, the active-state UE initiates an *inactivity timer* to monitor new packet arrivals. If no new packets arrive before the timer's expiration, the UE transits into sleeping state. Otherwise, the UE restarts the timer whenever receiving newly arrived packets. This is the basic operation of legacy PSM, part of which is shown in Fig. 1.

It has been recognized that there is a tradeoff between the energy conservation of a UE and the delay performance of the transmitted packets in the PSM. The tradeoff between the energy consumption and the downlink buffering delay has been studied [2], [3]. This tradeoff arises from the fact that the longer the UE stays in the sleeping state, the more power is saved, but an additional buffering delay occurs. Moreover, the off period, during which no packet is transmitted and so the UE can sleep, is usually modeled as exponential distribution because it is originated from user behaviors, such as reading and silencing [4].

In this paper, we consider a network propagation delay as another factor that induces the traffic off period. Considering the attributes of network propagation delay, we propose a modified PSM that reversely applies the state transition of legacy PSM in the LTE standard. We analyze the legacy and proposed PSMs with respect to the energy consumption and the buffering delay, and characterize them as a simple *energy-delay tradeoff* (EDT) curve according to the value of inactivity timer.

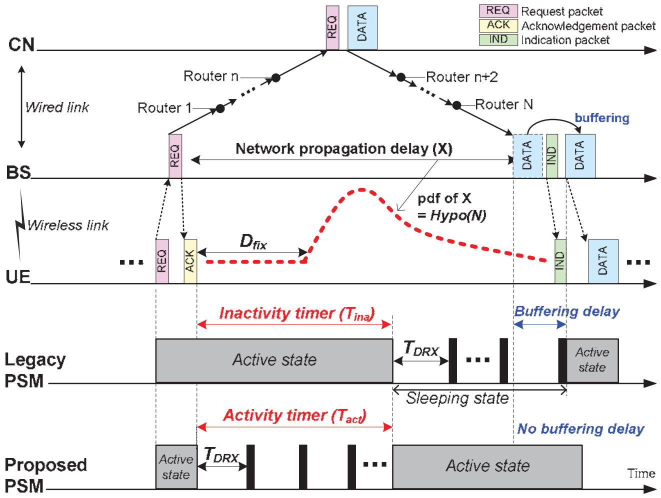


Fig. 1. Operations of legacy and proposed PSMs according to the traffic distribution with network propagation delay.

II. Network Propagation Delay Model

We consider request-response-based applications, such as web browsing and email synchronization. Fig. 1 illustrates an example of traffic arrivals of these applications. This traffic distribution involves the network propagation delay, which is defined as the time interval from the time when the base station (BS) sends the correspondence node (CN) a request packet to the time when it receives the data packet as a response. The network propagation delay is the sum of delays happening in each hop on the packet transmission path. Delay in each hop can be divided into minimal network delay and queueing delay. The minimal network delay consists of the propagation delay (5 μs/km), the transmission delay (which depends on the packet size and the link rate), and the route lookup delay. We denote this delay as D_{mnet}. Since D_{mnet} has a small variation compared to the entire propagation delay, its value is assumed to be fixed. On the other hand, the queueing delay in the n-th router, denoted as D^n_{que}, is variable and modeled as an exponential distribution based on empirical measurements [4], [5].

Let X be the random variable of network propagation delay. Then, X is expressed as

$$X = \sum_{n=1}^{N} (D_{mnet} + D^n_{que}) = \sum_{n=1}^{N} D^n_{que} + D_{fix} = Hypo\,(N) + D_{fix} \quad (1)$$

where D_{fix} is a fixed delay given by $N{\cdot}D_{mnet}$. Therefore, X follows a *hypoexponential* distribution composed of N exponential distributions with different rates in series. The

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0008565).

Wonjae Shin is with Signal & Systems Lab., Samsung Electronics Co. Ltd., Republic of Korea. Jung-Ryun Lee is with School of the Electrical Engineering, Chung-Ang Univ., Republic of Korea.

Hyun-Ho Choi is with Department of Electrical, Electronic, and Control Engineering, Hankyong National Univ., Republic of Korea, Corresponding author (e-mail: hhchoi@hknu.ac.kr).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

probability density function (pdf) of X is given by

$$f_X(t) = \begin{cases} \sum_{n=1}^{N} C_n \lambda_n e^{-\lambda_n (t - D_{fix})} & \text{if } t \geq D_{fix} \\ 0 & \text{if } t < D_{fix} \end{cases} \qquad (2)$$

where λ_n is a rate parameter of exponential distribution $D^n{}_{que}$ and C_n is a constant given by $\prod_{m \neq n} \lambda_m / (\lambda_m - \lambda_n)$.

III. Proposed Power-Saving Mechanism

As shown in Fig. 1, the network propagation delay conforms to a heavy-tailed asymmetric distribution (i.e., hypoexponential distribution) with a certain constant delay. Thus, the proposed PSM is motivated by the fact that there is no packet arrival during a certain time period after the BS sends the request packet to the CN. So, it can simply change the PSM operation from the conventional active-to-sleeping state transition into the sleeping-to-active state transition, as shown in Fig. 1. To do this, the proposed PSM additionally makes the UE go into the sleeping state immediately when the UE receives the acknowledgement packet for its request. Compared with the inactivity timer (T_{ina}), we similarly introduce *activity timer* (T_{act}) to control the period of sleeping state. Note that the standard LTE PSM is not flexible enough to support this proposed technique.

IV. Energy-Delay Tradeoff Analysis

First, the probabilities that the data packet arrives in the active state and sleeping state are respectively calculated as

$$P_{act} = \int_0^{T_{ina}} f_X(t)dt , \quad P_{slp} = \int_{T_{ina}}^{\infty} f_X(t)dt = 1 - P_{act} . \qquad (3)$$

Let T_{DRX} be the length of DRX cycle. Since the data packet arrives uniform-randomly from a viewpoint of the given state's period, the average buffering delay is a half of DRX length when the DRX length of each cycle is constant. Therefore, the average buffering delay of legacy PSM is given by

$$\overline{D}_{legacy} = P_{slp} T_{DRX} / 2. \qquad (4)$$

Let E_{act} and E_{slp} be the costs of energy consumption when the UE stays in the active state and sleeping state, respectively. By calculating the conditional expectation of network propagation delay given that the data packet arrives in the sleeping state ($\overline{T}_{arr}$), the average energy consumption of legacy PSM is obtained as

$$\overline{T}_{arr} = E[t \mid t > T_{ina}] = \int_{T_{ina}}^{\infty} \frac{t f_X(t)}{P\{t > T_{ina}\}} dt = \int_{T_{ina}}^{\infty} \frac{t f_X(t)}{P_{slp}} dt, \qquad (5)$$

$$\overline{E}_{legacy} = P_{act} E_{act} + P_{slp} \left\{ \frac{E_{act} T_{ina} + E_{slp} (\overline{T}_{arr} - T_{ina})}{\overline{T}_{arr}} \right\}. \qquad (6)$$

Since the proposed PSM is the opposite of the legacy PSM, its buffering delay and energy consumption can be obtained by changing P_{slp} into P_{act} in (4) and by exchanging the cost value between E_{act} and E_{slp} in (6), respectively.

V. Results and Discussions

We use T_{DRX}=1.28 s, E_{act}=20 and E_{slp}=1 as relative energy costs, D_{mnet}=10 ms, and λ_n=(mean of 100 ms, variance of 1 ms) [1], [4]. We vary the number of routers (N) within 1~8 and the timer values within 0~5 s.

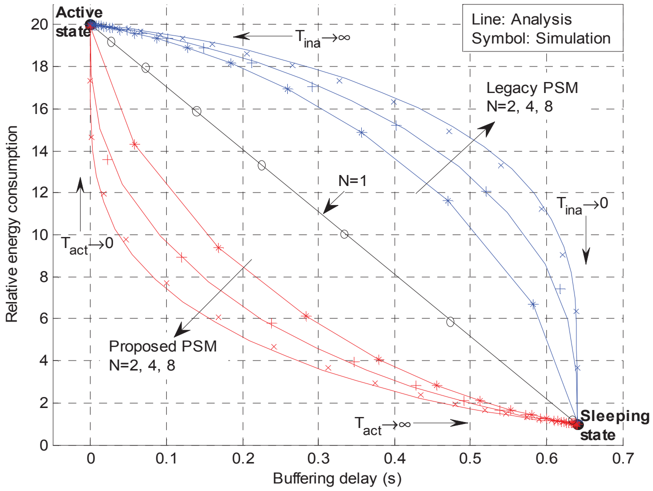


Fig. 2. Energy-delay tradeoff curves.

Fig. 2 plots the energy-delay tradeoff (EDT) curves. Every point on each curve can be achieved by adjusting the inactivity or activity timer. Therefore, based on this EDT curve, we can select an optimal timer value that minimizes the energy consumption while satisfying the requirement of buffering delay. An inner curve shows better tradeoff performances than outer one, i.e., both energy consumption and buffering delay become smaller at the same time. So, the proposed PSM significantly outperforms the legacy PSM from the perspective of EDT. As the number of routers (N) increases, the EDT of proposed PSM becomes improved, but the EDT of legacy PSM becomes degraded. This is because the pdf of network propagation delay moves to the right and so it gives more opportunity to sleep before awake as N increases. Notably, when N=1 (i.e., the network propagation delay just follows the exponential distribution), the tradeoff curve becomes a linear line connecting the energy-delay costs of two states, regardless of the PSM type.

VI. Conclusions

The proposed PSM modifies the legacy LTE PSM by considering the attribute of network propagation delay. The derived EDT curves clearly show that the proposed PSM outperforms the legacy PSM in various network environments and also facilitate to select an optimal parameter in order to minimize the energy consumption of mobile CE device while guaranteeing its quality of service (QoS).

References

[1] C. S. Bontu and Ed Illidge, "DRX Mechanism for Power Saving in LTE," *IEEE Communications Magazine,* pp. 48-55, June 2009.

[2] D. Nga, *et al*, "Delay-guaranteed Energy Saving Algorithm for the Delay-sensitive Applications in IEEE 802.16e Systems," *IEEE Trans. on Consumer Electronics*, vol. 53, no. 4, pp. 1339-1347, Nov. 2007.

[3] J.-R. Lee and D.-H. Cho, "Performance Evaluation of Energy Saving Mechanism Based on Probabilistic Sleep Interval Decision in IEEE 802.16e," *IEEE Trans. on Veh. Tech.*, vol. 56, no. 4, pp. 1773-1780, July 2007.

[4] *IEEE 802.16m-08/004r5*, "IEEE 802.16m Evaluation Methodology Document (EMD)", 2009-01-15.

[5] T. Yensen, *et al*, "HMM delay prediction technique for VoIP," *IEEE Trans. on Multimedia*, vol. 5, no. 3, pp. 444-457, Sep. 2003.

Background Scene Classification Robust to the Influence of Human Regions

Ryota Mase, Ryoma Oami, and Toshiyuki Nomura
Information & Media Processing Labs., NEC Corporation
1753, Shimonumabe, Nakahara-Ku, Kawasaki, Kanagawa 211-8666, Japan.

***Abstract*-- We propose a background scene classification method robust to the influence of human regions. Conventional methods classify scene of an image by using image features extracted from entire region in the image. Therefore, in these methods, the influence of the human region such as color of the skin and the clothes reduces classification accuracy of the background scene. Our method classifies background scene of an image by using image features extracted from only background region except detected human regions. The experimental results show that the proposed method improves average of the rate at the balance point between recall rate and precision rate in almost all background scenes compared to the conventional method.**

I. INTRODUCTION

With the spread of digital cameras and camera-equipped mobile-phones, an automatic image quality enhancement technology is highly demanded for simple and convenient photography. Especially, the image quality enhancement method based on scene category classification is widely used in most cameras [1]. It automatically analyzes the scene of an image, and improves the image quality by using multiple-image-processing functions with correction parameters that are adaptively controlled according to scene of an image. Snapshots are frequently taken with one of various backgrounds such as “night scene”, “flower” and “landscape”. Therefore, it is indispensable to detect human regions and correctly classify the background scene of an image for quality enhancement of both human and background regions based on optimal correction parameters for each region.

Scene classification methods have been proposed so far [2]-[4]. By combining the conventional method with human region detection, human regions can be detected and background scene of an image can be classified. However, in the conventional scene classification methods, the influence of the human region such as color of the skin and the clothes reduces classification accuracy of the background scene, since they classify scene of an image by using image features extracted from entire region in the image.

In this paper, we propose a background scene classification method robust to the influence of human regions. The proposed method classifies background scene of an image by using image features extracted from only background region except detected human regions.

II. CONVENTIONAL METHOD

A conventional method [1] classifies background scene of an image into a predefined background scene category based on posterior probability. Posterior probability for each background scene category is calculated as follows:

$$p(\omega_i \mid \mathbf{x}) \propto p(\mathbf{x} \mid \omega_i) p(\omega_i)$$

where ω_i is a parameter of Gaussian Mixture Model (GMM) for background scene categories i, and $\mathbf{x}$ is feature vector compressed from image feature vectors by a linear discriminant analysis (LDA). The image feature vectors are extracted from an image as follows;

Color Layout: Color layout describes spatial distribution of color in an image. After a reduced image is generated from an image, this reduced image is transformed by DCT and a few low-frequency DCT coefficients are selected as color layout.

HSV Histogram: HSV histogram describes color distribution in an image. This feature is generated based on distribution of quantized pixel colors in HSV color space.

Edge Histogram: Edge histogram is the feature that expresses the local edge distribution in an image. After an image is divided into several blocks, this feature is generated based on distribution of some types of edge extracted from each block. It is an integration of histogram of all blocks.

LDA based GMM that is one of the good discrimination methods requires to be designed in advance by using many images [1]. However, there is a wide range of variations in location and size of human regions. Therefore, it is difficult to design appropriate model for them. As a result, in this method, the influence of the human regions such as color of the skin and the clothes reduces classification accuracy of the background scene of a snapshot.

III. PROPOSED METHOD

This paper proposes an image feature extraction method for background scene classification robust to the influence of human region. It first detects human regions in an image and then extracts the image feature vectors from only background region except detected human regions.

A. Human region detection

The proposed method defines human region as a combination of face region and body region. The face regions in an image are detected by a face recognition technology [5]. Then, the body regions in the image are detected for each face region based on general relation of size and location of human body to those of human face.

B. Feature extraction from region except human region

In the proposed method, the image feature vectors are

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

extracted from the region except the human regions detected by the process described in III-A. Each feature extraction is described below.

Color Layout: Color layout is generated by transforming the reduced image with size 8x8 on frequency domain and selecting 30 low-frequency DCT coefficients. Since extraction of this feature needs all pixels in an image, pixels of the human regions need to be regenerated by interpolation from the surrounding pixels. In many natural images, spatial distribution of color tends to be horizontally-correlated. Therefore, the pixels are first linearly interpolated in a horizontal direction, and then they are linearly interpolated in a vertical direction.

HSV Histogram: HSV histogram is generated based on distribution of pixel colors in the region except human regions in an image. The pixel colors are quantized to 128 bins in HSV color space.

Edge Histogram: After an image is divided into 4x4 blocks, edge histogram is generated as an integration of distribution of 5 types of edge extracted from each block. This results in a histogram of 80 dimensions. However, it is difficult to generate the histograms for blocks overlapping with the human regions. This is because there is poor correlation between histogram of a block and that of the contiguous blocks. Therefore, all bins of the histogram generated from the human regions are set to zero.

IV. EXPERIMENT

The proposed algorithm has been implemented and the performance was evaluated on a personal computer with Intel Core2 CPU (2.4 GHz) and memory (1 GB). Evaluation was carried out through experiments with 3092 images for testing, which include the images with 12 background scenes and the images with others. In these images, at least one person is shown up. And in this experiment, we used 496 images for training with 12 background scenes, in which no person is shown up. Table I shows the background scenes of images for training and testing, and the number of images for each background scene. Background scene classification accuracy of test images was evaluated with the proposed method and a conventional method [1] by recall and precision rates derived from threshold processing of posterior probability for each background scene category.

Table I: Background scenes of images for training and testing, and the number of images for each scene.

Background Scene	The number of images for training	The number of images for testing
Flower	52	212
Scenery	74	412
Autumn	38	232
Cherry Blossom	20	228
Snow	40	196
Sunset	33	240
Dish	54	200
Yellow-tinged	37	192
Night Cherry Blossom	35	240
Night	35	200
Firework	30	248
Illumination	48	164
Others		328
Sum	496	3092

Table II shows the rate at the balance point between recall rate and precision rate for each background scene category in the proposed method. In Table II, value in brackets shows the difference from the rate at the balance point between recall rate and precision rate for each background scene category in the conventional method. These results show that the proposed method improves classification accuracy for 8 background scenes out of 12 background scenes, and average of the rate at the balance point between recall rate and precision rate in all background scenes to about 64 percent from about 60 percent by the conventional method.

Table II: The rate at the balance point between recall rate and precision rate.

Background Scene	Accuracy	Background Scene	Accuracy
Flower	0.58 (+0.15)	Dish	0.62 (0.00)
Scenery	0.70 (0.00)	Yellow-tinged	0.85 (+0.04)
Autumn	0.81 (+0.04)	Night Cherry Blossom	0.47 (+0.02)
Cherry Blossom	0.61 (+0.06)	Night	0.86 (+0.04)
Snow	0.47 (+0.13)	Firework	0.44 (0.00)
Sunset	0.72 (+0.01)	Illumination	0.50 (-0.03)

V. CONCLUSION

In this paper, we have proposed a background scene classification method robust to the influence of human region. Conventional methods classify scene of an image by using image features extracted from entire region in the image. Therefore, in these methods, the influence of the human region such as color of the skin and the clothes reduces classification accuracy of the background scene. Our method classifies background scene of an image by using image features extracted from only background region except detected human regions. The proposed method has improved average of the rate at the balance point between recall rate and precision rate in almost all background scenes compared to the conventional method.

REFERENCES

[1] M. Tsukada et al., “Image Quality Enhancement Method based on Scene Category Classification and Its Evaluation”, Proc. of ICCE, 4.4-4, 2009, pp.1-2.

[2] A. Vailaya et al., “Image classification for content-based indexing”, IEEE Trans. on Image Processing, vol.10, no.1, 2001, pp.117-129.

[3] J. Shen et al., “Semantic-Sensitive Classification for Large Image Libraries”, Int. Multimedia Modeling Conf., 2005, pp.340-345.

[4] Wei Jiang et al., “Effective Semantic Classification of Consumer Events for Automatic Content Management”, ACM Multimedia Workshop on Social Media, Oct, 2009, pp.35-42.

[5] H. Imaoka et al., “NEC's Face Recognition Technology and Its Applications”, NEC TECH. J., Vol.5, No.3, 2010, pp.28-33.

Image Unsteadiness Correction in Archive Film Scanners

Kirill GUSEV

***Abstract*—This paper presents a method of image unsteadiness correction of digitized cinema materials. The method consists of two steps: correcting frame unsteadiness in a film scanner using sprocket hole images and correcting the unsteadiness of an image itself. Correction algorithms are based on calculating the cross correlation function of current image and the first image of the sequence, thus obtaining the value of the offset.**

I. INTRODUCTION

As traffic capacity of end-user internet channels grows online video hosting becomes more and more popular. Video hosting services give people an opportunity to get acquainted to cultural heritage of cinematography by hosting films of the past which otherwise would be known only to a small group of experts. One of the problems of digital restoration process is film frame unsteadiness. This sort of damage worsens the emotional perception of film by the viewer and lowers the efficiency of image compression algorithms.

Old film materials have high shrinkage value, often varying from part to part, so they need to be digitized on a pinless film scanner. To control the accuracy of frame positioning in such a scanner we propose the algorithm based on calculation of cross correlation function between the sprocket hole image of the current frame and the first frame of the scanned sequence. Offset of the cross correlation function maximum gives the offset of the current image against the first image of the sequence.

Image unsteadiness is generated not only in the film scanner but also in the preceding filmmaking process: in the film camera, and on the different stages of printing process. Due to this unsteadiness the position of the image against the sprocket hole is not accurate from frame to frame. So the second step of correction is based on the image itself. Again, we use cross correlation algorithm. But as the image may have moving objects or intentional camera panning, the algorithm must be revised.

II. SPROCKET HOLE IMAGES STABILIZATION

Archive film materials may be severely degraded and usually have high shrinkage value. When digitizing such materials film scanner with pinless film transportation system is the only choice. To provide precise image positioning we propose using digital processing of sprocket hole images.

To utilize this algorithm film scanner must have optical system and image sensor capable of capturing not only film frame but also corresponding sprocket holes.

We define region of interest containing image of one of the sprocket holes to reduce the calculation load. As the sprocket hole shape is clearly defined in standards [1][2] we can generate a binary pattern of a sprocket hole image, either BH or KS type depending on the type of film we are scanning. The size of the pattern is defined by a known scale factor of film scanner optical system. The scale factor also assumes the maximum expected shrinkage value, so that the pattern would never be larger than the sprocket hole image of a scanned film.

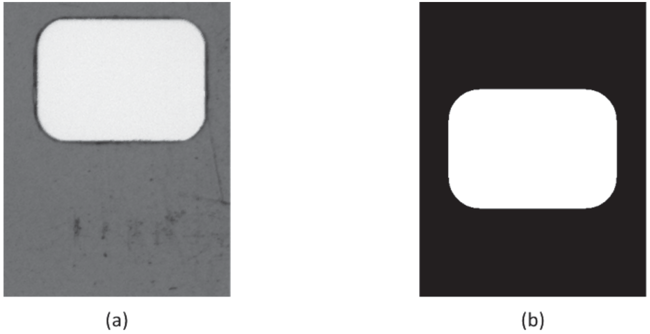


Fig. 1. Sprocket hole image (a) and its corresponding generated pattern (b).

Color image is transformed into a grayscale one. Cross correlation function between the pattern and the sprocket hole image is then calculated. Computation is realized in the frequency domain. First we calculate the 2D fast Fourier transform (FFT) of the pattern and the image. Then we multiply them. Multiplication result is brought back into spatial domain using inverse 2D FFT.

We find the maximum of the cross correlation function. Its 2D coordinates correspond to the value of the shift between the pattern and the image. We assume the offset of the first sprocket hole to be the anchor point and shift every next image so that the position of current sprocket hole on the image matched the position of the sprocket hole on the first image of the sequence.

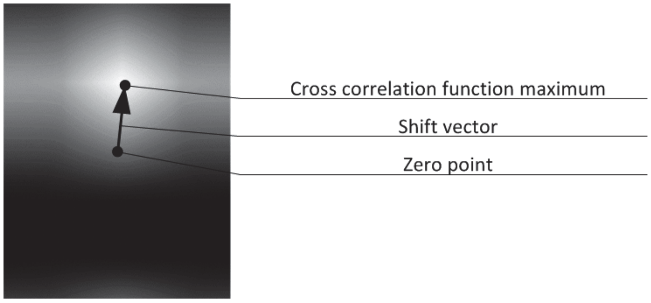
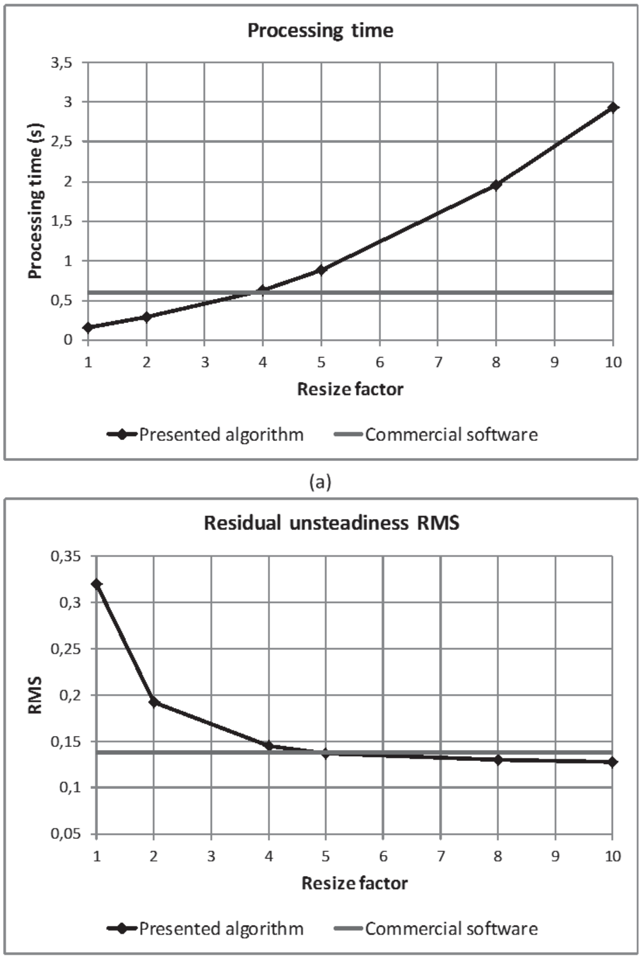


Fig. 2. 2D cross correlation function of the image and the pattern of Fig. 1.

We compared the performance of our method to that of one of the commercial software systems for compositing. Such software usually has built-in motion tracking functions for

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

stabilization and motion matching. To increase the accuracy of our algorithm we applied bicubic resizing to the image before correction. We used same commercial software as an instrument to measure the residual unsteadiness. Though this way of measurement is not strict it can give a comparison of two algorithms in the same conditions. We processed the sequence of 1000 frames scanned together with sprocket hole area with total resolution of 2592×1944 with scale providing the image size of 2K (2048×1536 for style C133 frame [3]). Fig. 3(a) shows the comparison of processing time while Fig. 3(b) shows the comparison of residual unsteadiness.



(a)

(b)

Fig. 3. Comparison of performance of presented algorithm and commercial software: processing time (a) and residual vertical unsteadiness root mean square (RMS).

Curves on Fig. 3 prove that presented algorithm with four times bicubic resize has the same processing time and accuracy as commercial software. If we further increase the resize factor the processing time grows rapidly while the accuracy stays at almost the same level.

The advantage of proposed method is its ability to process images as they come from the film scanner whereas stabilization options of compositing software need the whole sequence at once and work in two stages, first calculating the track and then shifting and rendering the output image.

III. IMAGE UNSTEADINESS CORRECTION

Unsteadiness correction using sprocket hole images solves the task of accurate frame positioning in the scanning process. But if there are inaccuracies in frame positioning against sprocket holes due to unsteadiness in the film camera or in the printing processes from negative to positive we need to correct the unsteadiness using the image itself.

The image may contain moving objects or intentional camera panning. To overcome these difficulties we need to add some modifications to the algorithm. We use the first image of the sequence as a pattern and compute cross correlation function between it and the current image. But moving objects generate their own peaks on the cross correlation function and blur the peak corresponding to the global motion thus lowering the accuracy of correction (Fig. 4). Normalizing of modulus of multiplication of the two FFTs improves the accuracy (Fig. 5).

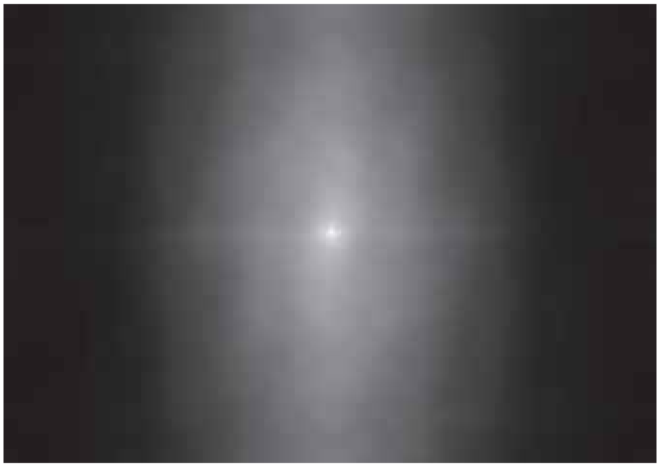


Fig. 4. 2D cross correlation function of the two consequent frames.

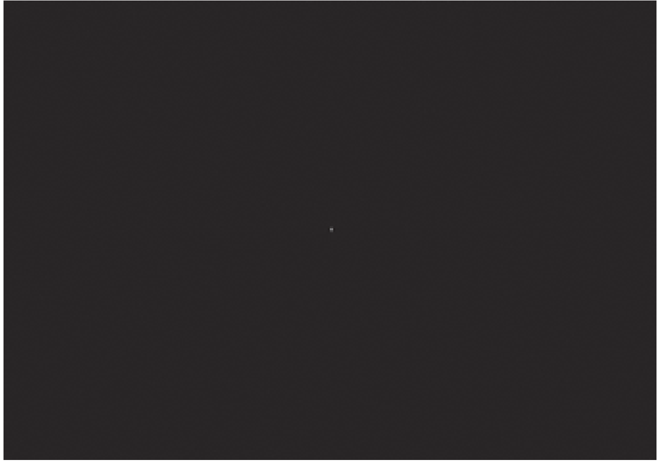


Fig. 5. Same 2D cross correlation function from Fig. 4, but with normalization.

Another way to improve the sensitivity of the method is to emphasize the high frequencies of the image and get rid of noise on the even regions of the image. To achieve this we propose to use edge detecting algorithms, such as Sobel filtering, as preprocessing (Fig. 6).

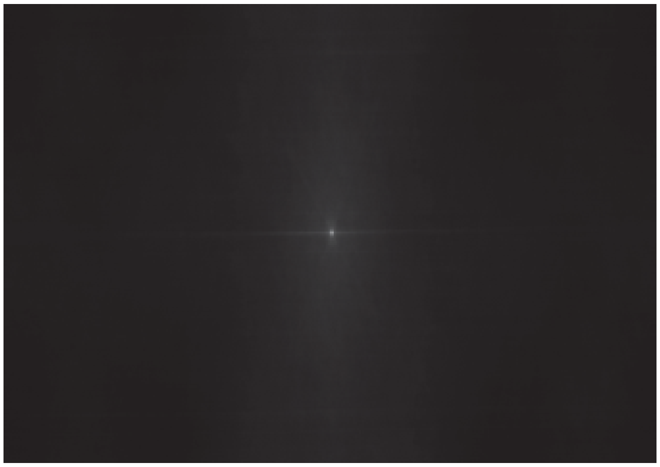


Fig. 6. 2D cross correlation function of the two consequent frames with Sobel pre-filtering.

At last, to eliminate the impact of the moving objects we may limit the region of the image in such a way, that the region of interest includes only stationary part of the image (usually its background).

The results of the algorithm modifications are shown in the Table 1.

TABLE I
COMPARISON OF ALGORITHMS MODIFICATIONS

Modification	Residual unsteadiness RMS
Basic	0.5996
With normalization	0.3173
With Sobel pre-filtering	0.1938
With region-of-interest	0.2170

Though the results of the method on stationary frames are encouraging, we still need to work on filtering of the frame coordinates data to distinguish high frequency unsteadiness from smooth intentional camera movement.

IV. CONCLUSIONS

The proposed combined method of image unsteadiness correction provides correction of both frame unsteadiness in a pinless archive film scanner and unsteadiness of the frame against the sprocket holes generated in the preceding filmmaking process.

The accuracy of the first stage (sprocket hole based correction) is as good as that of the commercial software for motion tracking. RMS value of residual unsteadiness is less than 0.15 pixels.

Resize factor of bicubic interpolation was found providing optimum between processing time and accuracy of correction.

The method corrects the images as they come from the film scanner before they are saved in the files in contrast to the motion tracking software, which needs the whole sequence to be on the hard drive before processing. This advantage of the proposed method results in time saving of the whole process.

The work on the second stage of the method (allowing unsteadiness correction by the image itself) is not yet finished: though showing good results on stationary frames it needs to be adapted to frames with panning.

REFERENCES

[1] SMPTE 93–2005 SMPTE STANDARD for Motion-Picture Film (35-mm) – Perforated BH.

[2] SMPTE 139-2003 SMPTE STANDARD for Motion-Picture Film (35-mm) – Perforated KS.

[3] SMPTE 59-1998 SMPTE STANDARD for Motion-Picture Film (35-mm) – Camera Aperture Images and Usage.

On an Implementation of HEVC Video Decoders with DSP Technology

F. Pescador, *Member, IEEE*, M.J. Garrido, E. Juarez, *Member, IEEE*, C. Sanz, *Member, IEEE*

***Abstract*—High Efficiency Video Coder (HEVC) will become a new MPEG International Standard by the end of 2012. HEVC is targeted to provide the same quality as H.264 at about a half of the bit-rate and will replace soon to its predecessor in multimedia consumer applications. In this paper, a preliminary implementation of an HEVC video decoder based on a DSP is presented and compared with a formerly developed H.264 DSP-based decoder.**

I. Introduction

As it is well known, the video decoder plays a central role in the consumer multimedia terminals. In the last years, the introduction of HDTV and the new 3DTV formats have increased the need for more efficient video compression standards. In this scenario, the ISO MPEG group has been working on a new standard, High Efficiency Video Coder [1] (HEVC), since 2010. The final standard approval is scheduled by the end of 2012 and it is expected that it will replace H.264 in Set-Top Boxes and DTV receivers in a few years.

Nowadays, Digital Signal Processor (DSP) technology [2] allows the implementation of very flexible video decoders at a relative low cost. In the last years, we have developed optimization techniques to implement MPEG-2, H.264 and H.264/SVC video decoders based on DSP technology with excellent results [3][4]. In this paper, a preliminary implementation of an HEVC video decoder based on a DSP is presented and compared with a former H.264 implementation based on the same DSP and in the same optimization status. Section I includes a short HEVC reference. In section II the DSP architecture is outlined. In section III, the HEVC decoder implementation is explained and its performance results are compared with those of an H.264 decoder implemented with the same DSP. Finally, section IV concludes the paper.

II. HEVC

HEVC standard is based on the same motion-compensated hybrid coding than their predecessors, from H.261 to H.264. The new standard has not a revolutionary design; instead, it has a lot of small improvements that, when put together, conduct to a considerable bit-rate reduction. The tests performed during the standardization process show that HEVC may compress until half the bit-rate of H.264 with the same quality [5], at the expense of a higher complexity.

The main differences among HEVC and its predecessor H.264 can be summarized as follows [5]:

- The Macroblock structure is replaced with a more flexible one, based on coding units (CUs). The CUs may support block sizes up to 64x64 pels.
- The shape of the prediction units (PUs) may be asymmetrical (i.e., two rectangles of different sizes).
- The transform units (TUs) may be up to 32x32 pels.
- Up to 33 intra prediction modes.
- Advanced skip modes and motion vector prediction.
- A new Adaptive Loop Filter (ALF).
- A Sample Adaptive Offset (SAO) is applied to the reconstruction signal after the Deblocking Filter.
- Tools oriented to parallel processing (WPP).

Otherwise, HEVC has the same CAVLC/CABAC entropy coding schema as well as the same Deblocking Filter (DF).

III. DSP architecture

The implementation of the HEVC decoder has been carried out with the same DSP [6] used in previous works [3][4] in order to ease the comparison with the H264 decoder. In Fig. 1, a simplified block diagram of the DSP is shown. The main processor is a fixed-point VLIW core with two levels of memory (L1 and L2) and an internal DMA (IDMA). L1 memory is splitted between L1P for program (32 KB) and L1D for data (80 KB). L2 memory (128 KB) can be used for data or program. All internal memory levels can be divided between general purpose memory and cache memory. A switched central resource interconnects the core with a set of standard peripherals and a video processing subsystem with a video capture (VPFE) and a video display (VPBE) processor.

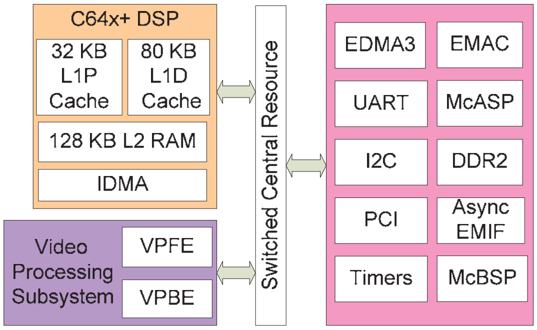


Fig. 1. Internal architecture of the DSP.

IV. The DSP-based HEVC decoder implementation

Now, the implementation and test of the HEVC decoder are explained and the results are compared with those of a former implementation of an H.264 decoder based on the same DSP.

[*]PccMuTe: Power Consumption Control in Multimedia Terminal is a research project supported by the Spanish Ministry of Science and Technology under grant TEC2009-14672-C02-01.

F. Pescador, E. Juarez, C. Sanz and M. J. Garrido are with the Electronic and Microelectronic Design Group (GDEM) at the *Universidad Politécnica de Madrid*, Spain. ({pescador, ejuarez, cesar, matias}@sec.upm.es).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

A. *Implementation*

The HEVC HM5.0 reference software [7] has been migrated to the DSP. Only basic optimization techniques have been used. Some changes have been done in the code of the decoder to obtain a DSP-based functional version:

- Integration of the library *string*.
- Redefinition of the function find included in the class *TComIterator* to adapt the input parameters.
- Redefinition of the function used to calculate the absolute value and inclusion of the *math* library.
- Redefinition of some C++ classes to adapt the member variables unsupported by C++ compiler.
- Redefinition of the *bool* type using a namespace.
- Removal of the functions used to measure the performance using the timers of the PC.
- Development of the function *strdup* unavailable in the DSP development environment.

The RTOS has been configured with the following features:

- Integration of the decoder in an OS task and definition of the stack (16 MB) and heap (24 MB) of this task.
- Configuration of the size of the cache memories to 32 KB (L1P and L1P) and 128 KB (L2).
- Configuration of the compiler output format to ELF, needed to support some features of the C++ code.
- Inclusion of internal timers to measure the performance.

B. *Testbench of the HEVC DSP-based decoder*

A development board [8] based on the DSP with a 594 MHz system clock, the CCS V5.1 [9] framework and a DSP emulator have been used to measure the decoder performance. The decoding time of each picture has been measured in system clock cycles by using DSP internal timers. As can be seen in Fig. 2, a SYS/BIOS [10] task has been implemented with a File Processing process to read the input stream from a file, a process with the migrated decoder and a File Processing process to store the decoded frames on a file.

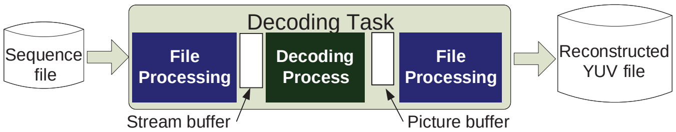


Fig. 2. Testbench used to measure the decoder performance

Four input streams have been generated by encoding the *Akiyo* CIF resolution sequence with the HM 5.0 encoder. The following settings, recommended in [11] for comparing H.264 to HEVC, have been used: GOP 8, ALF, SAO, RDOQ and QPs 26, 32, 37 and 44. In the first row of Table I, the number of frames per second (fps) processed by the DSP-based HEVC decoder are shown.

C. *Comparison with the H.264 decoder*

The same testbench have been used to measure the performance of an H.264 decoder based on the FFMPEG library [12] that was migrated to the same DSP in a previous work [3]. In this case, the *Akiyo* sequence has been encoded with the following settings: GOP 8, CABAC and QPs 26, 32, 37 and 44. In Table I, in the 2nd and 3rd rows, the number of *fps* processed by the H.264 decoder both, after migration and after the full set of optimization was applied, are shown.

As can be shown, the optimization process may speed the decoder around 3.5 times. With this data, we can forecast that, if the full set of optimizations were implemented on the HEVC migrated code, more than 50 CIF *fps* could be decoded. This is a conservative prediction because the FFMPEG library for H.264 is more optimized than the HEVC reference code.

TABLE I
PERFORMANCE OF THE HEVC & H.264 DECODERS (FPS)

Decoder	QP26	QP32	QP36	QP44	GOP16	One I	SAO
HEVC (migrated)	14.4	15.4	16.3	17.1	14.7	16.3	16.6
H.264 (migrated)	53.9	57.4	58.9	63.1	59.3	61.4	N/A
H.264 (optimized)	177.4	206.5	210.0	231.7	208.2	213.8	N/A

V. CONCLUSION

A preliminary implementation of a HEVC video decoder based on the HM5.0 and using a DSP has been presented. The performance of the decoder has been measured and compared with the performance of a H.264 decoder in both, the same stage of optimization and with a full set of optimizations. With the results of this comparison, we forecast that the HEVC decoder could decode more than 50 CIF fps if the full set of optimizations were applied.

Our future work will be focused on implementing the HEVC decoder with a multi-core processor [13] in order to decode higher resolutions.

REFERENCES

[1] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand, “High efficiency video coding (HEVC) text specification draft 6,” JCTVC-H1003, Feb. 2012.

[2] Texas Instruments. OMAP3530 Technical Reference Manual. Literature Number SPRUF98X. June 2012.

[3] F. Pescador, G. Maturana, M.J. Garrido, E. Juárez and C. Sanz “An H.264 video decoder based on a DM6437 DSP”. IEEE Trans. on Consumer Electronics. Vol. 55, Nº 1. Pp. 205-212. February 2009.

[4] F. Pescador, E. Juarez, M. Raulet, C. Sanz "A DSP based H.264/SVC decoder for a multimedia terminal," Consumer Electronics, IEEE Transactions on , vol.57, no.2, pp.705-712, May 2011

[5] M.T. Pourazad, C. Doutre, M. Azimi, P. Nasiopoulos, "HEVC: The New Gold Standard for Video Compression: How Does HEVC Compare with H.264/AVC" IEEE Consumer Electronics Magazine, vol.1, no.3, pp.36-46, July 2012.

[6] Texas Instruments. TMS320DM6437 Technical Reference Manual. Literature Number SPRS345D. June 2008.

[7] HEVC Reference Software HM50. http://hevc.hhi.fraunhofer.de/

[8] DM6437 Digital Video Development Platform (DVDP). http://www.spectrumdigital.com/product_info.php?cPath=37&products_id=196&osCsid=0abf0072f9687529d1d010374287bd64

[9] Code Composer Studio v51. http://www.ti.com/tool/ccstudio&DCMP=dsp_ccs_v4

[10] SYS/BIOS 6.x. Real Time Operating System. http://www.ti.com/tool/sysbios&DCMP=B.

[11] Joint Call for Proposals on Video Compression Technology. ISO/IEC JTCI/SC39/WG11, N11113. Jan 2010.

[12] FFMPEG audio and video codec library. http://www.ffmpeg.org/

[13] Texas Instruments. TMS320C6472 Technical Reference Manual. Literature Number SPRS612G. July 2011.

Vision-based Sleep Mode Detection for a Smart TV

Yeong Nam Chae, Suwon Lee, ByungOk Han, and Hyun S. Yang, *Member, IEEE*
CS Dept., KAIST, Daejeon, Korea

***Abstract*—Sleep mode detection is one of the significant features of power management and green computing. However, for a television or a smart TV, it is difficult to detect a deactivation event because the user can use these devices without input from an input device. We propose a robust method to detect deactivation events based on a vision approach involving face detection and motion detection for a smart TV. Experiments are performed on a large dataset. The proposed approach significantly reduces false detections of faces and complement missed humans by means of motion detection.**

I. INTRODUCTION

The sleep mode is a low-power mode for electronic devices such as computers and smartphones. The sleep mode for computers is a significant feature of power management, which is an aspect of green computing. Currently, Smart TV, which is either a television set with integrated internet capabilities or a set-top box for television that offers more advanced computing abilities and connectivity characteristics than a contemporary basic television set, is becoming more widespread. Therefore, power management schemes for smart TVs are becoming important from the perspective of green computing. A deactivation event for conventional computing devices such as computers and smartphones can be easily detected because these devices interact with the user using only input devices such as a keyboard, a mouse or a touch-screen. If there is no input for a certain period of time from the input devices, a conventional computing device deactivates the current software and changes the current status to the sleep mode. However in the case of a television or a smart TV, it is difficult to detect a deactivation event because the user can use these devices without input from an input device. In this paper, we propose a robust method to detect deactivation events based on a vision approach involving face detection and motion detection. The human face is the important feature here for detecting a deactivation event. However, current face detection technology is not intended for commercial products due to shortcomings in the detection rate and number of false alarms when used in such applications. To reduce the number of false alarms, we adopt skin color filtering and background verification schemes. To enhance the ability to find missed humans by the face detector, we adopt a motion detection.

II. ALGORITHM

The overall deactivation detection procedure is shown in Fig. 1. The deactivation event is n times checked during the waiting time ω. The waiting time is ordinarily selected as 30 to 60 minutes by the user. The sampling number n is a design parameter which determines power consumption level and the deactivation detection accuracy. At every detection step, the proposed method analyzes an image at time t for face detection and analyzes image sequences from $t-\delta$ to $t+\delta$ for motion detection. If a face or motion is detected, the deactivation timer DT is initialized as 0. If both a face and motion are not detected, the deactivation timer is increased by ω/n. If the deactivation timer is greater than the waiting time ω, a deactivation event is detected and system goes into sleep mode.

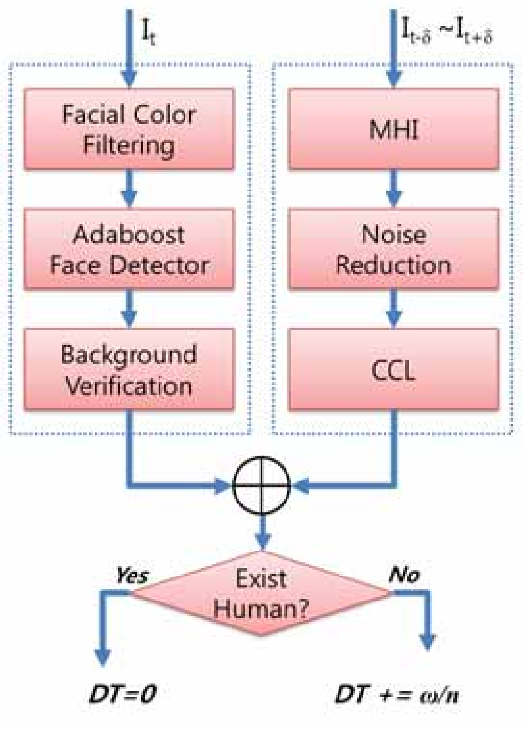


Fig. 1. Overall Procedure

In order to reduce the number of false alarms and the computation cost, we adopt our previous approach [1], which filters out non-facial color regions efficiently by means of sparse region scanning based on the facial color density in a given region. The proposed facial color model can include rare facial colors in the dataset; hence, this model can be robust against illumination variation. We enhance the facial color model by adding faces in difficult lighting conditions to the

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0005793). And this work was supported by the IT R&D program of MKE & KEIT [10041610, The development of the recognition technology for user identity, behavior and location that has a performance approaching recognition rates of 99% on 30 people by using perception sensor network in the real environment].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

training dataset. Though we reduced false detections using facial color filtering, there were still falsely detected faces in a real environment. Thus, we adopt background verification for robust face detection. The detected faces are verified through comparisons with the same region in the background image. The last background image is verified itself with other background images that are periodically archived during the sleep mode or when the power is off. To verify the detected faces with the same region in the background image, the Chi-square distance of the local binary pattern-based histogram representation [2], which is robust to illumination variations occurring by ambient lighting changes, is adopted as the measuring criteria.

$$\chi^2(x,\xi) = \sum_i \frac{(x_i - \xi_i)^2}{x_i + \xi_i}$$

In (1), x and ξ are the normalized enhanced histograms to be compared, and the indices i refer to the $i\,th$ bin in the histogram.

In order to detect motion from image sequences, we adapt the representation of the MHI (motion history image) to difference images from $t - \delta$ to $t + \delta$. The MHI collapses an image sequence into a 2-D image that captures spatial and temporal information pertaining to motion [3]. The MHI is known for its fast processing speed and its ability to represent short-duration motion. To reduce noise that arises from a vision sensor or an illumination condition, we perform the morphology operations [4] of erosion and dilation repeatedly. Next, we undertake CCL (connected component labeling) to find motion blobs that exceed a certain size.

III. EXPERIMENT

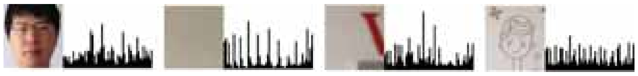
In this experiment, we verify the proposed face detection approach, as the motion detection approach adopted here is used widely in computer vision fields. To verify the facial-color-filtering based face detector with a large dataset, the Caltech 10,000 web faces dataset [5] was used as a test set. The Caltech dataset contains 7092 color images and has 10524 faces of various resolutions and all complexions in different settings. Included are portrait images, groups of people, and other configurations. The test set consists of only color images from the database. This test set includes 5525 color images and 8382 faces. The proposed method was applied to an AdaBoost face detector, which is state-of-the-art in terms of the detection rate and computational time. The AdaBoost face detector used in this experiment is supported by OpenCV [6]. In order to detect both frontal and profile faces, we combined one frontal classifier and two profile classifiers sequentially.

Table 1. Overall Results on the Caltech Dataset

		Conventional AdaBoost	Proposed face detector
Caltech Dataset 5525 images 8382 faces	DR	86.77%	86.95%
	FA/image	2.13	1.31
	Time	2473ms	1684ms

The overall result of the experiment is shown in Table 1. We compared the proposed method with the conventional AdaBoost face detector in terms of the detection ratio (DR), false alarm (FA) rate, and computation time. The experiment was conducted on a Pentium 4 2.4 GHz single-core PC.

As shown in Table 1, the overall false alarm rate was considerably reduced by 61%. Moreover, the computational time that determines power that is consumed was diminished remarkably. The detection rate of the proposed approach was slightly higher compared to that of the conventional AdaBoost face detector, as some false alarms led to a missed detection.



(a) *(b)* *(c)* *(d)*

Fig. 2. LBP Histogram Samples

To verify the proposed background verification method, we show some examples of LBP-based histogram representation. Fig. 2 (a) shows the face region in the real environment and its histogram representation. Fig. 2 (b), (c), (d) shows the sample background regions in the last background image and its histogram representations. This example shows that the false alarm region in the background image is distinguishable with a real face. Hence, we can eliminate false alarms by means of background verification.

IV. CONCLUSION

In this paper, we propose a vision-based method to detect deactivation events for sleep mode detection on a smart TV. In order to detect a deactivation event, we use a vision approach involving face detection and motion detection. To reduce the false alarm rate of the face detector, we adopt facial color filtering and background verification schemes. To enhance the finding of missed human faces by the face detector, we adopt a motion detection scheme. An experiment with a large dataset involving real environments was conducted. The proposed approach significantly reduces the false alarm rate of face detection and complements missed humans by means of motion detection. However, the proposed method can detect only upright faces; thus, we are currently expanding the proposed method to cover faces which appear at an angle.

REFERENCES

[1] Y. N. Chae, J. Chung, and H. S. Yang, "Color filtering-based Efficient Face Detection", *Proc. Of the International Conference on Pattern Recognition*, 2008, Tampa, USA

[2] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, No. 12, Dec., 2006, pp. 2037-2041.

[3] A. Bobick and J. Davis, "The recognition of human movement using temporal templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, No. 3, 2011, pp. 257-267.

[4] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image Analysis Using Mathematical Morphology", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, No. 4, July, 1987, pp. 532-550.

[5] M. Fink, R. Fergus, and A. Angelova, “Caltech 10, 000 web faces,” http://www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces/.

[6] “Open CV,” http://opencv.willowgarage.com/wiki/

Real-time Multi-Person Tracking in Fixed Surveillance Camera Environment

Jin-Woo Choi and Jang-Hee Yoo
Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea.

***Abstract*—In this paper, we propose a real-time multi-person tracking system operating in fixed surveillance camera environment. We adopt particle filtering as our object tracking framework. Background subtraction is used to generate the ROI. And pedestrian detection is used to initialize each tracker. Object size estimation and tracking failure detection is proposed to improve tracking accuracy and robustness. Experimental results demonstrate that the proposed algorithm tracks multiple persons efficiently in real-time.**

I. Introduction

Multi-person tracking is a critical problem in the intelligent video surveillance system. There are many challenges that make tracking a difficult problem such as illumination changes, occlusions, scale and shape changes, fast motions, and real-time processing. In order to solve these problems, several tracking algorithms have been proposed. One of the promising methods is *Particle Filter* based tracking method [1], [2]. Particle filter based tracking methods represent tracking uncertainty in a *Markovian* manner, thus it is suitable for online applications. In addition, particle filter tracking is robust to partial occlusion, rotation, and rapid motion. There are multiple objects tracking approaches utilizing both particle filtering and object detection technique. Okuma *et al*. [3] proposed a color-based particle filtering method which initializes each tracker by final detection result. Breitenstein *et al*. [4] extends this idea by using the detector confidence term. The detector confidence term is taken into observation likelihood and it makes tracking more robust.

In this paper, we propose a real-time multi-person tracking system, which is robust to depth variation in the surveillance video. The proposed method efficiently tracks multiple persons without tracker lost by using object size estimation and tracking failure detection technique.

II. Proposed Algorithm

A. Our object tracking approach

The proposed multiple objects tracking method tracks objects by particle filtering framework [2]. However, we extend the conventional particle filtering based tracking methods by several new ideas. These ideas will be explained in the following subsections. Frame difference based background subtraction is used to generate ROI (Region of Interest). HOG (Histogram of Oriented Gradients) and SVM (Support Vector Machine) based pedestrian detection algorithm [5] is used to initialize each tracker.

The state $\mathbf{s} = (x, v_x, y, v_y, w, h, \dot{s})$ consists of the 2D image position, the 2D velocity components, width and height of the object, and corresponding scale change rate. We use the *bootstrap filter* to approximate the probability distribution. Then the importance weight $w_t^{(i)}$ for each particle i at time step t is calculated by

$$w_t^{(i)} \propto w_{t-1}^{(i)} \cdot p(z_t \mid \mathbf{s}_t^{(i)}) . \tag{1}$$

In our method, similarity between the target color histogram and a sample color histogram is used to calculate the observation likelihood. The color histograms are constructed in the HSV color space using 6x6x6 bins. Then the observation likelihood of the each sample can be calculated by

$$p(z_t \mid \mathbf{s}_t^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1-\rho[p_{s_t^{(i)}}, q]}{2\sigma^2}\} , \tag{2}$$

$$\rho[p,q] = \sum_{u=1}^{m} \sqrt{p^{(u)} q^{(u)}} . \tag{3}$$

Where, $p_{s_t^{(i)}}$ is the color histogram of the sample $s_t^{(i)}$, q is the target histogram, and $\rho[p,q]$ is the Bhattacharyya coefficient, and σ is standard deviation of the color noise. We use a first order dynamic model for moving object described by,

$$\mathbf{s}_t = \mathbf{T}\mathbf{s}_{t-1} + \mathbf{w}_{t-1} . \tag{4}$$

A state transition matrix $\mathbf{T}$ propagates the samples with a first order motion model. It is assumed that velocity, width, height and scale change rate remain constant. $\mathbf{w}_{t-1}$ is a multivariate Gaussian random variable which gives perturbations to the state components.

B. Object size estimation

Particle filter based object tracking algorithms have weaknesses in object size estimation especially for the case that there is severe depth variation. Although we can define object width and height as state variables of the particle filter, it cannot estimate the exact size of the object with relatively small number of particles. Our method employs an explicit size estimation module to solve the problem. The object size estimation module estimates the current size of the object particle by

$$\begin{aligned} \hat{w}_t^i &= \overline{w}_{t-1} , \\ \hat{h}_t^i &= \overline{h}_{t-1} , \end{aligned} \tag{5}$$

This work was supported by the Industrial Strategic Technology Development Program, 10039149, Development of Basic Technology of Human Identification and Retrieval at a Distance for Active Video Surveillance Service with Real-time Awareness of Safety Threats funded by the Ministry of Knowledge Economy (MKE, Korea)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

where, $\bar{w}_{t-1}$ and $\bar{h}_{t-1}$ are the estimated object width and height of the previous frame. Then the size estimated object particles are distributed by Gaussian perturbation model as described in the previous subsection.

C. Tracking failure detection

The proposed criterion for tracking lost or success decision is described as follows.

$$Status(k) = \begin{cases} 0 & \text{if } \rho[p_{E[\mathbf{s}_t^k]}, q_k] < \rho_{th} \\ 0 & \text{if } \rho[p_{E[\mathbf{s}_t^k]}, q_k] - \rho[p_{E[\mathbf{s}_{t-1}^k]}, q_k] < \Delta_{th} \\ 1 & \text{otherwise} \end{cases} . \qquad (6)$$

$p_{E[\mathbf{s}_t^k]}$ is the color histogram constructed at the expected object position of k th object, q^k is the target color histogram of k th object, ρ_{th} and Δ_{th} are predefined threshold values. If the tracking failure detection module detects tracking lost by (6), lost signal 0 is transmitted to the pedestrian detection module to request a redetection of the lost target. Otherwise, tracker estimates the state of the object by particle filtering as usual.

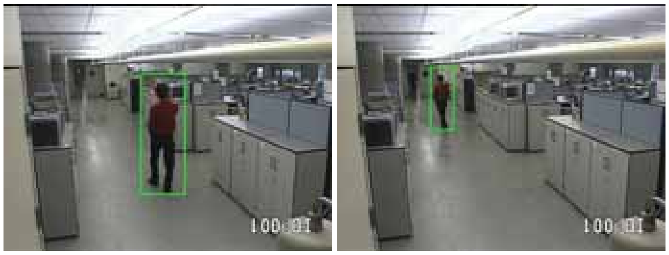
III. Experimental Results

The proposed people tracking algorithm is tested at a PC with quad-core 3.2GHz CPU. The number of particles per tracker is fixed to 100. The number of samples to construct a color histogram is set to 250. Standard deviation of the color noise σ is fixed to 0.2. Object size estimation threshold values ρ_{th} and Δ_{th} are set to 0.8 and -0.2 respectively. Test sequences are real-world surveillance video such as *ETRI hallway* made by us, *PETS 2009* and *AVSS 2007* which are publicly available.

First, we test the proposed object size estimation method using *ETRI hallway* sequence. Because this sequence shows severe depth variation, object scale changes drastically. As shown in Fig. 1, the proposed method successfully tracks the target and robustly estimates the size of the object although its scale changes.

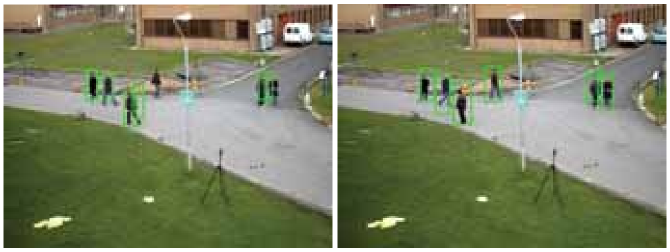
Second, we test the tracking failure detection module. The proposed method detects the tracker failure, and then requests the lost person redetection. As described in Fig. 2 (a), if tracking failure detection is off, the tracker for the man next to the street light fails and drifts to the wrong target. However, if tracking failure detection is on, our method detects the tracking failure and immediately tracks the original target again as can be seen in Fig. 2 (b).

Third, we compare our tracking method with naïve color-based particle filter [2]. Test sequence is *AVSS 2007* which includes severe depth variation. In Fig. 3 (a), [2] shows poor performance not only for estimating the size of the target but also for tracking the location of the target. In contrast, the proposed algorithm shows satisfying result both for location and the size of multiple persons as describe in Fig. 3 (b).



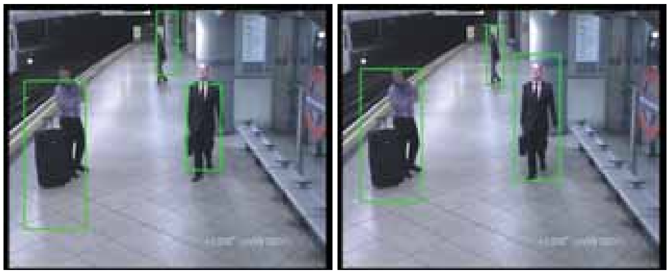
(a) *ETRI hallway* frame #366 (b) *ETRI hallway* frame #493

Fig. 1. Effectiveness of proposed object size estimation method.



(a) Tracking failure detection off (b) Tracking failure detection on

Fig. 2. Comparison of tracking failure detection on/off.



(a) Color-based particle filter [2] (b) Proposed algorithm

Fig. 3. Comparison with color-based particle filter [2].

In terms of processing speed, the proposed system shows 9.8 to 24.4fps depending on the number of targets in a sequence. Because our implementation was not optimized, processing speed can be increased further.

IV. Conclusion

In this work, we proposed a real-time multi-person tracking system for fixed surveillance camera environment. The proposed object size estimation method efficiently estimates size of the object. By the proposed tracking failure detection method, tracking accuracy and robustness is improved. The experimental result demonstrated that our real-time people tracking system outperforms the conventional algorithm.

References

[1] M. Isard and A. Blake, “Condensation-conditional density propagation for visual tracking,” in *Int. Journal of Computer Vision*, vol. 29, No. 1, pp. 5–28, 1998.

[2] K. Nummiaro, E. Koller-Meier, and L. V. Gool, “An adaptive color-based particle filter,” in *Image and Vision Computing*, vol. 21, No. 1, pp. 99–110, 2003.

[3] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little, and D. G. Lowe, “A boosted particle filter: multitarget detection and tracking,” in *ECCV*, 2004.

[4] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. V. Gool, “Robust tracking-by-detection using a detector confidence particle filter,” in *ICCV*, 2009.

[5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *CVPR*, 2005.

Exposure Guided Background Subtraction for Smart Cameras

Bin Wang[1], Zhihui Xiong[1], Yu Liu[1], Zhengshi Zhang[2], Wei Wang[1], Maojun Zhang[1]

1 College of Information System and Management, 2 College of Basic Education for Commanding Officers
National University of Defense Technology, Changsha, China

***Abstract*—An exposure guided background subtraction (EGBS) model is proposed for smart cameras to handle illumination change due to auto-exposure in visual surveillance. To reduce false foreground pixels caused by auto-exposure, EGBS compensates background illumination directly utilizing the information generated by auto-exposure module without extra illumination change estimation. Hence, it is very preferable for smart camera without any extra hardware resources. Experimental results indicate the proposed model efficiently reduces false foreground pixels caused by auto-exposure.**

I. INTRODUCTION

Background subtraction is one of the most widely used techniques to segment moving objects for static cameras in visual surveillance system. Many useful background subtraction methods such as mixture of Gaussian (MoG)[1], codebook and ViBe [2] have been proposed. However, most of them can't cope with fast illumination change [3,4] caused by light change or auto-exposure. In visual surveillance application, almost every surveillance camera supports auto-exposure to adapt to different illuminations. For example, in indoor visual surveillance, auto-exposure frequently occurs due to object moving while the background scene illumination doesn't change. When auto-exposure occurs, both the fixed background pixels and moving object pixels are all detected as foreground pixels.

In order to handle the fast illumination change, some illumination invariant features such as gradient or edge are used for background modeling. Gradient information is most discriminating at the boundaries of the objects, but does not provide a clear difference for large, untextrued objects (e.g. a white bus on the road surface). Furthermore, gradient information used for pixel-based background subtraction is highly unreliable when camera moves (such as shaking). Another way to handle illumination change is background illumination compensation[5,6], which considers the illumination compensation as an inverse problem. They [5,6] firstly estimate the illumination change and then compensate the background illumination. However, these algorithms require high estimation accuracy and high computational cost to estimate the illumination change. Therefore, these methods are not suitable for smart cameras (Fig.1.a) because of the limited resources, such as energy, processing power (eg: battery-powered camera) and memory in the cameras[7].

This work was supported by National Natural Science Foundation (NSFC) of China under project No.61175006, No.61175015, No. 60803101 and No.60872150; iCORE (Alberta Innovates) Alberta, and NSERC Canada.

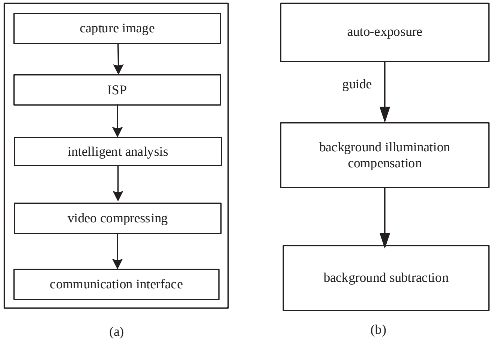


Fig.1 a: smart camera, b: exposure guided background subtraction

II. PROPOSED METHOD

In this paper, we propose a method that does not need to estimate the illumination change, and directly utilize the information generated by auto-exposure model of ISP to guide background illumination compensation (Fig.1.b). We propose an exposure guided background subtraction algorithm based on MoG. It directly uses exposure information to guide the MoG background model precisely updating before background subtraction, reducing the false foreground pixels in background. To avoid the non-linearity due to gamma function, we model the background on image dada before gamma correction. To the best of our knowledge, this is the first effort exploring utilizing the front-end device information (such as image signal processor, ISP) to improve intelligent analysis performance in smart cameras.

III. EXPOSURE GUIDED MOG BACKGROUND MODELLING

In this paper, we take account the auto-exposure module in ISP which adjusts the global image brightness by change the sensor gain, some local brightness adjust methods aren't take account in this paper. In visual surveillance, many auto-exposure methods adjust global image brightness with brightness change ratio R. It means that, if R is known, then the changed image brightness is also known, regardless of the adjustment process is linear or non-linear.

In typical video cameras, the brightness change ratio R is calculated as equation (1). The changed background illumination can be precisely compensated by utilizing R.

$$R = CurBrightness / T \arg etBrightness \tag{1}$$

Mixture of Gaussians (MoG) is a commonly used background subtraction method in visual surveillance since it can cope with periodic disturbances (swaying vegetation or flowing water). However, it can't cope with fast illumination change caused by auto-exposure. So we choose MoG as the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

baseline method and handle the fast illumination change caused by auto-exposure by utilizing the exposure information of camera.

The MoG method models background pixel's value distribution using K-Gaussians, and describe the probability of observing a pixel value X_t at time t as (2). K is the number of Gaussians, which is set to be 3 in our experiment. $\omega_{i,t}$, $u_{i,t}$ and $\Sigma_{i,t}$ are weight, mean and the covariance matrix of the i-th Gaussian in the mixture at time t, which are learned from a background sequence $B=(b_1,b_2...b_n)$.

$$P(X_t)=\sum_{i=1}^{K}\omega_{i,t}\eta(X_t,u_{i,t},\Sigma_{i,t}) \tag{2}$$

At time t, when an objet moves and changes the average brightness of current frame (in Fig.2.b), auto-exposure of ISP in smart camera will occur. The value of pixel in background X_t will be changed to X_t' with R as (3). R is the adjust factor (equation.1).

$$X_t'=RX_t \tag{3}$$

Then, X_t' no longer meets the background distribution which is only supported by original background sequence B. It leads to be set to false foreground. If the background sequence B is adjusted as (4), X_t' will meet to the adjusted background distribution which is supported by B'. And X_t' will avoid to be set to false foreground.

$$B'=(b'_1,b'_2...b'_n)=R(b_1,b_2...b_n) \tag{4}$$

Now, the K-Gaussians background model also needs to be adjusted to fit the new distribution of the adjusted background sequence B'. The K-Gaussians model is linear, in which, every component is a single Gaussian model with mean u_i and variance σ_i. After being adjusted as (4), the background pixels mean u_i and variance σ_i are respectively changed to u'_i and σ'_i as (5) and (6), where n_i is the number of pixels belong to i-th Gaussian, b_j^i and b'^i_j are the pixels belong to i-th Gaussian before and after auto-exposure adjusting.

$$u'_i=\frac{1}{n_i}\sum_{j=1}^{n_i}b'^i_j=\frac{1}{n_i}\sum_{j=1}^{n_i}Rb_j^i=Ru_i \tag{5}$$

$$\sigma'_i=\sqrt{E(b'^{i\,2}_j)-(u'_i)^2}=\sqrt{\frac{1}{n_i}\sum_{j=1}^{n_i}(b'^i_j)^2-(u'_i)^2}=\sqrt{\frac{1}{n_i}\sum_{j=1}^{n_i}(Rb_j^i)^2-(Ru_i)^2}=R\sigma_i \tag{6}$$

Since K-Gaussian mixture model is linear, so when auto-exposure occurs, the probability of observing a pixel value X'_t at time t can be described as (7). $\Sigma_{i,t}$ is equivalent to σ_i, since we assume that each color channel of RGB is independent.

$$P(X'_t)=\sum_{i=1}^{K}\omega_{i,t}\eta(X'_t,u'_{i,t},\Sigma'_{i,t}) \tag{7}$$

$$u'_{i,t}=Ru_{i,t} \tag{8}$$

$$\Sigma'_{i,t}=R\Sigma_{i,t}=R\sigma_{i,t} \tag{9}$$

The following steps to classify the pixels to foreground or background are like MoG.

IV. Experiment and results

We designed a smart camera which is an embedded system based on FPGA to evaluate our proposed method EGBS. We capture 10 test videos in indoor surveillance; Fig.2 illustrates one of these video pictures, and the foreground detection results are also shown in Fig.2. The quantization performance indicators of 10 test videos are shown in Table 1. We can conclude that, compared with the MoG method, the proposed method improves segmentation performance with a higher F-score against auto-exposure from the experiments.

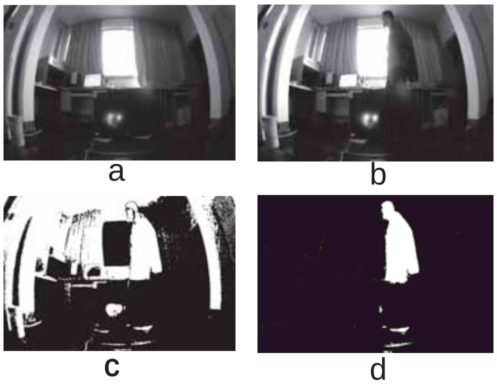


Fig.2 Results of foreground detection in one test video, a: Background image, b: Auto-exposure image, c: MoG, d: Exposure guide MoG

TABLE 1: PERFORMANCE INDICATORS OF MOG AND PROPOSED METHOD

AVERAGE (%)	MOG	EGBS (EGMOG)
PRECISION	22.31	80.15
RECALL	42.22	72.24
F-SCORE	28.94	69.56

V. Discussions

Directly using the exposure information to guide background model illumination compensation is an effective method to handle illumination change caused by auto-exposure. It's a light-weight and efficient method with less computational cost. And it is suitable for smart cameras. In the future work, we will extensively explore the combination of ISP and intelligent analysis algorithms to enhance the smart camera performance with less resource consumption.

Reference

[1] C.Stauffer.et al. "Learning patterns of activity using real-time tracking". IEEE T PATTERN ANAL, vol. 22, no. 8, pp: 747-757, 2000.

[2] Olivier Barnich.et.al, "ViBe: A universal background subtraction algorithm for video sequences". IEEE Transactions on Image Processing, vol.20, no.6, pp.1709-1724. 2011.

[3] Liu,L.Y. et al. "Background subtraction using shape and colour information". Electronics Letters, vol.46, no.1, pp: 41-43.2010

[4] JinMin Choi.et al. "Robust moving object detection against fast illumination change". COMPUT VIS IMAGE UND, vol.116, pp:179-193, 2012.

[5] Vasu Parameswaran. et al. "Illumination Compensation Based Change Detection Using Order Consistency", CVPR, 2010

[6] Darnell Janssen Moore, "Dynamic illumination compensation for background subtraction", United States Patent, 2012

[7] Casares,M. et al. "Light-weight salient foreground detection for embedded smart cameras", COMPUT VIS IMAGE UND, vol.114, pp:1223-1237, 2010

Intelligent control for adaptive video streaming

V. Menkovski, *Student Member*, *IEEE*, A. Liotta, *Member*, *IEEE*

***Abstract*—We present an autonomous learning agent for adaptive video streaming in best effort networks. The agent learns an optimal control strategy in regards to the delivered quality of experience without the need for implementation of a complex heuristics.**

I. INTRODUCTION

Video streaming functionality is present in most, if not all, web enabled devices. For a device to achieve delivery of a high quality streaming service it needs to continuously reproduce the video with sufficient bit-rate and without any errors. However, in best-effort networks multiple sources are competing for the same resources and therefore no guarantees are given that resources will be available when needed [1]. Since video streaming is a data-intensive process it is particularly susceptible to variations in throughput. If the resources are not sufficient the video playback will freeze or, otherwise, the video will have to be streamed with lower bit-rate.

To address the variability in available resources, adaptive streaming technologies are developed such as HTTP streaming [2] and SVC [3]. These technologies allow for continuous adaptation of the bit-rate as the video is being streamed so that a controlled degradation in quality, or quality of experience (QoE), can be achieved.

Adaptive streaming clients available today predominantly demonstrate a strategy of streaming at the highest possible bit-rate, for as long as possible [4]. Only when the buffer becomes depleted they switch to lower bit-rate levels. This approach does not take into account the subjective perception of quality [5], nor the degradation to QoE that comes from frequent changes between quality levels [6]. These ‘greedy’ strategies can further lead to over-provisioning of bit-rate with little to no positive effect on the QoE, due to the nonlinear dependence between bit-rate and user perception (Fig 1). Further, this leads to a higher probability of buffer depletion and playback freeze, which overall results in a lower QoE.

In this paper we propose a solution that addresses these issues by developing a QoE estimation function [7] for adaptive video streaming that incorporates: 1) the subjectively perceived quality; 2) the impairments from playback freeze; and 3) the effect of the frequency and amplitude of change in quality. Then we propose a design for an intelligent streaming agent that optimizes its decision strategy based on the subjective QoE delivered to the customer.

The agent uses a reinforcement learning [8] method to infer the optimal decisions trained in an environment of simulated background traffic. This approach does not require design of a control heuristic for the agent, which can adapt its strategy based on the understanding of the subjectively perceived quality. The intelligent adaptive streaming client hence provides better utilization of the network resources and higher QoE, for a wide range of network conditions.

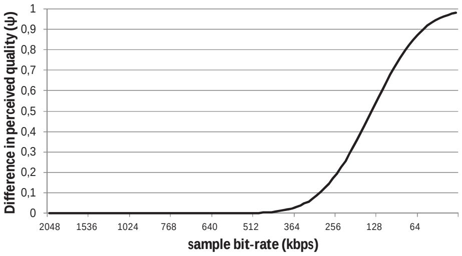


Fig 1. Subjective degradation in quality

II. METHOD

The HTTP adaptive streaming architecture is composed of a HTTP server, with access to a video database, and a client application, running on a connected device (Fig. 2). The client requests chunks of video from the server and reproduces the video on the device display. The client can choose to request chunks at different quality levels (L_1, L_2, etc), based on its estimate of network throughput and its own control strategy.

Our agent, residing on the client, evaluates its performance based on a broader view of delivered QoE given in (1).

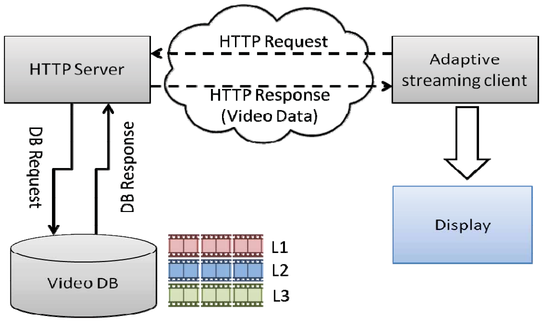


Fig 2. System description

$$\begin{aligned} QoE \;=\; & w_s f_{subjective}(bit\_rate, video\_si, video\_ti) + \\ & + w_f f_{freeze}([(len_1, T_1), (len_2, T_2)...(len_n, T_n)]) + \\ & + w_l f_{lvlChange}([(delta_1, T_1), (delta_2, T_2)...(delta_k, T_k)] \end{aligned} \tag{1}$$

The $f_{\text{subjective}}$ function calculates the degradation due to restrictions in bit-rate. It takes into account the characteristics of the video (spatial and temporal information) and returns a

This work has been partially supported by ARTEMIS project DEMANES (Design, Monitoring and Operation of Adaptive Networked Embedded Systems, contract 295372).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

relative value of degradation. A typical subjective quality curve obtained with the Maximum Likelihood Difference Scaling method [9] is presented in Fig. 1.

The f_{freeze} function calculates the degradation incurred by a freeze in playback. The value is based on research done on the psychological effects of this type of impairment. The degradation value given is proportional to the length of the freeze and to the frequency of these occurrences. The f_{freeze} inputs a list of pair values. The first (len_i) is the length of the event and the second (T_i) is the time at which it occurred. This way the recent events have bigger effect and the older ones have smaller (decayed by $e^{-\lambda \Delta t}$). The total impairment is a sum of the effect of each event from the beginning. The same approach is taken for the $f_{\mathrm{lvlChange}}$, where the $delta_i$ is the distance between the levels and T_i is the moment of occurrence. Since the three types of impairments have different amount of impact they are weighted differently: w_s, w_f and w_l. These weights can be adjusted according to results from subjective trials.

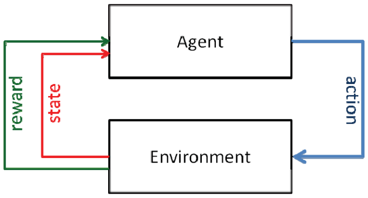


Fig 3. Reinforcement learning loop

Generally, a reinforcement learning framework consists of an agent working in an environment (Fig. 3). The agent executes actions over the environment and receives state changes and reward feedback from it.

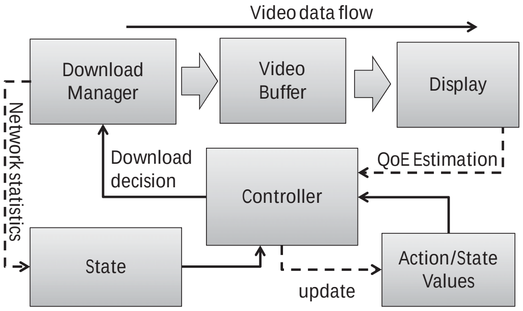


Fig 4. Design of the RL Intelligent agent

In our case the actions are the download requests at different bit-rates; the reward is the QoE estimation; and the state is a combination of the video buffer condition, network throughput estimation and the position in the video stream (Fig. 4).

The estimation of the network throughput is implemented by a set of filters: Exponentially Weighted Moving Average with fast and slow tracing and stability estimator. The filters combine historical throughput measurements to predict the current throughput availability. For the training of the agent we implemented the linear gradient-descent SARSA(lambda) algorithm [8].

III. RESULTS

The agent is trained in an environment of simulated background traffic. The values for the weights of the QoE function were selected as 1 for w_s, 2 for w_l and 10 for w_f. The background traffic is modeled with a self-similar process, where the number of file transfers is sampled from a Poisson distribution, and the length of each transfer sampled from a Paretto distribution. The Hurst parameter for the background traffic is set to 0.7. The agent learns to avoid freezes quickly, since this is heavily penalized in the QoE function. The estimated QoE incurred over after each episode is given in Fig. 5. We trained the agent on 1000 episodes, and found that the efficiency of the inferred control strategy improves considerably as new training episodes are included.

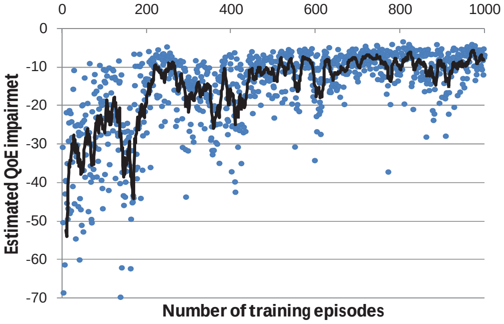


Fig. 5. Performance of the RL intelligent agent

IV. CONCLUSION

Our approach provides a flexible solution for the problem of adaptive streaming, but could also be used more generally in other control problems where it is hard to model the system deterministically. Another direct benefit is the inclusion of human perception (subjective) factors in the decision. Future work should explore the depth of predictive capabilities and more complex QoE non-linear dependencies between the given factors, as well as training on network traces.

REFERENCES

[1] F. Agboma and A. Liotta, “Quality of experience management in mobile content delivery systems,” *Telecommunication Systems*, vol. 49, no. 1, pp. 85–98, 2012.

[2]“ISO/IEC 23009-1:2012 - Information technology -- Dynamic adaptive streaming over HTTP (DASH) -- Part 1: Media presentation description and segment formats.”

[3] G. Van der Auwera, P. T. David, M. Reisslein, and L. J. Karam, “Traffic and quality characterization of the H. 264/AVC scalable video coding extension,” *Advances in Multimedia*, vol. 2008, no. 2, pp. 1–27, 2008.

[4] F. Bertone, V. Menkovski, and A. Liotta, “Adaptive P2P Streaming,” in *Streaming Media with Peer-to-Peer Networks*, IGI Global, 2012, pp. 52–73.

[5] V. Menkovski and A. Liotta, “Adaptive psychometric scaling for video quality assessment,” *Signal Processing: Image Communication*, vol. 27, no. 8, pp. 788–799, Sep. 2012.

[6] P. Ni, A. Eichhorn, C. Griwodz, and P. Halvorsen, “Frequent layer switching for perceived quality improvements of coarse-grained scalable video,” *Multimedia Systems*, vol. 16, no. 3, pp. 171–182, 2010.

[7] V. Menkovski, G. Exarchakos, and A. Liotta, “Online QoE prediction,” in *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*, 2010, pp. 118 –123.

[8] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1. Cambridge Univ Press, 1998.

[9] V. Menkovski, G. Exarchakos, and A. Liotta, “The value of relative quality in video delivery,” *J. Mob. Multimed.*, vol. 7, no. 3, pp. 151–162, Sep. 2011.

Intelligent Document Capturing and Blending System Based on Robust Feature Matching with an Active Camera

Wan-Yu Chen, Jia-Lin Chen, Yu-Chi Su, and Liang-Gee Chen, *Fellow, IEE*E
DSP/IC Design Lab., Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan

Abstract-- **We propose an intelligent document capturing and blending system based on robust feature matching for efficient document management. The proposed system not only supports handwritten text and figure extraction, but also provides image blending mechanism to automatically merge the extracted handwritten texts and figures into electronic documents for the user. The proposed system addresses camera shake and luminance variation problems caused by active cameras. Besides, we adopt robust feature matching techniques to improve the system accuracy. Experimental results show that our system supports 95.65% detection rate and achieves 88.3% compression ratio reduction compared with the previous work. Besides, we also compare system performance considering Scale Invariant Feature Transform (SIFT) [1] and Speeded-Up Robust Features (SURF) [2]. We derive 71.2% complexity reduction and 4.3% detection rate degradation with SURF feature matching.**

I. INTRODUCTION

Recently, because of the facility of information indexing, retrieval, storage and exchange, documents in electronic form are more and more popular. However, a lot of paper documents still exist in our daily life. For example, when attending a meeting or course, we usually print the presentation documents and write some notes or comments on the paper documents. Such behavior helps us to understand and recall the presentation easily. In addition, we usually print technical papers and write some innovation idea on the papers when we read them. This action helps us to record our instant idea immediately and conveniently. Hence such handwritten record is valuable and needs efficient management. However, paper documents consume large storage space. And it is time-consuming to search a note from a lot of paper documents. The inconvenient management of paper documents makes our valuable idea hard to be maintained.

Along with the development of hand-held camera and scanner, people choose to convert paper documents into electronic form. Traditional flatbed scanner has better image quality but users usually feel cumbersome to use it. Hand-held camera is more popular than flatbed scanner. People usually carry a cellphone with camera every day and can capture a photo anytime. While the image sensor converts the paper documents into electronic image files, the captured image is constituted with image pixels without registration. And the image size is proportional to image quality. Several previous works have been presented to register the captured image for active camera applications [3]-[5]. Park et al. developed a vertical line detection method for image registration [3]. Kim et al. adopted rectangle line feature detection method using Hough transform [4]. Yuan et al. proposed a robust feature matching approach with pre-defined layout for better registration quality [5]. However, such capturing approach only assumes the document with the pre-defined layout and cannot adapt to dedicated document layout. Besides, the previous works only reduces the storage of paper document, users still feel inconvenient to search item from a lot of image files.

In this paper, we propose a novel document capturing and blending system for efficient document management. With robust feature matching over the original electronic documents, we successfully register the captured images with dedicated layout. Furthermore, the handwritten texts and figures are extracted and blended with original electronic documents by the proposed texture extraction and blending mechanism. The system aims to provide the user key-word search function of electronic documents and keep the handwritten texts and figures simultaneously. Users can share their handwritten record easily with cloud and local device cooperation. Besides, we address registration instability problems with an active camera and improve the registration quality by matching the dedicated layout. Furthermore, the compression ratio can be largely reduced due to handwritten texture extraction. Finally, we speed up the whole system with SURF feature matching with acceptable accuracy degradation.

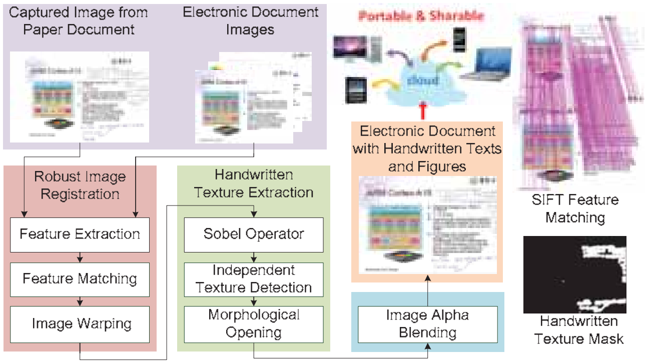

Captured Image from Paper Document
Electronic Document Images
Portable & Sharable
cloud
Electronic Document with Handwritten Texts and Figures
SIFT Feature Matching
Robust Image Registration
Feature Extraction
Feature Matching
Image Warping
Handwritten Texture Extraction
Sobel Operator
Independent Texture Detection
Morphological Opening
Image Alpha Blending
Handwritten Texture Mask

Fig. 1 System block diagram

II. PROPOSED METHOD

The overall block diagram of the system is shown in Fig. 1. The captured image is processed through feature extraction and matching to derive the global camera motion for robust image registration. The following is an image warping process to register the captured document based on estimated camera homography matrix. Second, we propose a handwritten texture extraction mechanism considering both original electronic document image and transformed captured image. Finally, we apply image alpha blending with the extracted handwritten texture mask on original electronic image to preserve handwritten texts and figures for the user.

A. Robust Image Registration

To register the captured image in a dynamic environment, a robust feature extraction and matching approach is employed as the first step of the framework. For each image, global camera homography matrix needs to be estimated before handwritten texture extraction. This is because the global camera motion mounted on a mobile agent differs frame by frame. By taking the advantage of robust feature extraction and matching techniques, we match corresponding feature points of the captured image with the original electronic document regardless of environment variation. To complete robust image registration, we separate the process into three steps. Firstly, SIFT features [1], a local descriptor with good scale, rotation, and luminance invariance, are extracted from the input image. Next, for each feature in the input image, feature matching is performed to find the nearest neighbors among reference images in the database. To speed up the processing time of the matching stage, kd-tree [6] is adopted as the index of the image database. After this step, hundreds of matching pairs for each frame in the input video are obtained. In addition, to remove false matching among matching pairs, RANSAC [6] is employed to filter outlier pairs. RANSAC algorithm iteratively selects samples at random among matching pairs and estimates their

978-1-4673-1363-6/13/$31.00 ©2013 Crown

homography matrix as the fitting model. Finally, we apply image warping using the estimated homography matrix H on the captured image as equation (1). $(\tilde{x},\tilde{y})$ is image coordinate after image warping.

$$H=\begin{bmatrix} h11 & h12 & h13 \\ h21 & h22 & h23 \\ h31 & h32 & h33 \end{bmatrix} \qquad \begin{bmatrix} w*\tilde{x} \\ w*\tilde{y} \\ w \end{bmatrix} = \mathrm{H}*\begin{bmatrix} \mathrm{x} \\ \mathrm{y} \\ \mathrm{z} \end{bmatrix} \tag{1}$$

B. Handwritten Texture Extraction

A Sobel filter-based texture detection technique combined with morphological opening operation is used to provide robust handwritten texture extraction. We design an independent Sobel filter-based mechanism to extract the texture area regardless of luminance and color variance between captured image and electronic document image. The idea of Sobel filter-based mechanism comes from the fact that handwritten texture detection reveals its strong edge response on captured image and weak edge response on original electronic document image. To reduce the image blur effect caused by homography transform, 10x10 window calculation is adopted in the proposed system. In this situation, handwritten texture mask is extracted more accurately.

Equation (2) shows Sobel filter operation. Equation (3) derives independent texture extraction based on the transformed paper document image *f(x,y)* and electronic document image *f'(x,y)*. On the other hand, even if most handwritten texture can be successfully extracted by the independent texture detection step, it usually fails to distinguish noise of captured image from handwritten texture, especially under severe camera noise. In this case, morphological opening operation is adopted to reduce handwritten texture misdetection caused by camera noise.

$$\begin{aligned} &G_x = f(x-1,y-1)+2f(x-1,y)+f(x-1,y+1))-(f(x+1,y-1)+2f(x+1,y)+f(x+1,y+1)) \\ &G_y = f(x-1,y-1)+2f(x,y-1)+f(x+1,y-1))-(f(x-1,y+1)+2f(x,y+1)+f(x+1,y+1)) \\ &Sobel\,(f(x,y)) = |G_x|+|G_y| \end{aligned} \tag{2}$$

$$\begin{aligned} &If\,((\sum_{(i,j)\in N} Sobel\,(f'(x+i,y+j)) < th) == 0)\ \&\&\ (Sobel\,(f(x+i,y+j)) > th) \\ &\quad texture\,(x,y) = true \\ &else \\ &\quad texture\,(x,y) = false \end{aligned} \tag{3}$$

C. Image Alpha Blending

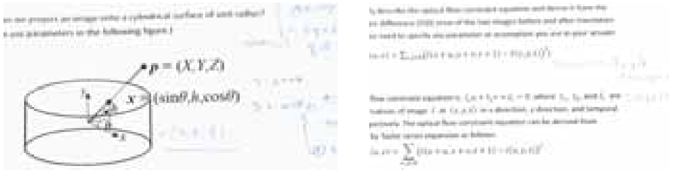
The image alpha blending is applied on the original electronic document with the handwritten texture mask extracted from the previous steps. Thus handwritten texts and figures are blended with the electronic document according to the texture mask.

When all the procedures illustrated above are done, the next captured image is processed by the system in a similar way. Most importantly, handwritten texture extraction and blending in this work are employed to provide both the convenient management and instant recording facility of electronic and paper documents for users.

III. EXPERIMENTAL RESULTS AND ANALYSIS

Two experiments are conducted to evaluate performance of the proposed system. All the images tested in experiments are captured from a CMOS front-mounted Canon S100 camera with 640×480 resolution.

Table I shows the system improvement on detection rate involving SIFT or SURF image registration mechanism, respectively. With the proposed image registration mechanism, the detection rate is improved to 95.65% and 91.3% considering SIFT and SURF techniques. Fig. 2 depicts the success and failure case of SURF based image registration. From Fig.2 (a), we can see that an image with figures could be easily detected by SURF. But an image containing a lot of texts and no distinguished layout, as shown in Fig. 2 (b), is difficult for SURF matching. In this case, the SIFT technique can distinguish the feature difference of individual texts and achieves higher accuracy.



(a) (b)

Fig. 2 Results of success and failure images. (a) Success image with SURF approach. (b) Failure image with SURF approach.

TABLE I ACCURACY COMPARISON

Detection method	Accuracy (%)
SIFT feature matching	95.65
SURF feature matching	91.30

The second experiment evaluates the compression ratio of the proposed system. In the proposed system, user only needs to store the handwritten texture area and reduce the disk space. This experimental result demonstrates that our system provides 88.3% compression ratio reduction compared with the previous method [4].

TABLE II TIME COMPLEXITY ANALYSIS

Module	SIFT(%)	SURF(%)
Feature Extraction	88.64	74.39
Feature Matching	5.08	0.95
Image Warping	1.09	3.79
Others	6.09	21.07

Table II summaries the workload ratio of each module by software implementation for the proposed system. The operating environment is under Win7 operating system with Intel Core I7 3.4G CPU and 4GB DDR RAM. This result shows feature extraction of robust image registration dominate 88.64% of the whole process. The system process takes 5.29 second in average to deal with a frame with SIFT techniques. To speed up the system performance, we adopt SURF feature extraction. Thus, our system is accelerated to 1.53 second in average per frame and this processing speed is acceptable by user experience.

IV. CONCLUSION

We propose an efficient document capturing and blending system based on robust feature matching for active camera applications. With the proposed system, we can digitalize handwritten texts and figures on the paper documents and record it on electronic documents automatically. Thus, we can exploit both the instant recording advantages of paper documents and convenient management of electronic documents.

Our system achieves 95.65% detection rate with SIFT feature matching technique and 88.3% compression ratio reduction compared with the previous work [4]. In addition, we also derive 71.2% complexity reduction and 4.3% detection rate degradation with SURF features.

REFERENCES

[1] D. G. Lowe,"Distinctive image features from scale-invariant keypoints", published by International Journal of Computer Vision, 2004.

[2] H. Bay, et al., "SURF: Speeded Up Robust Features", ECCV, Vol. 110, pp. 407-417, 2006.

[3] A. Park, et al., "Intelligent document scanning with active camera", ICDAR, pp. 991 – 995, vol. 2, 29 Aug.-1 Sept. 2005.

[4] W. H. Kim, et al., "Document Capturing Method with a Camera Using Robust Feature Points Detection", DICTA, pp.678 – 682, Dec. 2011.

[5] V. G. Edupuganti, et al., "Registration of camera captured documents under non-rigid deformation", CVPR, pp. 385-392, 2011.

[6] N. Y. Khan, et al., "Performance Evaluation against Various Image Deformations on Benchmark Dataset", DICTA, pp.503 – 506, Dec. 2011.

Pose Estimation of a Depth Camera Using Plane Features

Seon-Min Rhee[1], Yong-Beom Lee[1], James D. K. Kim[1], Taehyun Rhee[2]
[1]*Samsung Advanced Institute of Technology, Samsung Electronics, Republic of Korea*
[2]*Victoria University of Wellington, New Zealand*

***Abstract* — We present a novel method for pose estimation of a depth camera through plane features. Since conventional features for color images, mainly points and lines, are not applicable to a depth image, we propose a new type of feature utilizing planar structures in a scene. To measure the accuracy of our method, we generated a synthetic scene and calculated a position error between an estimated location and its ground truth. We also applied our method to the real world scene captured by a depth camera verifying practical usage.**

I. Introduction

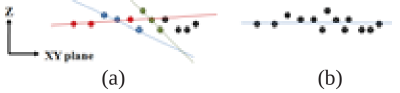
A depth camera is proliferating and emerging as a powerful consumer electronic device recently. Many commercial products have already been released and propagated to various technical fields such as user interface, robot navigation, gesture recognition [1] and so on. Pose estimation of a depth camera, however, is a still challenging technical problem since conventional features, mainly points or lines which have been used for a color image such as [2], are not directly applicable to a depth image.

In this paper, we propose a novel approach for pose estimation of a depth camera using a new type of feature, i.e. planes. Our approach utilizes geometry information such as planar structure in a scene captured by a depth camera. Since there may be several natural planes in an indoor environment such as walls and ceilings for example, we assume that a certain scene has enough planes to be referred. In our method, first we detect multiple planar structures in a scene through a split and merge approach [3]. Then, we select reference planes and they are tracked in the following consecutive frames by similarity comparison to estimate camera pose.

II. Proposed Method

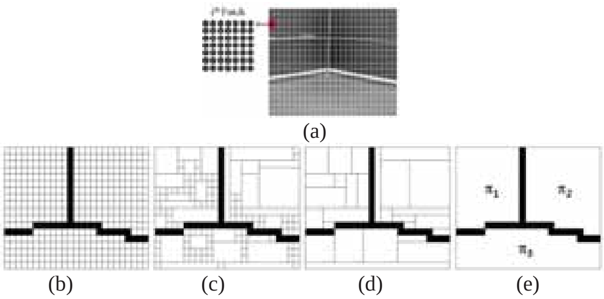
A. Extract Plane Features

First we detect planar structures in a scene using a set of 3D points captured by a depth camera to define a plane feature. While RANSAC [4] is a widely known method to estimate a plane model in color images, it is not suitable for a depth image due to depth noise characteristic. Since only three points are used for estimating a plane in RANSAC method, an influence of the depth noise is critical as shown in Fig 1(a). From our observation, reliable model estimation is expected when multiple points are considered as illustrated in Fig 1(b). To estimate more robust plane models, we split an entire depth image into small patches containing multiple points as shown in Fig 2(a).



(a) (b)

Fig 1: Model estimation of a plane using (a) three points and (b) multiple points by least square fitting



(a)

(b) (c) (d) (e)

Fig 2: Detect planar structures (a) split a depth image into small patches (b)-(e) iterative merging: patches not on a planar structure are discarded which are depicted as a black.

Next upon an initial assumption that all patches are on a planar structure, we estimate their parametric plane models using least square fitting [5]. Then we perform R^2 test [6] to check *goodness of fitting* and discard the patches whose value of R^2 is under a certain threshold. This implies that the estimated plane model cannot properly explain the distribution of the points in the patch and therefore the patch is not on a planar structure. The rest of patches are regarded as being positioned on a planar structure and iteratively merged in case they have similar orientation and the distance from the origin. The plane models are then re-estimated by the least square fitting using all the points in the merged patches. This procedure is iterated until all the patches in the same planar structure are merged together. This process is illustrated in Fig 2(b)-(e), where the discarded patches not on planar structures are depicted as black squares.

The plane feature of a detected planar structure $\boldsymbol{\pi}_\mathbf{i}$ is represented by $\mathrm{N}_\mathrm{i}(a_i, b_i, c_i)$ and d_i, the plane normal and the distance between the origin and the plane respectively.

B. Track Plane Features

We define *reference planes* as the ones which are tracked continuously in the following frames. To estimate 6-DOFs camera pose of a current frame, at least three pairs of reference planes and the corresponding planes in the current frame are required. Since there may be several planes in a scene, it is critical to select proper reference planes among all detected planes for expecting robust pose estimation. To select the proper reference planes, two major points are considered. A size of a plane is considered first because a larger plane is easy to be detected robustly in the following frames. Thus a plane which has the highest number of points in the planes is considered as a reference plane. Two additional planes which scores 2nd and 3rd highest number of points are considered as candidates of reference planes. To guarantee feature distinctiveness, if the candidate planes have similar orientation and the distance with those of already selected reference planes, we consider the plane as a reference plane which has the next highest number of points.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Once the reference planes are decided in a reference frame, the first frame for example, corresponding planes are tracked in the consecutive frames by comparing plane features of all detected planes in the current frame. The corresponding plane *m* of the reference plane *n* can be found by:

$$\mathrm{argmax}_{\mathrm{m \in detected\ planes}}(\mathbf{N}_\mathrm{n} \cdot \mathbf{N}_\mathrm{m}) \ < \varepsilon_1 \ \& \ dist|\ \mathrm{d_n} - \mathrm{d_m}| < \varepsilon_2,$$

where ε_1 and ε_2 are thresholds.

C. Estimate Camera Pose

Using the plane correspondence, we can calculate the pose (rotation and translation) of the camera. The reference planes are notated by $\boldsymbol{\pi}_\mathbf{k}^\mathbf{r}$ using plane feature and their corresponding planes in i$^{\text{th}}$ frame are notated by $\boldsymbol{\pi}_\mathbf{k}^\mathbf{i}$, where *k* is the number of the planes from 1 to 3:

$$\boldsymbol{\pi}_\mathbf{k}^\mathbf{r}\!: \ \mathrm{a_k^r x + b_k^r y + c_k^r z + d_k^r = 0},$$
$$\boldsymbol{\pi}_\mathbf{k}^\mathbf{i}\!: \ \mathrm{a_k^i x + b_k^i y + c_k^i z + d_k^i = 0}.$$

Using the plane correspondences, we calculate a rotation matrix $\mathbf{R}_{3\mathrm{x}3}$ of the camera by solving the linear equation:

$$\begin{bmatrix} a_1^r & a_2^r & a_3^r \\ b_1^r & b_2^r & b_3^r \\ c_1^r & c_2^r & c_3^r \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} a_1^i & a_2^i & a_3^i \\ b_1^i & b_2^i & b_3^i \\ c_1^i & c_2^i & c_3^i \end{bmatrix},$$

$$\begin{bmatrix} a_1^i & b_1^i & c_1^i & 0 & 0 & 0 & 0 & 0 & 0 \\ a_2^i & b_2^i & c_2^i & 0 & 0 & 0 & 0 & 0 & 0 \\ a_3^i & b_3^i & c_3^i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_1^i & b_1^i & c_1^i & 0 & 0 & 0 \\ 0 & 0 & 0 & a_2^i & b_2^i & c_2^i & 0 & 0 & 0 \\ 0 & 0 & 0 & a_3^i & b_3^i & c_3^i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_1^i & b_1^i & c_1^i \\ 0 & 0 & 0 & 0 & 0 & 0 & a_2^i & b_2^i & c_2^i \\ 0 & 0 & 0 & 0 & 0 & 0 & a_3^i & b_3^i & c_3^i \end{bmatrix} \begin{bmatrix} R_{11} \\ R_{12} \\ R_{13} \\ R_{21} \\ R_{22} \\ R_{23} \\ R_{31} \\ R_{32} \\ R_{33} \end{bmatrix} = \begin{bmatrix} a_1^r \\ a_2^r \\ a_3^r \\ b_1^r \\ b_2^r \\ b_3^r \\ c_1^r \\ c_2^r \\ c_3^r \end{bmatrix}. \quad (1)$$

Since the computed **R** in (1) does not satisfy the property of a rotation matrix generally, it is necessary to approximate **R** to the best rotation matrix. Let the singular value decomposition of **R** be $\mathbf{UVD}^\mathrm{T}$. Then the best rotation matrix is known to be $\mathbf{R} = \mathbf{UD}^\mathrm{T}$ [7].

To estimate the translation $\mathbf{T}_{3\mathrm{x}1}$, we use a distance from a point to a plane. Let D_k^i denotes the distance between a point whose position moves from the origin of the reference frame to the current camera position $(\mathbf{-R^{-1}T})$ and a plane π_k^1. Then D_k^i is same with d_k^i and therefore, **T** can be calculated using:

$$D_k^i = \frac{\left| -a_k^r(\mathbf{R}^{-1}\mathbf{T}) - b_k^r(\mathbf{R}^{-1}\mathbf{T}) - c_k^r(\mathbf{R}^{-1}\mathbf{T}) + d_k^r \right|}{\sqrt{(a_k^r)^2 + (b_k^r)^2 + (c_k^r)^2}},$$

$$\begin{bmatrix} d_1^i \\ d_2^i \\ d_3^i \\ 1 \end{bmatrix} = \begin{bmatrix} -R_{11}a_1^r - R_{12}b_1^r - R_{13}c_1^r & -R_{21}a_1^r - R_{22}b_1^r - R_{23}c_1^r & -R_{31}a_1^r - R_{32}b_1^r - R_{33}c_1^r & d_1^r \\ -R_{11}a_2^r - R_{12}b_2^r - R_{13}c_2^r & -R_{21}a_2^r - R_{22}b_2^r - R_{23}c_2^r & -R_{31}a_2^r - R_{32}b_2^r - R_{33}c_2^r & d_2^r \\ -R_{11}a_3^r - R_{12}b_3^r - R_{13}c_3^r & -R_{21}a_3^r - R_{22}b_3^r - R_{23}c_3^r & -R_{31}a_3^r - R_{32}b_3^r - R_{33}c_3^r & d_3^r \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \\ 1 \end{bmatrix}.$$

III. Experimental Results

We generated a synthetic *office* scene having several planar structures to measure the accuracy of our method. We also experimented with a real world scene captured by MESA SR4000 containing three planes to verify practical usage. The test scenes are shown in Fig 3.

To estimate the accuracy of the suggested method, position error is calculated for 1500 frames quantitatively for the synthetic scene according to the change of the camera pose. We calculate average Euclidean distance between ground truth position and its estimated position:

$$\frac{1}{n}\sum_{i=0}^{n} \mathrm{abs}(\mathbf{P}_i^g - \mathbf{P}_i^e), \qquad (5)$$

where *n* is the number of frames, $\mathbf{P}_i^g$ is a position of 1000 unit distance from the location of the ground truth camera and $\mathbf{P}_i^e$ is a estimated position of $\mathbf{P}_i^g$ by the suggested method. Ideally $\mathbf{P}_i^e$ is identical with $\mathbf{P}_i^g$. The position error was 1.59 ± 0.056 unit with the 90% confidence intervals.

Fig 3(b) and 3(c) compares plane detection results from our approach with RANSAC implemented by [8]. From the aspect of computing time, it takes 99.8msec by our method and 359.2msec by RANSAC respectively. Obviously our method outperforms RANSAC in terms of the accuracy as well as the processing time. Fig 3(d),(e) shows the effect of the estimated camera pose by augmentating a virtual object (yellow teapot) into the depth images along with the camera movement.

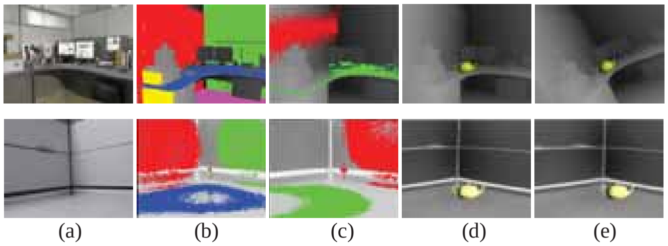


Fig 3 Test scenes: (a) snapshot of the synthetic scene and real scene. (b) detected planar structures by our method, (c) detected planar structures by RANSAC, (d),(e) augmentation of a virtual object according to the camera motion.

IV. Conclusion

A depth camera becomes a powerful consumer electronic device and expected to be used widely in the everyday life. In this paper, we present a novel approach to estimate a pose of the depth camera using a new type of a feature, i.e. plane features. The result also can be combined with the color image for the better performance.

References

[1] J. Shotton et al, “Real-time human pose recognition in parts from single depth images,” *IEEE Computer Vision and Pattern Recognition,* 2011.

[2] T. Tuytelaars et al, “Local invariant feature detectors,” *Journal of Foundations and Trends in Computer Graphics and Vision*, 3(3), 2008.

[3] S.-M. Rhee et al, “Split a merge approach for detecting multiple planes in a depth image,” *IEEE Int’l Conference on Image Processing*, 2012.

[4] M. A. Fischler et al, “RANSAC: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *ACM Communications*, 24(6):381–395, 1981.

[5] C.R. Rao, H. Toutenburg, A. Fieger, C. Heumann, T. Nittner and S. Scheid, “Linear models: Least squares and alternatives,” *Springer Series in Statistics.*

[6] A. C. Cameron and F. A. G Windmeijer, “An r-squared measure of goodness of fit for some common nonlinear regression models,” Econometrics, 77(2): 329–324, 1997.

[7] Z. Zhang, “A flexible new technique for camera,” *Technical Report MSR-TR-98-71*, 1998.

[8] The Mobile Robot Programming Toolkit, http://www.mrpt.org/.

Interpolation Method for ToF Depth Sensor with Pseudo 4-tap Pixel Architecture

Tae-Chan Kim, Kwanghyuk Bae, Kyu-Min Kyung, and Shung Han Cho
Samsung Electronics Co., Ltd.

Abstract—**This paper presents an interpolation method for ToF depth sensor with pseudo 4-tap pixel architecture. Pseudo 4-tap pixel architecture uses different modulation signals for different row pixels. While faster data acquisition reduces motion artifacts, depth calculation with two vertical pixels lowers depth resolution. The proposed method uses offset values for similarity weights to improve depth resolution. Experimental results show that edge artifact is improved in the vertical direction.**

I. INTRODUCTION

Recently, depth information has been used in three-dimensional applications such as gesture recognition, console game or user interface for TV. Also, depth information can be acquired by various methods such as stereoscopic, structured light, and Time-of-Flight (ToF). As compared to structured light depth sensor, ToF depth sensor has a more cost-effective and smaller module size with accurate depth information.

Phase-shift depth measurement is one of the widely used ToF methods in consumer electronics. Phase-shift based ToF sensors have various pixel architectures. Pixel architectures are classified according to the number of taps that derive the phase of the incoming modulation signal in one pixel. There are three different pixel architectures based on the four-phase algorithm; 1-tap, 2-tap, and 4-tap. 1-tap pixel architecture has the advantages of a small pixel structure and a large fill factor compared to the others. However, it has some drawbacks such as motion artifacts caused by the required four frames for one depth map and high depth errors caused by changing light condition during acquisition. On the other hand, 4-tap pixel architecture is less sensitive to motion and changing light condition but it has a large pixel structure and a small fill factor [1][2]. 2-tap architecture is generally preferred due to the middle characteristics, but motion artifacts are still remained because of two frames per one depth map.

Ovsiannikov *et al*. proposed the architecture acquiring four correlation measurements simultaneously with 2-tap [3]. In this paper, we refer to this architecture as the pseudo 4-tap pixel architecture for convenience. Pseudo 4-tap pixel architecture reduces motion artifacts but has low spatial resolution in the vertical direction.

This paper proposes an interpolation method in ToF sensor with pseudo 4-tap pixel architecture. In order to improve the spatial resolution in the vertical direction, the proposed method uses offset value to derive similarity weights in calculating depth data. The experimental results show that edge artifact is improved in the vertical direction.

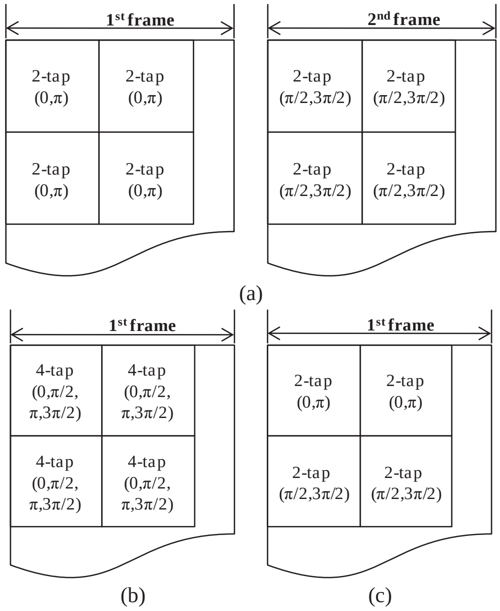
II. INTERPOLATION METHOD

A. Pseudo 4-tap Architecture

2-tap pixel architecture is able to acquire two correlation measurements simultaneously that are π radian out of phase as shown in Fig. 1(a). In the next frame, the other two measurements are obtained by shifting the phase of the sensor modulation signal by π/2 radian. Therefore, 2-tap needs two frames in order to compute one depth map.

4-tap pixel architecture acquires four-phase data simultaneously and one frame is used to calculate one depth map as shown in Fig. 1(b). 4-tap pixel design increases the complexity of each pixel and reduces a fill factor.

Ovsiannikov *et al*. proposed pseudo 4-tap architecture of ToF sensor in [3] and it acquires four-phase data using one frame with 2-tap pixel architecture. Four correlation measurements are obtained by shifting the phase of the sensor modulation signal by π/2 radian in each odd and even row. Motion artifacts are significantly reduced because pseudo 4-tap can compute depth in shorter time. However, spatial resolution in the vertical direction is reduced because one depth calculation needs two pixels in the vertical direction.



(a)

(b) (c)

Fig. 1. Tap architecture; (a) 2-tap, (b) 4-tap, (c) pseudo 4-tap

B. Interpolation Using Weights Calculated from Offset Data

An interpolation method is needed in pseudo 4-tap pixel architecture to improve the spatial resolution in the vertical ddirection. Nearest neighbor method and bilinear interpolation method distort depth data in the boundary region of objects. Assuming that four correlation measurements are $P_{odd\ row,0}$, $P_{odd\ row,\pi}$, $P_{even\ row,\pi/2}$, and $P_{even\ row,3\pi/2}$ in equations (1)~(4). Fault phase ($\tilde{\theta}$) is calculated because measurements are combined from spatially different regions in the boundary, and *sin* and *cos* terms are calculated from different spatial positions. To solve this problem, we propose phase calculation with similarity weights from offset value *B* which is a sum of offsets from background (ambient) light and illumination source. Offset value *B* is not lost in the spatial domain and offset value at each pixel is calculated by simple addition, equation (6).

$$P_{odd\,row,0} = A'\cos\theta' + B' \tag{1}$$

$$P_{odd\,row,\pi} = -A'\cos\theta' + B' \tag{2}$$

$$P_{even\,row,\pi/2} = A''\sin\theta'' + B'' \tag{3}$$

$$P_{even\,row,3\pi/2} = -A''\sin\theta'' + B'' \tag{4}$$

$$\tilde{\theta} = \tan^{-1}\left(\frac{P_{even\,row,3\pi/2} - P_{even\,row,3\pi/2}}{P_{odd\,row,0} - P_{odd\,row,\pi}}\right) = \tan^{-1}\left(\frac{A''\sin\theta''}{A'\cos\theta'}\right) \tag{5}$$

$$B = P_{0\,or\,\pi/2} + P_{\pi\,or\,3\pi/2} \tag{6}$$

We use a bilateral filter in the interpolation of correlation measurements before depth calculation. A spatial filter is applied to the correlation measurements *P* and a similarity filter is jointly applied based on the offset value *B*. Let Ω denotes odd or even rows depending on current correlation position. f and g are Gaussian function, and h is $1-e^{-|ax|}$ where a is weight of Laplace function. $\hat{\theta}$ is phase calculated by bilinear interpolation. Interpolated correlation measurements *P* is then obtained as:

$$\hat{P} = \frac{1}{k}\sum_{j\in\Omega} P_j\, f(i-j)\, g\left(B_i - B_j\right) h\left(\hat{\theta}_i - \hat{\theta}_j\right) \tag{7}$$

Weights for the filter consist of position similarity, offset similarity, and phase similarity.

III. EXPERIMENTAL RESULTS

For comparison, 2-tap and pseudo 4-tap data are generated from the synthesized “Teacup” depth and image [4]. Offset data is generated from gray image, and amplitude data is generated from both offset data and depth map. Finally, we calculate correlation data using equations (1)~(4). Fig. 2(a) and (b) show the enlarged part of generated 2-tap data and pseudo 4-tap data respectively. Fig. 2(b) shows that the spatial resolution is preserved in the offset data calculated by equation (6). The result of the proposed method is compared with the results of 2-tap, nearest neighborhood, and bilinear. The Results of nearest neighborhood and bilinear methods in pseudo 4-tap data show artifacts in edge region while the proposed method preserves edge data.

IV. CONCLUSIONS

This paper presents an interpolation method for pseudo 4-tap pixel architecture. Pseudo 4-tap pixel architecture has many advantages in moving conditions and under ambient light, but needs an interpolation method for depth calculation and compensation for low spatial resolution. Proposed method improves the spatial resolution using the offset similarity in the bilateral filter. With the proposed method, pseudo 4-tap can be widely used for gesture recognition applications.

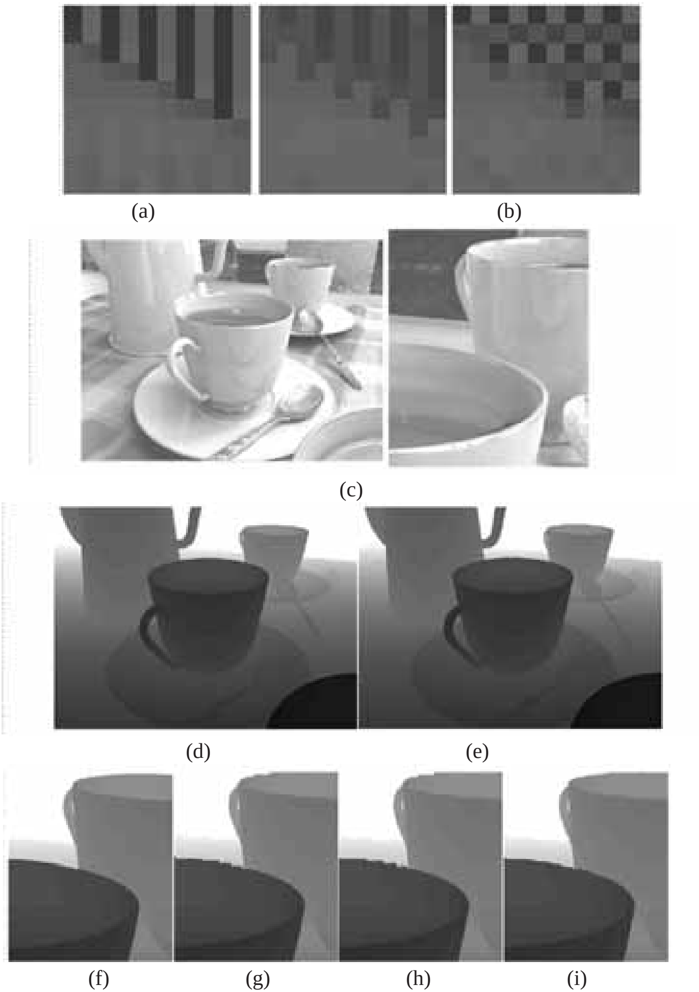
REFERENCES

[1] R. Lange, "3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology," *Ph.D. dissertation,* University of Siegen, Germany, 2000.

[2] R. Kaufmann, M. Lehmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, T. Oggier, N. Blanc, P. Seitz, G. Gruener, and U. Zbinden, "A Time-of-Flight Line Sensor – Development and Application," *Proceedings of the SPIE, Volume 5459*, pp. 192-199, 2004.

[3] I. Ovsiannikov, Y. Park, D. Min, and Y. Jin, "Image Sensors for Sensing Object Distance Information," U.S. Application Patent 2011/0129123 A1

[4] dofpro.com/cgigallery.htm


(a) (b)

(c)

(d) (e)

(f) (g) (h) (i)

Fig. 2. Experimental results; (a) 2-tap phase, 1st frame and 2nd frame, (b) pseudo 4-tap phase, (c) offset data from pseudo 4-tap, (d) full sized depth map from 2-tap, (e) full sized depth map calculated by proposed method from pseudo 4-tap, (f) depth form 2-tap, (g) depth calculated by nearest neighborhood from pseudo 4-tap, (h) depth calculated by bilinear from pseudo 4-tap, (i) depth calculated by proposed method from pseudo 4-tap

How to Extend Battery Life in Location Alarm

Jeong-Gwan Kang, Namhoon Kim, Sungmin Park, Kyongha Park, HyunSu Hong
Samsung Electronics, Suwon Korea

***Abstract*-- Location Alarm application notifying user of an alarm at the locations where user registered events causes serious battery drain because the current location should be continuously monitored with turning on the GPS or WPS. This paper proposes a scheme to extend battery life that minimizes the access to the GPS or WPS considering the user's moving state recognized by various sensors in mobile device.**

I. Introduction

Location-based alarm is also getting in demand with the smart devices containing the locating functions like GPS or Wi-Fi since the spatial information is so important factor in people's life. Some location alarm applications can be downloaded from the market (or store), but these applications may usually drain battery life [1]. The main reason for the battery drain is that the getting the user's location information consumes battery power significantly and that most of location alarm applications keep tracing the user's location until the user arrives at the target location.

The battery life can be extended by reducing the access to the GPS or Wi-Fi modules while keeping the accuracy of location as one of service requirements. Existing approaches introduce the dedicated server into the alarm service to handle this problem [2, 3, 4]. But, it requires more cost to build and manage the server system, which also raises the privacy problem because the user's private information is collected by the server.

In this paper, we proposed a solution that can save the power consumption without any server system by controlling the frequency of accessing to locating modules adaptively depending on user's moving state recognized by the sensors in a mobile device.

II. Algorithm

The proposed algorithm achieves the goal with steps as followings – 1) to determine the inactive period for which the location alarm application need not to check the location, 2) to let the device go to sleep for the calculated inactive period, 3) to wake up the device and repeat step 1) and 2) when the inactive period elapsed or when moving situation is changed.

The inactive period in step 1) is determined with the user's current moving state and distances from the current location to the target locations on which user registered events. Only the accurate inactive period can guarantee the minimum power consumption and the precise location alarm.

A. *Determining the Inactive Period*

When the user sets the target position for location alarm, the application requests the current location of device from GPS or WPS, and calculates the distance from the current location to the target location. The inactive period is determined as followings:

$$P_i = SF_{type} \frac{d(L_{user}, L_i)}{v(M_{type})} \tag{1}$$

where

P_i : Inactive period for i^{th} location alarm

L_{user} : User's current location

L_i : i^{th} target location for the location alarm

M_{type} : Moving state type such as walk, bicycle, or vehicle.

$d(A,B)$: Distance between location A and B

$v(C)$: Nominal speed of the moving state type C

SF_{type} : Scale factor

Finally, the wake-up time on which the device should wake up and update the inactive period is determined as followings:

$$W = T + \min\{P_i\} \tag{2}$$

where

W : Wake-up time

T : Current time.

B. *Updating the Inactive Period*

The device wakes up and updates the inactive period when the wake-up time reaches or when the user's moving state is changed. Followings are the representative cases with the detailed update procedures.

1) *No transition Case in moving state*

If there is no change in moving state during the inactive period, the device wakes up at the wake-up time determined previously and checks whether the current location satisfies the nearby threshold of the target location or not. If not, the application updates the wake-up time with (1) and (2), and then lets the device go to sleep again.

2) *Transition Case in moving state*

When a change in the moving state is recognized, the device wakes up, and the inactive period is updated to the target location with a newly recognized moving state. The 3 kinds of moving states are considered in our verification for this algorithm; walk, bicycle, and vehicle.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

3) *No movement Case*

If there is no meaningful change in the user location after the inactive period sets or updates, at the time, the application updates the period by the previous value without calculation, and lets the device go back to sleep.

III. Performance Evaluation

In order to study the performance of our scheme, in terms of success rate and the number of GPS access, simulations were performed using MATLAB. The success of location alarm service means that the alarm notification is given when the user arrives at nearby the target location (< 100 meters). The number of GPS access is a touchstone of battery consumption. Usually, one GPS access consumes 60mA battery. The scale factor, SF_{type}, is set to 0.6, which is determined through our simulation results. The proposed scheme was compared against the periodically location alarm, which has the period time from 1 minute to 50 minutes.

In our simulation, we assume the device registered 100 different target locations for the location alarm. And we assume three kinds of the user's moving state type: walk, bicycle, and vehicle. For computing inactive period, the reference speed of each moving state type is set to 5, 15 and 50Km/h, and the actual moving speed is randomly determined within 30% margin of reference speed.

A. *Success Rate*

Fig. 1 shows the alarm success rate of the proposed scheme and the periodically location alarm. The results of proposed scheme show almost 100% regardless of moving state type, while those of periodically location alarm are getting decreased by increasing moving speed and the period time. Therefore, it is required that the period is set to less than 1 minute in order to get more than 80% success rate in periodically location alarm service. But, if the period is set to less than 1 minute then it causes the battery consuming problem.

B. *The number of GPS access*

Fig.2 shows the number of GPS access of the proposed scheme and the periodically location alarm. As we mentioned before, the number of GPS access is directly proportional to the power consumption. Our scheme did access about 7 times to GPS until location alarm service was given, while the periodically location alarm which was set to 1 minute as the period time did access about 360 times to GPS. In order words, the periodically alarm incurs more battery power consumption to get the precise location alarm service.

Our scheme requires a fewer number of GPS access as much as the alarm which the period time is set to 50 minutes, while it guarantees the almost 100% success rate.

IV. Conclusion

This paper presents an energy saving location alarm scheme by converting spatial distance to the inactive period. When the target position is registered for the location alarm service by user, the device determines the inactive period that it is not necessary to check the current location with considering user's moving state. Then, if the inactive period ends, the device checks whether the user is nearby the target position or not, instead of periodically checking it. In conclusion, the proposed scheme makes mobile device provide location alarm with minimum power consumption and maximum alarm accuracy.

In order to recognize a moving-state at any time even when the device is in sleep state, the recognizer module should be always driven on the microprocessor or a dedicated core which is independent from the main mobile device's processor. The microprocessor manages the sensors with little battery power so that the recognizer can wake up the device.

References

[1] Apple, Optimize your settings. http://www.apple.com/batteries/iphone.html.

[2] B. Bamba, L. Liu, A. Iyengar, and P. S. Yu. Distributed processing of spatial alarms: A safe region-based approach. In ICDCS, 2009.

[3] M. Doo, L. Liu, N. Narasimhan, and V. Vasudevan. Efficient indexing structure for scalable processing of spatial alarms. In GIS, 2010

[4] B. Bamba, L. Liu, P. S. Yu, G. Zhang, and M. Doo. Scalable processing of spatial alarms. In HiPC, 2008.

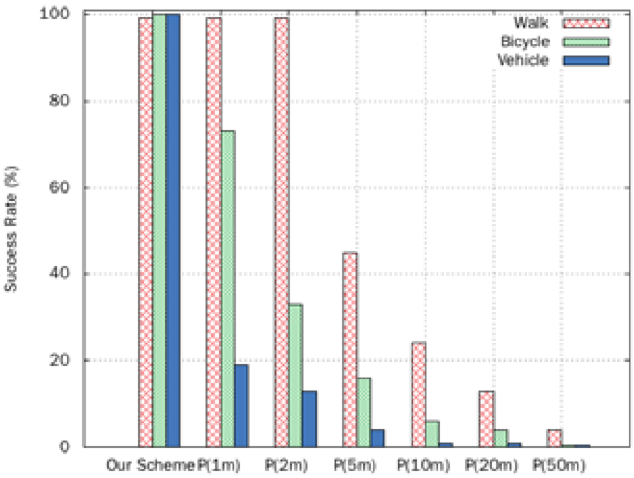


Fig. 1 Success Rate

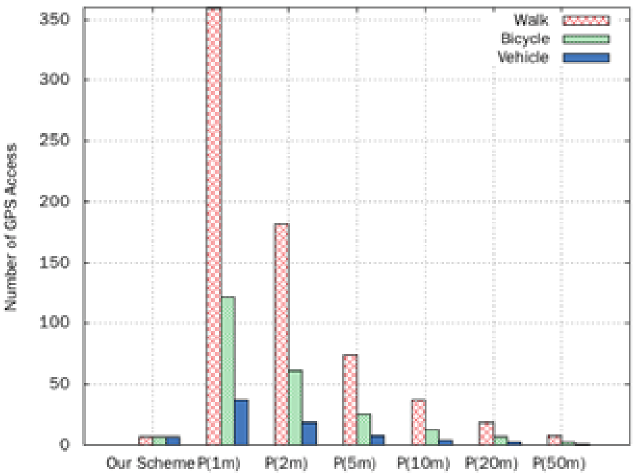


Fig. 2 The number of GPS access

Vision Based Motion Estimation Method using Ego-Exo Cameras

Taeyoung Uhm*, Ji-In. Jun*, and Jong-Il Park*[a], Member, IEEE
*Hanyang University, Seoul, Korea

Abstract—**Recently, vision based robust pose estimation method for human computer interaction has been developed by many researchers. It plays an important role in the development of interaction in consumer electronics. Existing pose estimation using monocular camera employed either ego-motion or exo-motion which are not acceptable for fine motion (e.g. a breath, wrist motion, and etc.). In this paper, we propose a hybrid vision method for fine motion estimation. For this method, we integrated ego-camera and exo-camera coordinates. We expect that the proposed method can provide a practical solution for robust natural interaction with digital information display systems and interactive serious game or remote rehabilitation care systems.**

I. INTRODUCTION

The camera pose estimation for motion tracking has been an important field of research for vision–based interface [1]. Especially, single- or multi-camera-based pose estimation method has been used in consumer electronics for achieving natural human computer interaction. The pose estimation method is roughly divided into two categories: ego-camera and exo-camera pose estimation. First, pose estimation using ego-camera is mainly employed for object self-motion (e.g. robot, vehicle). However, this method has the ambiguity of its own for fine motion. In other words, it is difficult to know whether it is translation or orientation when the movement is small. Second, exo-camera-based pose estimation by fixed external observation cameras is popular. However, it is very difficult to detect and estimate small rotations of target objects. Moreover, accurate pose estimation often requires optimization using non-linear equations [2]. In this paper, we propose a hybrid pose estimation method which takes advantages of ego-camera and exo-camera systems. Furthermore, there is no need to use non-linear equations to allow for more accuracy. The hybrid pose estimation method first estimates the exact position of ego-camera from two or more exo-cameras and then accurately estimates the fine rotation of ego-camera using the position with simple computation. Therefore, the method is well-suited for applications that require fine rotation and position in the user's movement (e.g. functional exercise therapy for rehabilitation game. See Fig. 1).

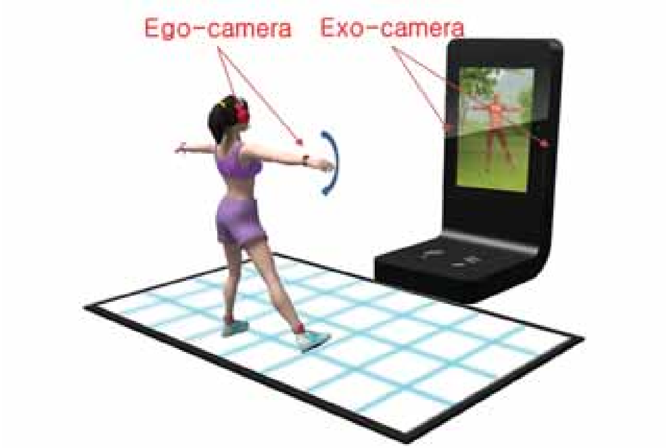


Fig. 1. Example of requiring fine motion estimation: physical therapy games, including a wrist movement.

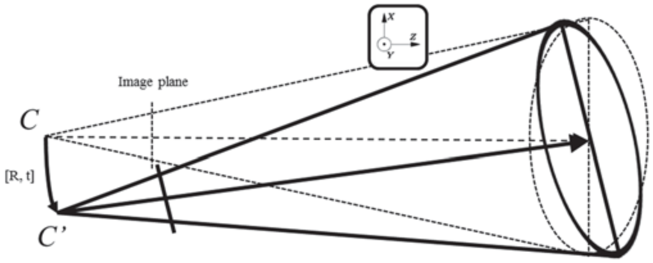


Fig. 2. Orientation and position model from the camera motion.

II. POSE AMBIGUITY & GEOMETRIC INTERPRETATION

The camera pose estimation with a calibrated camera is a problem of finding the six external parameters of the camera motion. Fig. 2 shows the external parameters which defined orientation $R = [R_1, R_2, R_3]^T$ and position $t = [t_x, t_y, t_z]^T$ of the camera with respect to a scene coordinate system. Without loss of generality, we assume n coplanar model $p_{(eg)i} = [p_{i_x}, p_{i_y}, 0]^T$ in scene coordinates which are transformed to camera coordinates $v_{(eg)i}$ by

$$v_{(eg)i} \propto R_{eg}\, p_{(eg)i} + t_{eg}. \tag{1}$$

where $\propto$ denotes "directly proportional" [2]. For ego-camera pose estimation, we employed a general marker based pose estimation algorithm [3].

On the other hands, exo-camera motion based on stereo vision method can extract $p_{(ex)i} = [p_{i_x}, p_{i_y}, 0]^T$ in scene coordinates which are transformed to camera coordinates

This work was supported by Seoul Education Program for Creative Human Development (HM120006).

a) Corresponding author: Jong-Il Park, Hanyang University, Korea. Email: jipark@hanyang.ac.kr

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$v_{(ex)i}$ by

$$v_{(ex)i} \propto R_{ex} p_{(ex)i} + t_{ex}. \qquad (2)$$

These external parameters corresponding to the ego-motion parameters are more reliable as shown in Fig. 3. Then, the value of orientation which is based on the reliable value of position from exo-motion is corrected. This hybrid method enables more accurate estimation of the fine motion than a single ego-motion-induced estimation.

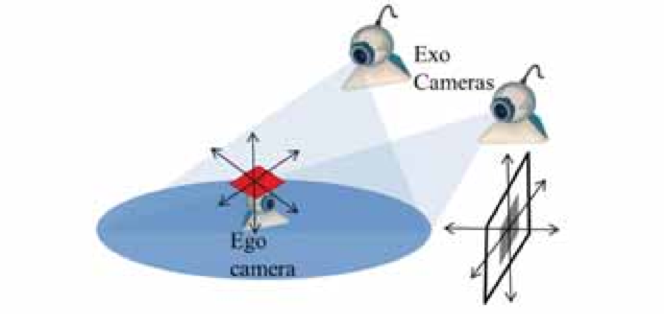


Fig. 3. The relation between ego-camera and exo-cameras.

III. ROBUST POSE ESTIMATION METHOD

The novel estimation method consists of ego-camera and exo-cameras. First, ego-camera based projective transformation matrix P is defined as follows,

$$\mathbf{P} = K[R_{eg}, t_{eg}] \rightarrow K[R', t_{ex}]. \qquad (3)$$

Here, K is internal matrix and R' indicates the value of new orientation based on the position value from exo-camera. For ego-camera-based estimation, if we use the position, the formulation is changed as follows:

$$p_{eg} = K[R_{eg} \mid t_{eg}]P_{eg} \rightarrow p_{eg} = K[R' \mid t_{ex}]P_{eg}. \qquad (4)$$

Finally, we calculate the orientation matrix R' from eq. (4), which can be extracted using 3 correspondences from ego-camera coordinate as,

$$R' : K^{-1} p_{(eg)i} = [R' \mid t_{ex}]P_{(eg)i}, \; i = \{1,2,3\}. \qquad (5)$$

IV. EXPERIMENTAL RESULTS

We demonstrate preliminary results using one ego-camera(320x240) and two exo-cameras(1280x720) and markers. The ego- and exo-cameras are positioned 0.8m away facing each other. The camera and marker are separated by 0.85m. Figure 4 shows the results of accuracy. The error tendencies between ego-camera only and our proposed method depend on the size of the markers (66mm, 90mm and 150mm) as shown in Fig. 5. The proposed method clearly shows less error especially when the size of marker is small, i.e. the ambiguity of rotation is big.

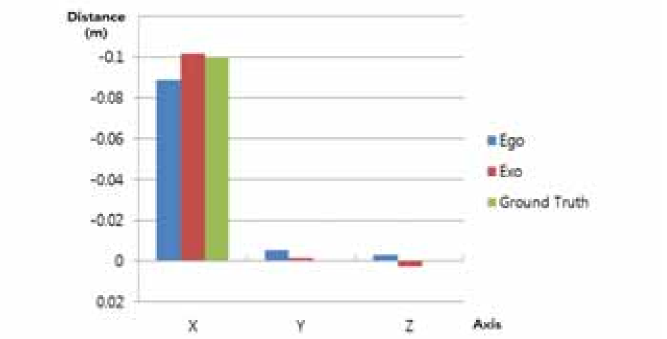


Fig. 4. The value of position from ego- and exo-cameras. (The case of movement in the direction of the X-axis by -0.1m).

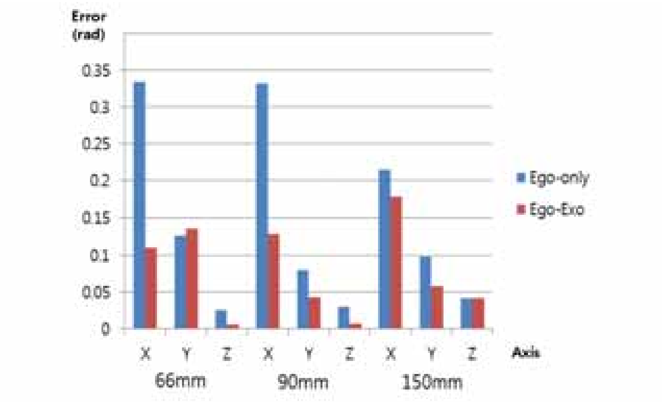


Fig. 5. The error comparison between ego-camera only and hybrid cameras (ego- and exo- cameras). (The case of movement in the direction of the X-axis by -0.1m)

V. CONCLUSION

In this paper, we propose a hybrid pose estimation based on ego- and exo-cameras without the complex computation of non-linear error minimization. The advantage of the hybrid method is clear and we expect the method can be applied to various human-computer interaction scenarios that require fine and accurate motion estimation.

In the future, non-marker-based motion estimation method based on optical flow methods is to be applied. In addition, we are doing rigorous performance analysis using Cramer-Rao Lower Bound

REFERENCE

[1] Microsoft Corp. Redmond WA. KinectforXbox360.

[2] S. Gerald and P. Axel, "Robust Pose Estimation from a Planar Target," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pp. 2024-2030, 2006.

[3] L. Davis, and E. Clarkson, "Predicting Accuracy in Pose Estimation for Marker-based Tracking,," *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)*. pp. 01-08, 2003.

Cost-Effective Rasterization Using Valid Screen Space Region

Yeong-Kang Lai, *Member, IEEE* *and* Yu-Chieh Chung, *Student Member, IEEE*

***Abstract*-- Due to the progress of consumer electronics, 3D-Graphics system has become mobilizing and attractive. In order to render 3D graphic in efficiency, the rasterization techniques are developed. Traditional clipping techniques using the six-planes of view volume to split outside part of primitive are complicated and not cost-effective. For 3D graphics gaming application with high resolution in mobile device, the cost-effective hardware is crucial and necessary. This paper proposes a novel cost-effective strategy for primitives clipping in rasterization. In the entire process, no expensive clipping action is involved and no extra clipping-derived polygons are produced. The proposed architecture which processes the valid screen space region of each primitive in 8 cycles and the gate-count is 20k only, using the TSMC 65nm 1P9M process and the throughput reach 25 M Triangles/Sec.**

I. INTRODUCTION

In recent years, 3D graphics applications have become increasingly popular for consumer electronics, particularly, the market for 3D graphics gaming applications on mobile devices. Especially, 3D graphics-intensive applications are predicted to become widely available on a variety of portable mobile devices ranging from Tablets to Notes to Smart Phones.

These features intend multi-core SOCs, higher resolution, and lower power requirement to conserve battery life. Consequently, low power design is the most important strategy in order to become competitively portable mobile consumer electronics. Clipping is an important task in projection process but leads to much computation overhead. A lot of research efforts have been spent for promoting clipping performance.

In our proposed algorithm, there have neither any of pre-clip process nor clipping on any of view volume plane when doing the converts perspective frustum to a cube, canonical view volume. Instead, we use the valid 2D homogeneous region for the vertex's Z outside the view volume concept to project the valid screen space region then after view port transform with simple depth test & scissor test to render out the final region of primitive.

II. PROPOSED ALGORITHM

The proposed algorithm in rasterizer used of two parts: build valid 2D screen space region and revise the parameter for edge equation based tile-traversal. We adopt new clip-less method in 3D graphic pipeline of primitive clipping, the concept is from the top view of normalize the viewing frustum to a cube, canonical view volume & Marc Olano's[1] model. Fig.1 shows the top view of model define the near and far planes to be at z = f0 = − near and z = f1 = − far, respectively. The points at infinity are normalized and mapped to the plane z = f0 + f1. Notice in particular that the region labeled JKL on the left is normalized and mapped to the regions LKJ on the right, and this will help the basic concept of doing linear extrapolation for our new clip-less edge equation. Due to the traditional process of sequential parsing six planes of view volume clipping, it causes the challenges to implement clipping function in low-cost hardware design. When one or two vertices behind the camera, that even will cause flip projection and get the wrong region in screen space. To understand the connection between 2D homogeneous triangles and 3D triangles, we look at a single 3D triangle and its projection onto the screen from Marc Olano's model. To project the 3D point (X, Y, Z), set x = X, y = Y, and w = Z to get the 2D Homogenous point (x, y, w). We use these two cube space with normal mapping & project properties with near clipping plane to do linear extrapolation to find out valid region in screen space.

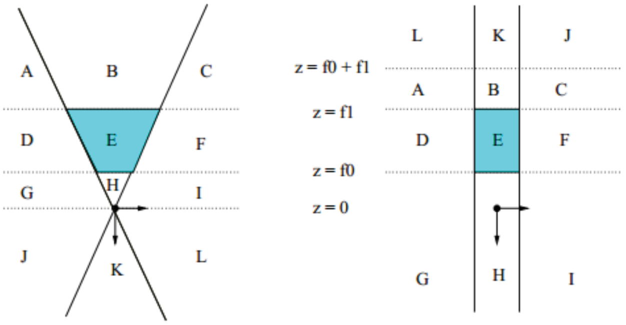


Fig.1 Top view of normalizing the viewing frustum to a cube and canonical view volume.

Thus, we build the clip-less process in screen space and notice only at the number of negative value w of the vertices. Before we got the valid region in screen space the next-step is to revise the sign of parameter for edge equation by which the opposite the vertex having a negative value w. Fig.2 shows that one vertex with Z <0 is mapped to screen space will get one vertex with W<0 & its valid region with red color. In Fig.4, the screen space shows that When A (w) <0, we use of the original vertex A relative to the vertex BC to do linear extrapolation and to find new effective vertex B 'C' in the near plane, this B-C-C'-B' is the valid region of this primitive.

In the other case, the Fig.3 shows that two vertices with Z <0 are normalized and mapped to screen space will get two vertices W<0 & its valid region with red color. Fig.5 shows that when both B (w) <0 & C (w) <0 , we use of the original vertex BC relative to the vertex A to do linear extrapolation and to find new effective vertex B 'C' in the near plane, this A-B'-C‘ is the valid region of this primitive. We show the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

dataflow of the proposed algorithm in Fig.6 .

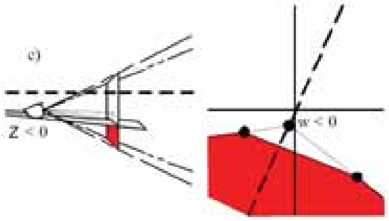


Fig.2 Canonical eye space: one vertex with Z <0 is normalized and mapped to screen space: one vertex with W<0 & its valid region with red color.

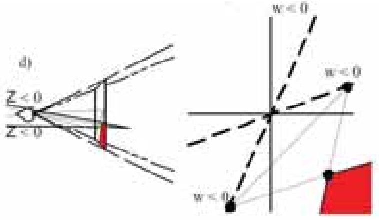


Fig. 3 Canonical eye space: two vertices with Z <0 are normalized and mapped to screen space: two vertices with W<0 & its valid region with red region.

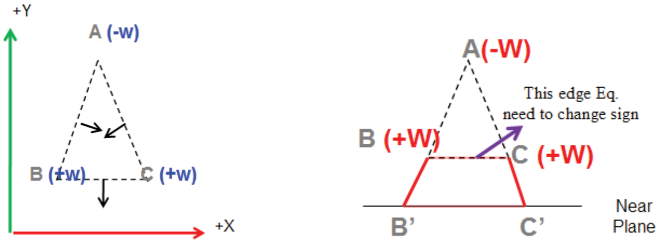


Fig.4 When A (w) <0 , we use of the original vertex A relative to the vertex BC to do linear extrapolation and to find effective vertex B 'C' in the near plane, this B-C-C'-B'is the valid region of this primitive

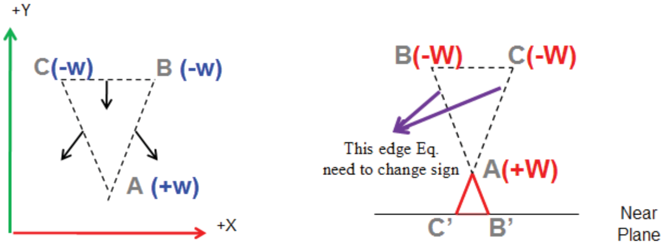


Fig.5 When B (w) <0 & C (w) <0 , we use of the original vertex BC relative to the vertex A to do linear extrapolation and to find effective vertex B 'C' in the near plane, this A-B'-C' is the valid region of this primitive

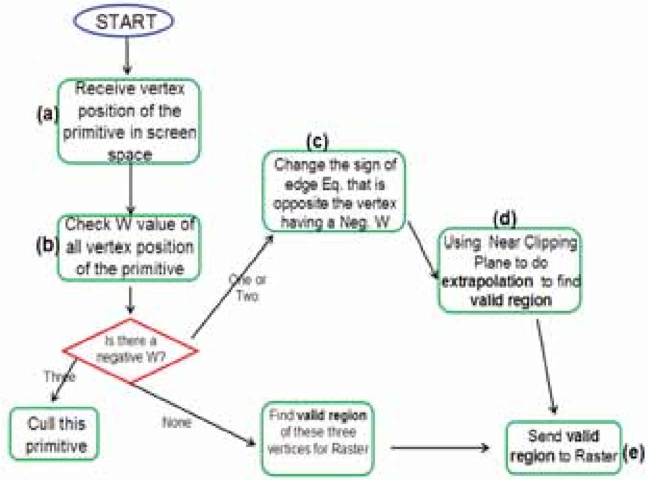

START
(a)
Receive vertex position of the primitive in screen space
(b)
Check W value of all vertex position of the primitive
Is there a negative W?
Three
Cull this primitive
None
One or Two
(c)
Change the sign of edge Eq. that is opposite the vertex having a Neg. W
(d)
Using Near Clipping Plane to do extrapolation to find valid region
Find valid region of these three vertices for Raster
Send valid region to Raster
(e)

Fig. 6. Dataflow of the proposed algorithm.

TABLE I
PERFORMANCE,THROUGHPUT & GATE COUNTS COMPARISION

Clipping Engine	Hardware Architecture	Area (gate counts)	Support Clip Plane	Throughput (M/Triangles/sec)	Process /Cycle time
[3]	Single Path	155.89k	X,Y,Z Planes	11	0.13um /6ns
[4]	Dual Path	79k+3kB SRAM	Z Planes	none	0.18um /6ns
[5]	Single Path	213.2k	X,Y,Z Planes	none	0.18um /6ns
[6]	Dual Path	165.17k	X,Y,Z Planes	10.9	90nm/6 ns
This work	Finding valid region in screen space	20k	X,Y,Z Planes	25	65nm/5 ns

III. EXPERIMENTAL RESULTS

In order to propose a cost-effective strategy for primitives clipping in rasterization, we used the clip-less algorithm and move the clipping stage to screen space in graphics pipeline.

The proposed architecture which processes the valid screen space region of each primitive in 8 cycles and the gate counts is 20k only, using the TSMC 65nm 1P9M process and the throughput reach 25 M Triangles/Sec. This proposed algorithm can parallel implement with edge equation parameter building stage and suitable for cost-effective in hardware design, thus called the novel clip-less algorithm.

IV. CONCLUSION

Basic polygon cutting (Primitives Clipping) in the entire 3D graphics pipeline (Rendering Pipeline) is an important step in the traditional polygon cutting. The cost of the hardware takes substantial increase and not suitable for low-power mobility GPU design. This paper proposes a novel cost-effective strategy for primitives clipping in rasterization. In the entire process, no expensive clipping action is involved and no extra clipping-derived polygons are produced. The implementation cost is hence tremendously reduced.

REFERENCES

[1] Marc Olano, Trey Greer, "Triangle Scan Conversion using 2D Homogeneous Coordinates", HWWS '97 Proceedings of the ACM SIGGRAPH/EUROGRAPHICS workshop on Graphics hardware.

[2] S Laine, T Aila, T Karras and J Lehtinen "Clipless Dual-Space Bounds for Fast Stochastic Rasterization" ACM Transactions on Graphics, 30(4) ,2011 (SIGGRAPH 2011).

[3] J. Bae, D. Kim, and L.-S. Kim, "An 11M-Triangles/sec 3D Graphics Clipping Engine for Triangle Primitives", Proc. IEEE Intl. Symp on Circuits and Systems(ISCAS), Vol.5 , May 2005.

[4] J.-H. Kim, et al., "Clipping-Ratio-Independent 3D Graphics Clipping Engine by Dual-Thread Algorithm", Proc. IEEE Intl. Symp on Circuits and Systems(ISCAS), May 2008.

[5] S.-F Hsiao and T.-C Tien, "Hardware Design and Verification of Clipping Algorithm in 3D Graphics Geometry Engine", National Sun-Yet San Univerdity, July 2008.

[6] K.-H Lin "Design of an Efficient Clipping Engine for OpenGL-ES 2.0 Vertex Shaders in 3D Graphics Systems", National Sun-Yet San Univerdity, Aug 2009.

Quantization Parameter Control At Prediction Restriction for Fast Multistream Joiner in Multi-vision System

Naofumi Uchihara and Hiroyuki Kasai, The University of Electro-Communications, Japan

Abstract—Multi-vision systems and panoramic video systems are expected to produce a new multimedia paradigm. Our previously presented multi-vision system allows many users to view multiple videos by virtue of fast stream joining algorithms. However image degradation occurs because these algorithms need restricted constant coding parameters. As described herein, we propose a QP control scheme to decrease this degradation without sacrificing high speed in the joiner.

I. Introduction

Highly efficient encoding techniques used for high-definition video and transcoding techniques for various terminals have been studied for decades. These trends are anticipating a new paradigm of video delivery services. Multi-vision systems and panoramic video systems are expected to form the nexus of that new paradigm. Kimata et al. [1] proposed a panoramic video scheme using multiple tile videos: small partitioned video images of one video. These technologies enable users to access any view area in an entire video. Nevertheless, such methods do not accommodate complexity on the client side such as handling of multiple sessions, decoding of multiple streams, and synchronous rendering of images. We have proposed and implemented a system that enables viewing of multi-vision video at any view area at multiple resolutions [2]. We also proposed an encoding and joining technique to eliminate all decoding processes and bit-shift operations in the joining phase for much faster joining [3]. However these techniques engender image degradation caused by fixed coding parameters. We assume that the output bit stream after the joining process (called a joined stream) complies with the H.264/AVC baseline profile.

II. Fast Stream Joiner For Multi-Vision System

A. Tile Stream Joiner and Performance Problem [2]

This section briefly introduces a conventional tile stream encoding and joining method [2] for use in a multi-vision system. The tile stream joining process is to combine two MB-lines of two vertically and horizontally adjacent tile streams. However, simple joining causes inconsistent prediction on the boundary of the two streams, which leads to image quality degradation. Such prediction includes the number of non-zero coefficients (called TotalCoeff) prediction from neighbor blocks. We resolved this problem using Context-Adaptive Variable-Length Coding (CAVLC) re-encoding [2]. Nevertheless, these process loads render this system infeasible for providing multi-vision to many users in a practical environment.

B. TotalCoeff Fixing for Prediction Inconsistency and Issue

CAVLC uses TotalCoeff of the neighbor block (upper block and left block) as the CoeffToken table selection. CoeffToken is a pair of TotalCoeff and TrailingOnes (the number of coefficients of which the absolute value is equal to one), and a block coefficient coding parameter on H.264/AVC CAVLC. If a target block is located on the upper edge of tile stream in the encoding phase, then no block exists on that target block. A block might exist on that target block in the joining phase when another tile stream is joined above this tile stream. Consequently, the target block refers to a different CoeffToken table because the means of TotalCoeff of neighbor blocks differ in between the encoding and joining phase. For this problem, we proposed a TotalCoeff fixing approach that fixes the TotalCoeff at a certain value, X, of the blocks located on the right and bottom edge of each tile stream [3]. To adjust the original TotalCoeff Y to a predefined X, we eliminate the X-Y number of non-zero coefficients of the AC component when X > Y. When Y > X, then the Y-X number of coefficients which have value of “1” is newly added from the highest frequency.

III. Quantization parameter Selection for Error Reduction

Although this conventional fixing method achieved 80-times faster stream joining, it caused image degradation because some coefficients were dropped or added.

This paper presents a proposal of a new quantization parameter (QP) control scheme in the encoder to decrease image quality degradation in blocks where TotalCoeff is fixed. For this proposal, the encoder selects an appropriate QP at the MBs where TotalCoeff is fixed, and the joiner performs fast QP modification at the edge of the tile without bitwise copy operation. In the following sections, we investigate an error model for adjusting coefficients for an encoder.

A. Error Model in Adding Coefficients

As shown in (1), we consider the error model by assuming that the distribution of frequency domain coefficients complies with the Laplace distribution, instead of the Cauchy distribution [4] because it is easy to retrieve important parameters of the Laplace distribution.

$$f_X(x) = \frac{\lambda}{2}\exp(-\lambda|x|) \tag{1}$$

In the equation above, $\lambda=\sqrt{2}/\sigma$; σ^2 is the variance. From (1), the probability $P(iQ)$ that the coefficient is quantized to iQ ($i=\{0,\pm1,\pm2,\ldots\}$) is shown in (2), with Q as the quantization step size.

$$P(iQ) = \int_{(i-1/2)Q}^{(i+1/2)Q} f_X(x)\,dx = \begin{cases} \frac{1}{2}e^{-i\lambda Q}\sinh\left(\frac{\lambda Q}{2}\right) & \text{if i > 0} \\ 1-e^{-\frac{\lambda Q}{2}} & \text{if i = 0} \\ \frac{1}{2}e^{i\lambda Q}\sinh\left(\frac{\lambda Q}{2}\right) & \text{if i < 0} \end{cases} \tag{2}$$

From (2), the probability that TotalCoeff of a block is equal to *tc* is represented as shown below.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$P_{TC}(tc) = {}_{16}C_{tc} \times \{1 - P(0)\}^{tc} \times P(0)^{(16-tc)} \qquad (3)$$

The expectation of TotalCoeff is shown below.

$$E[tc] = \sum_{i=0}^{16} tc_i \times P_{TC}(tc_i) = 16 \times (1 - P(0)) \qquad (4)$$

Using (4), the number of coefficients to be added is regarded as follows. If *FTC* is larger than *E[tc]*, then the (*FTC-E[tc]*) coefficients with value of 1 are added to the target quantized block. These added coefficients are dequantized to “Q” after inverse quantization. Consequently, the error caused by adding (*FTC-E[tc]*) value “1”, D_{add}, is shown in equation (5).

$$D_{add} = \left(FTC - E[tc]\right) \times Q^2 / 16 \qquad (5)$$

B. Error Model in Deleting Coefficients

The absolute value of coefficients to be deleted and their number are unknown. Therefore, we first estimate them, assuming that the magnitude of coefficient values in the block decreases from low-frequency to high-frequency component.

The number of coefficients to be deleted can be estimated as (*FTC-E[tc]*) using (3). We consider the expected value of the number of occurrence times in 16 trials is represented as 16×P(iQ). Therefore, if the quantization step with index *i* equals ±1, then the expected value of the number of coefficients that have absolute value of 1 is 16×P(±Q). If the inequality above is not satisfied, then we obtain the coefficient with the absolute value of $R_{|1|}$, as expressed in the equation below. This concept is presented in Fig. 1. The X-axis shows the absolute value of the coefficient. The Y-axis shows the summation of the expected value. We obtain the coefficients to be deleted by summing the expected numbers of occurrences in 16 trials, as presented below.

$$R_{|1|} = \min\left\{ \; r \; \left| \left(\sum_{i=1}^{r} \{32 \times P(iQ)\} \right) > 1 \right. \right\} \qquad (6)$$

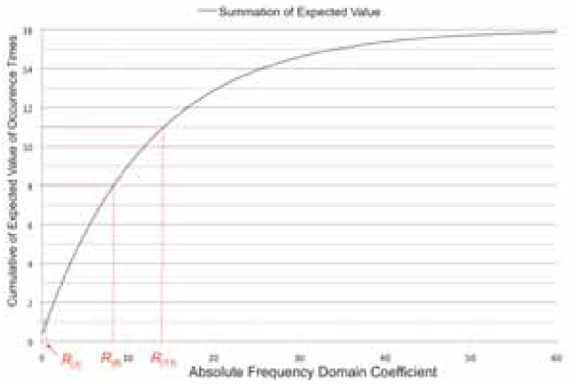


Fig.1 Concept of summation of occurrence times expectation

Similarly, $R_{|x|}$, the expected absolute value of the x_{th} coefficient, is calculated as shown below, provided that the range of *x* is *1≤x≤ (E[tc]-FTC).*

$$R_{|x|} = \min\left\{ \; r \; \left| \left(\sum_{i=1}^{r} \{32 \times P(iQ)\} \right) > x \right. \right\} \qquad (7)$$

Results show that the error caused by deleting the non-zero coefficient, $D_{decrease}$, is the following.

$$D_{decrease} = \sum_{i}^{E[tc]-FTC} \frac{(R_{|i|}Q)^2}{16} \qquad (8)$$

Equations (5) and (8) are useful to find QP adaptively to minimize D_{add} and $D_{decrease}$ in the encoder phase.

IV. QP Estimation Experiments

In this section, we evaluate the accuracy of the error model described from the viewpoint of MSE of the coded tile stream with selected QP. For the former evaluation, Mobile and Calendar with 352 [pel]×288 [line] was used for simple analysis.

The experimental conditions are that TotalCoeff of the rightmost blocks of the fifth MB column from the left edge is fixed. We present results of MSE of the target MB in Table I by comparing (i) when QP is constant, from 5 to 29, (ii) when QP is selected based on the proposed model. Gray cells in Table I show QP indicating the least MSE when *FTC* is 4–15. Therefore, if MSE in Proposal is close to MSE at the gray part, then we can say that the QP selection is ideal. From Table I, we understand that neither proposal always indicates a minimum MSE. However, the proposals can select QP adaptively that stably provides smaller MSE close to the minimum values, irrespective of any target TotalCoeff.

TABLE I
MSE with TotalCoeff fixing

TotalCoeff	4	7	10	13	15
QP=5	73.00	34.40	12.93	2.06	0.56
QP=8	72.50	33.20	12.97	2.05	0.73
QP=11	71.78	33.10	12.92	2.25	1.12
QP=14	72.69	33.86	13.15	2.91	2.08
QP=17	69.20	33.50	14.53	4.37	4.08
QP=20	67.64	33.38	16.18	8.24	8.71
QP=23	64.38	32.75	19.21	15.52	17.47
QP=26	70.12	42.66	27.92	32.69	36.74
QP=29	76.33	55.23	55.63	68.41	74.41
Proposal QP	68.68	30.23	13.43	2.03	0.38

VII. CONCLUSIONS

We proposed a QP control scheme that decreases image degradation caused by fixed TotalCoeff while maintaining higher speed in the stream joiner, which is effective for any fixing parameter and any sequences.

ACKNOWLEDGEMENT

This work is supported in part by KAKENHI 21760277, a Grant-in-Aid for Young Scientists (B).

References

[1] H. Kimata, S. Shimizu, Y. Kunita, M. Isogai and Y. Ohtani, “Panorama video coding for user-driven interactive video application,” *IEEE International Symposium on Consumer Electronics (ISCE2009)*, pp. 112–113, 2009.

[2] N. Uchihara and H. Kasai, “Fast H.264/AVC stream joiner for interactive free view-area multivision video,” *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1311-1319, August 2011.

[3] N. Uchihara and H. Kasai, “H.264/AVC encoding control for fast stream joiner in interactive multivision video,” *IEEE Transactions on Consumer Electronics, vol. 58, no. 3, pp. 1022-1030, August 2012*.

[4] Y. Altunbasak and N. Kamaci, “An analysis of the DCT coefficient distribution with the H.264 video coder,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’04)*, pp.17–21 vol. 3, pp. iii- 177–180, May 2004.

Face Recognition in Unconstrained Environments

Dong-Ju Kim, Sang-Heon Lee, Myoung-Kyu Sohn, Byungmin Kim, and Hyunduk Kim
Division of IT convergence, DGIST, Korea

Abstract--**This paper proposes a novel face recognition system having good recognition performance through elaborate face region detection in unconstrained illumination environments. The proposed system consists of face and eyes detection, rotated-angle compensation, face region cropping, preprocessing, and classification modules as sequential steps. In particular, the elaborate face region is obtained from automatic face cropping procedure based on distance information between the eyes. The performance evaluation was carried out with various pre-processing images on the Yale B and the CMU-PIE databases. From the experimental results, we confirmed that the proposed method showed the best recognition accuracy compared to different approaches.**

I. INTRODUCTION

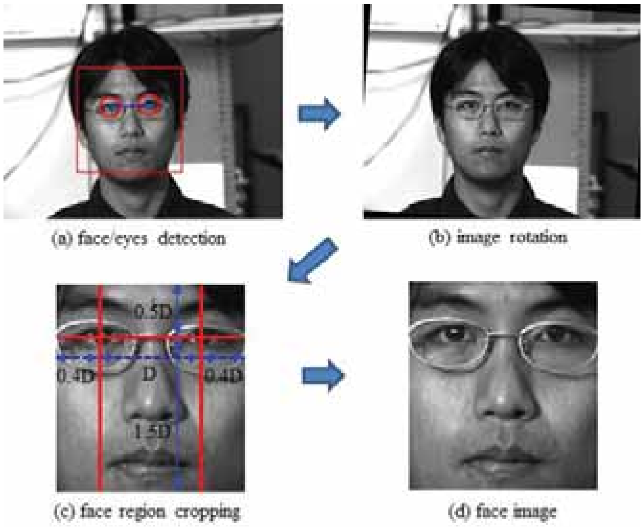
Face recognition has many applications such as biometrics systems, access control systems, surveillance systems, security systems, and content-based video retrieval systems [1]. Usually, face recognition systems can achieve good performance under controlled environments. However, face recognition tend to suffer when variations including illumination are presented in uncontrolled environments. Thus, this paper presents the novel face recognition system showing good performance through elaborate face region detection in unconstrained environments.

II. PROPOSED SYSTEM

This work has aimed to implement a novel face recognition system having good performance under unconstrained illumination environments. Generally, the most of traditional recognition approaches directly utilize a detected face image in the face recognition procedure. However, these approaches can be sometimes failed to recognize a user in practical environments, due to unelaborate detection of face region. To improve the face recognition performance on real environment, this paper proposes the novel face recognition system through elaborate face region detection.

A. Region detection and preprocessing

The first phase of proposed face recognition system is object detection including face and eyes region from the input image. To detect the face region, this paper employs the AdaBoost algorithm based on Haar-like features introduced by Viola and Jones [2]. After detecting the face and eyes region by using the AdaBoost algorithms based on Haar-like features (see Fig. 1 (a)), the coordinates of two eyes are utilized for rotated-angle compensation of the input image (see Fig. 1 (b)). In other words, the center coordinates of each eye can be obtained from the detection results, and we can calculate the rotated-angle with the horizontal line connecting both centers of eyes as show in Fig. 1 (a). The example of rotated image is also shown in Fig. 1 (b), and we can also know the converted center coordinates of both eyes on rotated image. Based on the center information of both eyes, we next compute the elaborate face region using image cropping as shown in Fig. 1 (c). Automatic image cropping of face region is done based on the distance, D, between two eyes. A distance between each eye and boundary is maintained as 0.4D. The height of the face region is set as 2.0D, here, the distance of from eye to bottom boundary is maintained as 1.5D. Finally, each face image was cropped by using boundary information and rescaled to a resolution of 60×54 pixels as shown in Fig. 1 (d).



(a) face/eyes detection (b) image rotation (c) face region cropping (d) face image

Fig. 1. Face acquisition procedure.

To improve recognition performance on unconstrained illumination environments, we carry out the preprocessing procedure, in which preprocessing denotes that the binary pattern operators including local binary pattern (LBP) [3] and local directional pattern (LDP) [4] are applied in face image. The LBP operator labels the pixels of an image by thresholding a 3x3 neighborhood of each pixel with the center value. While the LBP operator uses the information of intensity changes around pixels, LDP operator use the edge response values of neighborhood pixels and encode the image texture. This pattern is calculated by comparing the relative edge response values of a pixel by using Kirsch edge detector.

B. Classification

Automatic detected and preprocessed face images are

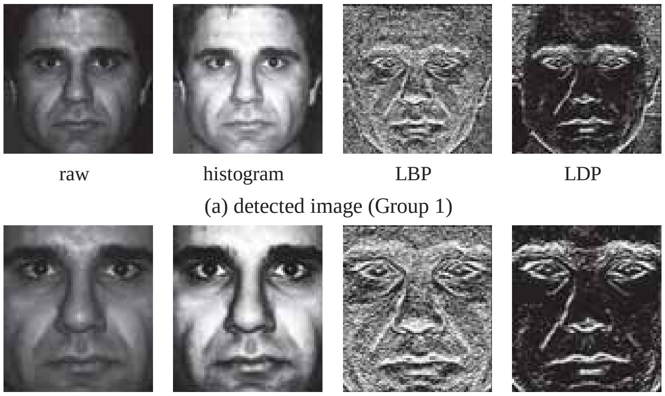
This work was supported by the DGIST R&D Program of the Ministry of Education, Science and Technology of Korea (12-IT-03).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

directly utilized as input images for facial feature extraction. In this work, we employ the well-known principal component analysis (PCA) [5] algorithm. PCA is a feature extraction and data representation technique widely used in the areas of pattern recognition, computer vision and signal processing. After feature extraction by PCA, nearest neighbor classifier based on the Euclidean distance is used to recognize an unknown face.

III. EXPERIMENTAL RESULTS AND CONCLUSION

To evaluate the recognition performance of the proposed method, we employed the images from the Yale B database and the CMU-PIE database. The Yale B database consists of 640 face images for 10 subjects representing 64 illumination conditions under the frontal pose. The CMU-PIE database contains 1,428 facial images of 21 illumination conditions for 68 individuals. First, we investigated the detection rates of two databases. Here, we only employed face images that were correctly detected the face and eyes regions from the input images in the experiments. In result, the detection rates revealed 91.71% and 94.57% for the Yale B and CMU-PIE databases, respectively. Fig. 2 showed an example of raw, histogram equalization, LBP, and LDP images for detected image and cropped image, respectively. Note that the cropped image (Group 2) are obtained from whole procedures including face and eyes detection, rotated-angle compensation, and image cropping, while the detected image (Group 1) denotes the face image obtained by the AdaBoost algorithms. The performance evaluation was carried out with each pre-processing image for each group.


raw histogram LBP LDP

(a) detected image (Group 1)

raw histogram LBP LDP

(b) cropped image (Group 2)

Fig. 2. Sample face image for Yale B database.

Next, we investigated the recognition performance of proposed system with two databases. To evaluate the recognition accuracy, we partitioned each database into training and testing sets. For the Yale B database, each training set comprised of seven images per subject, and the remaining images were used to test the proposed method. For the CMU-PIE database, three images from each person were also used for training and the remaining images were used for testing. For the Yale B database, the recognition results in terms of different preprocessing images and groups were shown in Fig. 3, and the results of the CMU-PIE database were also depicted in Fig. 4. As a result, the proposed method using LDP and PCA on cropped images showed a best recognition rate of 91.56% and 85.80% for Yale B and CMU-PIE database, respectively. Consequently, we confirmed the effectiveness of the proposed method under unconstrained environments through the experimental results.

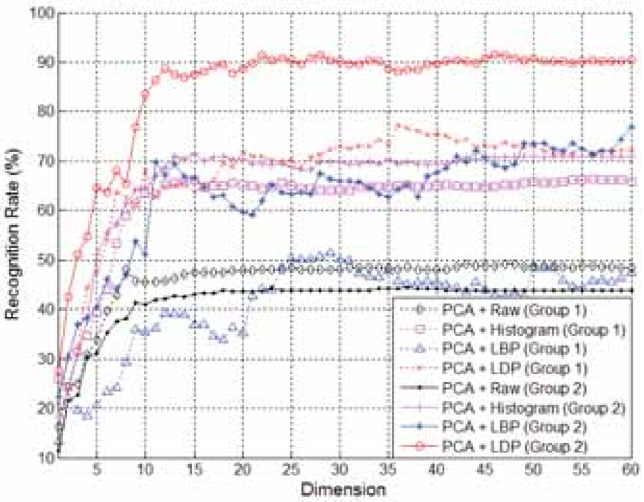


Fig. 3. Recognition rates of Yale B database.

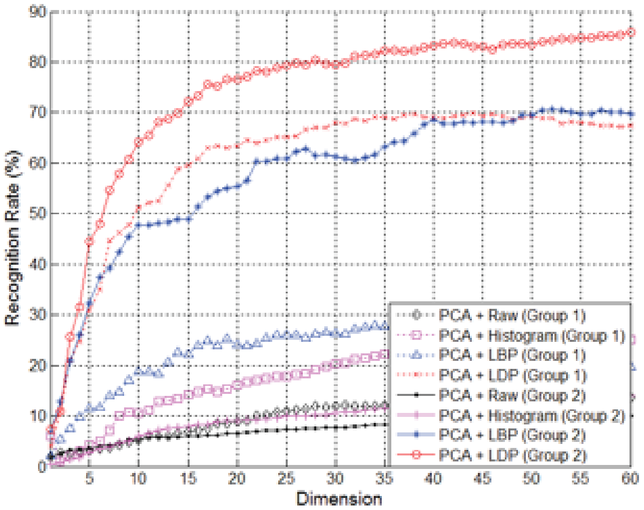


Fig. 4. Recognition rates of CMU-PIE illumination database.

REFERENCE

[1] N. B. Kachare and V. S. Inamdar, "Survey of face recognition techniques," International Journal of Computer Applications, vol. 1, no. 1, pp. 29-33, 2010.

[2] P. Viola and M. J. Jones, "Robust real-time object detection," Technical Report Series, Compaq Cambridge research Laboratory, CRL 2001/01, Feb. 2001.

[3] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.28, no.12, pp.2037-2041, 2006.

[4] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," ETRI Journal, vol.32, no.5, pp.784-794, 2010.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neurosci, vol. 3, no. 1, pp. 71-86, 1991.

Moving Object Detection Using Unstable Camera for Consumer Surveillance Systems

Seungwon Lee[1], *Student Member, IEEE*, Nahyun Kim[1], *Student Member, IEEE*, Inho Paek[2], *Member, IEEE*, Monson H. Hayes[1], *Fellow, IEEE*, and Joonki Paik[1], *Senior Member, IEEE*
[1]Image Processing and Intelligent Systems Laboratory, Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University, Seoul, Korea
[2]1st R&D Center/CP group, Nextchip Co., Ltd., Seoul, Korea

***Abstract*-- In this paper, we present a robust motion-based object detection system that corrects for the motion of an unstable camera. Assuming that the global camera motion may be modeled as an affine transform of the image between two successive frames, the proposed method is able to correct for camera motion using an elastic registration algorithm (ER). The local motion is then estimated from a current image and affine-transformed previous image. Finally, object regions are detected using the estimated local motions. Experimental results show that the proposed system is able to robustly detect moving objects in unstable imaging environment for consumer surveillance systems.**

I. Introduction

The detection of moving objects is a fundamental problem in computer vision and video processing, and has applications in many consumer electronics areas such as robot vision, and intelligent surveillance systems that use a variety of different types of cameras including pan-tilt-zoom (PTZ) and unmanned aerial vehicle (UVA) cameras [1]. Moving object detection may be considered as a low-level processing step for object recognition and analysis applications [2]. Generally, the first step that is used for objection detection is to remove the background using a background subtraction method. However, since background subtraction assumes that the camera is fixed, when the camera is moving there are difficulties with this approach.

Recently, an image signal processor (ISP) chip has been developed that provides block motion information as well as various other camera processing functions. Using the block motion information provided by this chip, Lee has proposed an adaptively partitioned block-based object detection method that estimates the direction of a moving object as well as object region [3]. However, this method fails to detect moving objects in the presence of global camera motion.

In this paper, we present a robust motion-based object detection method that is designed to work even in the presence of unstable camera motion. Under the assumption that the effect of the camera motion can be modeled as an affine transformation, we use the elastic registration (ER) algorithm proposed by Periaswamy to compensate for the effect of camera motion on successive frames [4]. More specifically, given the current frame, the previous frame is transformed using an affine matrix that is estimated using the ER algorithm in order to remove the effect of camera motion. Block-based local motion vectors are then estimated using optical flow between the current and the transformed previous frame. Finally, the object region is detected and the direction of motion is estimated using Lee's motion-based object detection method [3].

Experimental results show that the proposed method is an effective approach for moving object detection in unstable camera environments that can be embedded in an image signal processing (ISP) chip for high-level image processing functions for high-definition video surveillance systems.

II. Moving Object Detection Using Unstable Camera

When camera motion occurs, conventional motion-based object detection methods fail to detect objects because of global motion. Therefore, for robust object detection in an unstable imaging environment it is necessary to estimate and compensate for this global motion. Periaswamy has proposed an elastic registration algorithm that geometrically registers a source image to a target image using an affine matrix transformation [4]. Assuming that the camera motion is affine, we use the elastic registration algorithm without the brightness and contrast parameters in the general ER algorithm, to estimate the camera motion. More specifically, if $f(x, y, t)$ is the image in the current frame, and $f(\hat{x}, \hat{y}, t-1)$ the image in the previous frame, then the ER algorithm estimates the affine matrix that is used to model the motion between these two frames by minimizing the following quadratic error function:

$$E(\vec{a}) = \sum_{x,y\in\Omega} [\, f(x,y,t) - f(a_1x + a_2y + a_5, a_3x + a_4y + a_6, t-1)]^2, \tag{1}$$

where $\{a_1, a_2, a_3, a_4\}$ are the linear affine parameters, $\{a_5, a_6\}$ are the translation parameters, and Ω is the entire image. Since $E(\vec{a})$ is a nonlinear function of the parameters to be estimated, $E(\vec{a})$ is linearized using a first-order Taylor series expansion, which allows for the explicit solution for the vector $\vec{a}$, which is given by

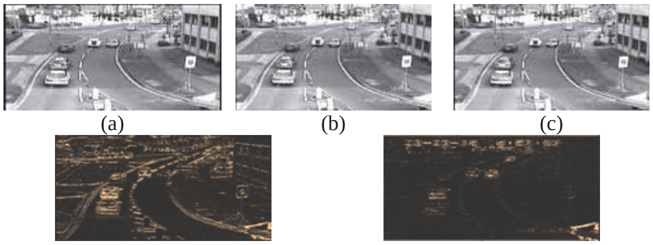
This research was supported by Basic Science Research Program through National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (2009-0081059), by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2011 (00045420-1), and by the Technology Innovation Program, Industrial Strategic Technology Development Program(Development of SR(Super Resolution) Image Scaler for 4K UHD), under Grant K10041900.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$\vec{a} = \left[\sum_{x,y\in\Omega} c(x,y)c^T(x,y) \right]^{-1} \left[\sum_{x,y\in\Omega} k(x,y)c(x,y) \right], \qquad (2)$$

Here, $c(x,y) = [f_x(x,y), f_y(x,y), -f_t(x,y), -1]^T$ where $f_x(x,y)$ and $f_y(x,y)$ are derivatives of $f(x,y)$ with respect to x and y, and $k(x,y) = f_t(x,y) - f(x,y)$ where $f_t(x,y)$ is the derivative of $f(x,y)$ with respect to "time."

Fig. 2 shows the results of the affine transformation of the input image using the estimated camera motion. The difference image between Figs. 2(a) and (b) is shown in Fig. 2(d), which shows a large amount of camera motion. On the other hand the difference image between Figs. 2 (b) and (c) is shown in Fig. 2(e), which has a smaller difference than Fig. 2 (d).


(a) (b) (c)

(d) (e)

Fig. 2. The results of the camera motion estimation using affine model; (a) input image, (b) previous image, (c) the result of the affine transformation of the input image, (d) the difference image between (a) and (b), and (e) the difference image between (b) and (c).

After correcting for camera motion, Lucas and Kanade (LK) optical flow is used for 8x8 block-based motion detection in the 8 times downscaled images [5]. In order to reduce the computational load, the estimated block motions are then classified into eight equispaced regions as shown in Fig. 3. If the magnitude of the motion is below a given threshold, then the block motion is considered to be equal to zero.

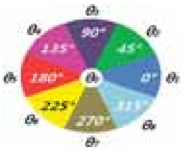


Fig 3. Eight equispaced regions for motion classification.

The proposed method uses intensity variance in each partitioned block instead of entropy used from [3], which has high computational cost for adaptive block partitioning. We compute the local variance of each initial block of size either 32×32 or 64×64. We assume that a block containing objects has a large variance. If the block has variance larger than a pre-specified threshold, then the block is divided into four sub-blocks.

If a partitioned block is larger than the 8x8 basic block, then the motion in the macro block is decided as follows: i) If a partitioned block has more than two zero basic motion blocks, it is considered as a static block, otherwise ii) the most frequent motion is assigned as the motion representing the block. Finally, the remaining motion blocks are labeled, which provides the moving direction, and small labeled object regions are removed for noise reduction.

III. Experimental Result

To evaluate the performance of the proposed method for moving object detection using an unstable camera, we tested sequences of size 765×512 at the rate of 24 frames per second.

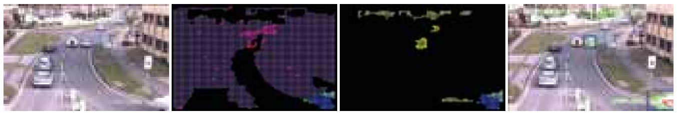

(a) (b) (c) (d)

Fig 4. The experimental result of the proposed method; (a) original image, (b) the detection result of the existing method, (c) the detection result of the proposed method, (d) labeling and estimated moving direction.

Fig. 4 compares the experimental result of the proposed method with an existing method [3], which fails to detect objects because of the global camera motion as shown in Fig. 4(b). On the other hand, the proposed method successfully detects the moving objects except overlaid characters on the image.

IV. Conclusion

In this paper, we proposed a robust motion-based object detection method using camera motion compensation in an unstable camera environment. Assuming that global camera motion results in an affine transformation of the image between two successive frames, an elastic registration algorithm (ER) is used to compensate for this motion. The local motion is then from two successive frames and, finally, object regions are detected using the estimated local motions.

Experimental results show that the proposed method can robustly detect moving objects in unstable imaging environment, and be embedded in an image signal processing (ISP) chip for high-level image processing functions in high-definition video surveillance systems.

Reference

[1] S. Lee and J. Paik, "Simultaneous object tracking and depth estimation using color shifting property of a multiple color-filter aperture camera," Proc. *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 1401-1404, 2011.

[2] R. Zhang, S. Zhang, and S. Yu, "Moving objects detection method based on brightness distortion and chromaticity distortion," *IEEE Trans. Consumer Electronics*, vol. 53, no. 3, pp. 1177-1185, 2007.

[3] S. Lee, J. Lee, E. Chon, M. Hayes, and J. Paik, "Moving object segmentation using motion orientation histogram in adaptively partitioned blocks for consumer surveillance system," *Proc. IEEE Int. Conf. Consumer Electronics*, pp. 202-203, 2012.

[4] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations." *IEEE Trans. Medical Imaging*, vol. 22, no. 7, pp. 865-874, July 2003.

[5] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *in Proc. DARPA Imaging Understanding Workshop*, pp. 121-130, 1981.

Fast Tile-Binning Method by Detecting 1D-Overlapped Primitives

Sang Oak Woo, Jeong-Soo Park, Seok-Yoon Jung and Shi-Hwa Lee

***Abstract*--Mobile 3D graphics use tile-based rendering algorithm in order to lower power consumption. For rendering tile by tile in the tile-based rendering, tile-binning process is inevitable and has to store the results of geometry processing into scene buffer, which causes memory access severely. Several methods were proposed to reduce memory bandwidth for tile-binning process. However, these memory-bandwidth reduction methods require more computations. Hsieh et al. attempted to eliminate edges in order to decrease the number of overlap test. Their method could only eliminate 0.02% of edges because the edges parallel to the axes can be eliminated. In this study, we propose a simple method to reduce the number of the overlap test by detecting 1D-tile overlapped primitives, which are overlapped with only one-dimensional tiles while keeping low memory bandwidth.**

I. INTRODUCTION

Memory-bandwidth reduction methods in tile-binning process aimed to reduce the number of false-overlapped tiles, which overlap with a bounding-box of a triangle but not with the triangle itself as shown in figure 1. When rendering each tile, the false-overlapped tiles make rasterizer load triangles, which do not generate any fragment visible.

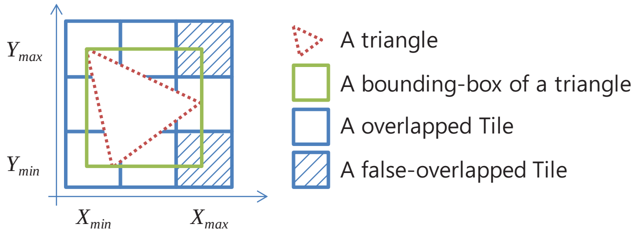


Fig. 1. Overlapped tiles and false-overlapped tiles

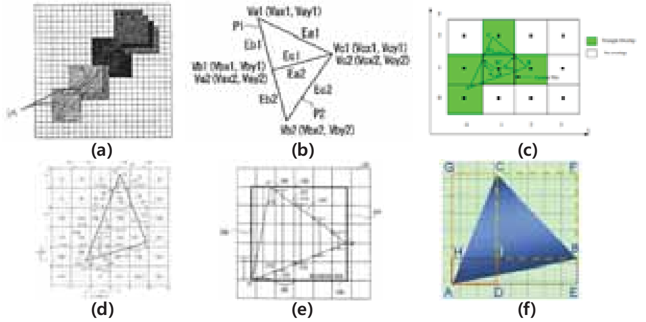


Fig. 2. Primitive subdivision methods by Junkins (a) and Naoi (b) and two-phased overlap test methods by Antochi (c), Baltaretu (d), Koneru (e) and Hsieh (f)

There are two kinds of methods for memory-bandwidth reduction in tile-binning process. The first type of methods is primitive subdivision method proposed by Junkins [1] and Naoi [2]. The second type of method is two-phased overlap test method by Antochi [3], Baltaretu [4], Koneru [5] and Hsieh [6] as shown in figure 2.

The primitive subdivision methods make the bounding-box smaller to reduce the number of false-overlapped tiles. The two-phased overlap test methods usually consist of a bounding-box overlap test and an advanced overlap test. Tiles overlapped with the bounding-box of a triangle are the input to the advanced overlap test in order to figure out whether tiles are really overlapped with the triangles. Any advanced overlap test requires more computations.

Hsieh et al. [6] introduced an edge elimination method as one module contained in the advanced overlap test in order to reduce computations. The edge that is considered to not have any false-overlapped tiles is eliminated from the edges list for the advanced overlap test. They thought only the edges parallel to the axes have no false-overlapped tiles.

In this study, we propose a simple method to reduce the number of the advanced overlap tests by detecting special primitives called “1D-tile overlapped primitives” and by bypassing overlap test of the 1D-tile overlapped primitives, which contain no false-overlapped tiles.

II. OVERLAPPING WITH 1D TILES

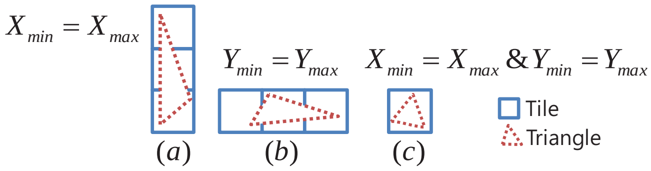


Fig. 3. Three types of 1D tile overlapped triangles: (a) X-1D tile, (b) Y-1D tile, and (c) one tile overlapped triangles

If triangles or edges are overlapped with 1D tiles as shown in figure 3, these primitives need not be tested for removing false-overlapped tiles because they contain no false-overlapped tiles. There are three types of 1D-tile overlapped triangles as follows.

X-1D tile overlapped triangle: $X_{min} = X_{max}$

Y-1D tile overlapped triangle: $Y_{min} = Y_{max}$

One tile overlapped triangle: $X_{min} = X_{max}$ & $Y_{min} = Y_{max}$, where X_{min}, X_{max}, Y_{min} and Y_{max} are the minimum/maximum tile coordinates of each axis overlapped with the bounding box as shown in figure 1.

A. 1D Tile Overlapped Triangles

As described, in the bounding-box overlap test phase the bounding box of the triangle is used to select only the tiles overlapped with the bounding box. Before stepping into the second phase of the advanced overlap test, if tiles resulting

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

from the first phase are one dimensional as shown in figure 3, the second phase can be skipped because there are no false-overlapped tiles.

B. 1D Tile Overlapped Edges

If the triangle is not 1D tile overlapped triangle, it means that the triangle may contain false-overlapped tiles. Hsieh's method [6] partitions triangle into three edges and the edges are investigated whether the edges contain false-overlapped tiles or not. After investigating, the edges with false-overlapped tiles are additionally tested in order to eliminate the false-overlapped tiles.

The proposed method investigates if edges making up a triangle are 1D-tile overlapped edges or not. 1D-tile overlapped edges do not have any false-overlapped tiles, similar to the 1D tile overlapped triangle. Hence, these edges do not need to be tested.

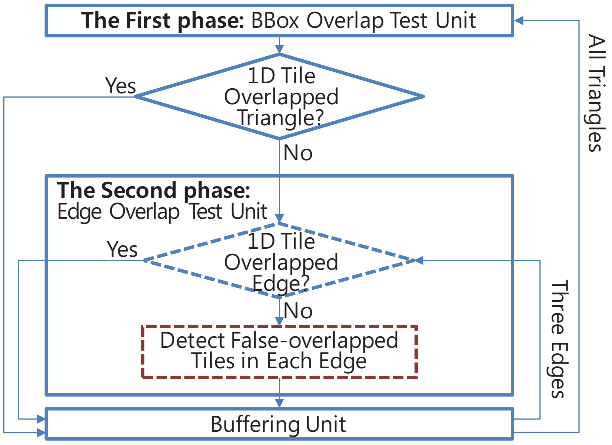

The First phase: BBox Overlap Test Unit
Yes
1D Tile Overlapped Triangle?
No
All Triangles
The Second phase: Edge Overlap Test Unit
Yes
1D Tile Overlapped Edge?
No
Detect False-overlapped Tiles in Each Edge
Three Edges
Buffering Unit

Fig. 4. Control flow of the proposed tile-binning algorithm

Figure 4 shows control flow of the proposed method. One of the merits of the proposed method is that it can replace "the second phase" with any other advanced overlap test. And further it can replace the red-dotted box in the second phase with any other false-overlapped tile detection algorithm.

III. EXPERIMENTAL RESULTS

Most popular benchmarks are used to verify the proposed method on the SRP-based GPU architecture [8]. More than 500 frames of each benchmark application are used as test scene. Our experiments shows that 90% of triangles are 1D-tile-overlapped triangles and additional 3% of edges out of 2D-tile-overlapped triangles are 1D-tile-overlapped edges. Finally 93% of primitives can skip overlap test because they are 1D-tile-overlapped primitives.

In the other hand, Hsieh's method shows that only 0.02% of edges are parallel to axis and can skip overlap test.

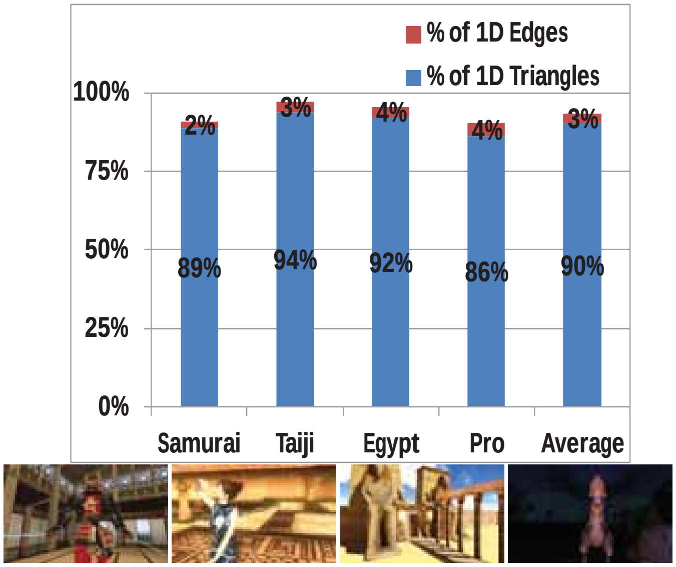


Fig. 5. Experimental results of four benchmarks: (a) the blue bar indicates the portion of 1D-tile-overlapped triangles, (b) the red bar indicates the portion of 1D-tile-overlapped edges.

IV. CONCLUSIONS

The proposed algorithm can bypass overlap testing of more than 90% of primitives with two simple integer comparisons. Therefore, it can reduce power consumption of tile binning module, because it can turn off overlap test module during almost 90% of running time. And one of the merits of the proposed method is that even if the proposed method is applied to any other overlap test, it could improve tile binning performance additionally.

REFERENCES

[1] Stephen Junkins, Oliver A. Heim and Lance R. Alba, "Polygon binning process for tile-based rendering," US 6,975,318 B2, 2002

[2] Junichi Naoi, "Image processing method," US 6,947,041 B2, 2002

[3] I. Antochi, B. Juurlink, S. Vassiliadis and P. Liuha, "Scene management models and overlap tests for tile-based rendering," Digital System Design, 2004. DSD 2004. Euromicro Symposium on, IEEE, pp. 424-431, 2004

[4] Oana Baltaretu, David L. Dignam and Sanjay O. Gupta, "Method for determining tiles in a computer display that are covered by a graphics primitive," US 6,437,780 B1, 2002

[5] Satyaki Koneru and Sajjad A. Zaidi, "Method and apparatus for determining bins to be updated for polygons, including lines," US 6,693,637 B2, 2001

[6] Hsiu-ching Hsieh, Chih-Chieh Hsiao, Hui-Chin Yang, Chung-Ping Chung and Jean Jyh-Jiun Shann, "Methods for precise false-overlap detection in tile-based rendering," Computational Science and Engineering, CSE'09. International Conf. on, vol. 2, IEEE, pp. 414-419, 2009

[7] RightWare, OpenGL|ES 2.0 performance benchmarking software, http://www.rightware.com/en/Benchmarking+Software, 2011

[8] Won-Jong Lee, Sang-Oak Woo, Kwon-Taek Kwon, Sung-Jin Son, Kyoung-June Min, Gyeong-Ja Jang, Choong-Hun Lee, Seok-Yoon Jung, Chan-Min Park and Shi-Hwa Lee, "A scalable GPU architecture based on dynamically reconfigurable embedded processor," HPG Poster, 2011

Gaussian Noise Image Restoration Using Local Statistics

Tuan-Anh Nguyen and Min-Cheol Hong, *Member, IEEE*
Department of Information and Telecommunication, Soongsil University, Korea

***Abstract*--We propose an efficient noise estimation-based method in removing Gaussian noise using local statistics. A proposed detector flexibly classifies the serious and mild noisy pixels prior to applying the strong and weak filters respectively.**

I. Introduction

In general, the obtained images are often corrupted by one or many kinds of noise due to the image system error and transmission effects. Among those artifacts, Gaussian noise-typed mostly appears on digital images that degrades its quality. Therefore, having the robust noise detection and filtering methods are strongly needed.

In this paper, we concentrate on the main problem in detection scheme design is the optimization of the tradeoff between noise removal and detail preservation.

II. Proposed Algorithm

A. Adaptive noise detector

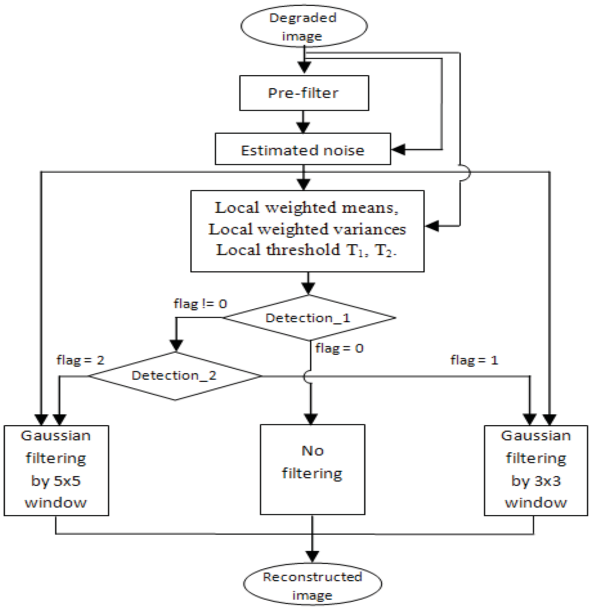


Fig. 1. Whole procedure of proposed algorithm

The degradation model is described by

$$y = x + n \tag{1}$$

where y, x and n represent, respectively, the *M*x*N* degraded, original image and the additive noise component. The local weighted statistics of each window of size (2U+1)x(2S+1) can be obtained as

$$\mu_{i,j} = \frac{\sum_{m=-U}^{U}\sum_{n=-S}^{S} w_{m,n} y_{i+m,j+n}}{\sum_{m=-U}^{U}\sum_{n=-S}^{S} w_{m,n}} \tag{2}$$

$$\sigma_{i,j} = \frac{\sum_{m=-U}^{U}\sum_{n=-S}^{S} w_{m,n} \left| y_{i+m,j+n} - \mu_{m,n} \right|}{\sum_{m=-U}^{U}\sum_{n=-S}^{S} w_{m,n}} \tag{3}$$

where the weighted matrix has the uniform entries, i.e. $w_{m,n} = 1,\ m \in [-U, U],\ n \in [-S, S] \cdot$

Under the assumption that an image is Gaussian distributed with different local activities, a noise detection function is defined as

$$flag_{i,j} = \begin{cases} 2, \text{ } if \ y_{i,j} < \mu_{i,j} - T_2 \ or \ y_{i,j} > \mu_{i,j} + T_2 \\ 1, \text{ } if \ \mu_{i,j} - T_2 < y_{i,j} < \mu_{i,j} - T_1 \\ \quad or \ \mu_{i,j} + T_1 < y_{i,j} < \mu_{i,j} + T_2 \\ 0, \text{ } otherwise \end{cases} \tag{4}$$

where $T_1 < T_2$ and obtained by

$$\begin{cases} T_1 = \text{cotg}\,\alpha = 1 / (\sigma_{i,j}^2 \sqrt{2\pi}) \\ T_2 = 2\sigma_{i,j} \end{cases} \tag{5}$$

Here, the angle α is defined as in Fig. 2 and used to represent the local activity of the current local window.

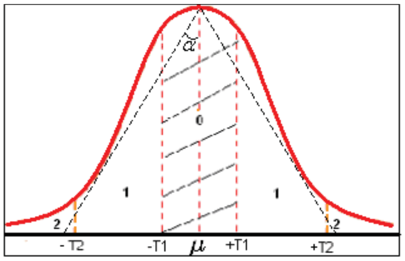


Fig. 2. Angle α and the two thresholds

Let assume that the same noise is added to both the flat area and the high activity region. Then the local standard deviation of the high activity is relatively higher than the flat area, causing the angle α be higher. This is how the local statistics affect the detection *flag*. Therefore, the thresholds T_1 and T_2 are calculated as in (5) to distinguish accurately between the serious noisy pixels and mild noisy ones.

Equation (4) means that when a pixel value is detected as noise free pixel, the $flag_{i,j}$ is marked as 0. On the contrary, if a pixel value is detected as corrupted one, the value of flag_{ij} equal to 1 for mild noised cases and equal to 2 for serious noised cases. Then the algorithm moves to the next pixel to

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2012M3C4A7032182)).

perform the same detection method for all pixels.

Moreover, the standard deviation of the noise component $\hat{\sigma}_{est}$ is calculated from the difference between the local standard deviations σ_D and σ_P of the same (2U+1)x(2S+1) local region in the degraded image and pre-filtered version which obtained as in [1]:

$$\hat{\sigma}_{est} = (MN)^{-1} \sum_{i=-U}^{U} \sum_{j=-S}^{S} \left|\sigma_D(i,j) - \sigma_P(i,j)\right| \tag{6}$$

B. Modified Gaussian filter

The estimated standard deviation is used to determine whether the image is seriously corrupted or not and to define the tuning parameter in Gaussian filter (8). Furthermore, we can obtain the new local statistics from (2) and (3) as

$$\begin{cases} \sigma' = \sigma - \hat{\sigma}_{est} \\ \mu' = \mu - \hat{\mu}_{est} \end{cases} \tag{7}$$

We propose a modified Gaussian filter

$$h_{m,n} = \frac{1}{Z}\exp\left(-T\,\frac{\left(\sigma'_{i,j}\right)^2 (m^2+n^2)}{\sqrt{\mu'_{i,j}+1}}\right) \tag{8}$$

where *Z* and *T* denote the normalizing constant and a tuning parameter, respectively. The support region (2U+1)x(2S+1), and the local information of the Gaussian filter in (8) are depended on the *flag$_{ij}$* in (4).

Finally, each pixel that flagged as 0 will be kept un-change in the output image whereas the noised one will be reconstructed as (9)

$$\hat{x}_{i,j} = \frac{\sum_{m=-U}^{U}\sum_{n=-S}^{S} h_{m,n}\, y_{i+m,j+n}}{\sum_{m=-U}^{U}\sum_{n=-S}^{S} h_{m,n}} \tag{9}$$

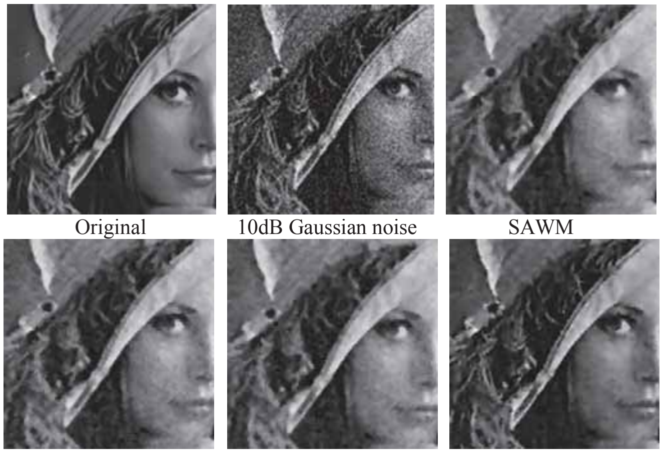
III. EXPERIMENTAL RESULTS

Gaussian distributed noise is added to the original 256x256 "Lena", “Cameraman”, "Goldhill" and “Bird” images with various SNRs (signal to noise ratio) and compared it with SAWM [2], BF [3], FAEA [4] methods. To evaluate the performance of the noise detection algorithm, the noise detection fidelity *(D_f)* [1], U = S = 1, and T = 0.05 are used.

Moreover, the denoising performance is assessed by using the peak signal to noise ratio (PSNR*)* and structural similarity index (SSIM) [5]. Tables 1 and 2 show that D_F of proposed algorithm outperforms the others for all cases, especially it is higher as the degree of noise is higher. Subjective comparisons in Fig. 3 and objective results of PSNR and SSIM are competitive among the above approaches.

IV. CONCLUSIONS

This paper presents a new effective algorithm for Gaussian noise detection and removal. In addition, the parameter used in this method has been automatically assigned based on the information obtained from the noise estimation step. From the experimental results, it is verified that the proposed algorithm effectively removes the Gaussian noise while keeping the image’s details.



Original 10dB Gaussian noise SAWM

BF FAEA Proposed

Fig. 3. Restoration results of different methods for Lena image

TABLE I
PERFORMANCE COMPARISONS FOR LENA IMAGE

Noise	Method	D_F	PSNR	SSIM
10dB	SAWM	84.96	27.60	0.752
	BF	N/A	29.43	0.824
	FAEA	92.16	28.08	0.774
	Proposed	94.98	29.62	0.854
20dB	SAWM	84.86	30.87	0.885
	BF	N/A	32.27	0.926
	FAEA	86.62	35.30	0.949
	Proposed	89.61	36.04	0.961
30dB	SAWM	70.03	31.66	0.909
	BF	N/A	32.63	0.938
	FAEA	71.52	39.61	0.975
	Proposed	70.04	41.04	0.982

TABLE 2
PERFORMANCE COMPARISONS FOR CAMERAMAN IMAGE

Noise	Method	D_F	PSNR	SSIM
10dB	SAWM	90.14	25.06	0.628
	BF	N/A	27.93	0.701
	FAEA	92.52	26.96	0.644
	Proposed	95.65	28.21	0.760
20dB	SAWM	86.02	27.34	0.755
	BF	N/A	31.22	0.886
	FAEA	87.84	34.13	0.899
	Proposed	93.62	34.50	0.902
30dB	SAWM	70.95	27.66	0.798
	BF	N/A	31.62	0.913
	FAEA	74.24	39.45	0.965
	Proposed	79.04	41.30	0.979

REFERENCES

[1] T. A. Nguyen, W. S. Song, and M. C. Hong, "Spatially adaptive noise detection and removal algorithm using local statistics," *IEEE Trans Consum. Electron.*, vol. 56, no. 3, Aug. 2010.

[2] X. Zhang, and Y. Xiong, “Impulse noise removal using directional difference based noise detector,” *IEEE Signal Process. Lett.,* vol. 16, no. 4, pp. 295-298, Apr. 2009.

[3] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” *Proc. IEEE Int. Conf. Computer Vision,* pp. 839-846, 1998.

[4] V. R. Vijaykumar, P. T. Vanathi, P. Kanagasabapathy, “Fast and Efficient Algorithm to remove Gaussian noise in digital images,” *Intl. J. Computer Science*, vol. 37, no. 1, pp. 78-84, Feb. 2010.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to Structural Similarity,” *IEEE Trans. Image Process.,* vol. 12, no. 4, Apr. 2004.

Proposal of a Universal Test Scene for Depth Map Evaluation

Istvan Andorko[1,2], *Member, IEEE,* Peter Corcoran[1], *Fellow, IEEE* and Petronel Bigioi[2], *Senior Member, IEEE*

College of Engineering and Informatics, National University of Ireland, Galway[1]; Digital Optics Corporation Europe, Ltd[2].

***Abstract*— Nowadays, a large number of depth map generation methods use additional devices, for example Infra-Red (IR) sensors, Time of Flight (ToF) cameras, etc. One of the disadvantages of these methods is, that they cannot use test images like the ones from the Middlebury database [1] to test their accuracy, for obvious reasons. We are planning to propose a universal test scene, which can be re-created by any researcher by providing a selection of universally accessible objects, scene layout measurements and test environment conditions such as light intensity and temperature.**

I. INTRODUCTION

The majority of the proposed depth map generation methods only rely on image processing algorithms that are either implemented on off-the-shelf CPUs, DSPs or SoCs. To test these methods, the researchers can access the test images from a specific database [1] and compare the accuracy of their methods with others. But, there are some cases, where beside the image processing algorithms implemented in software or hardware, some other special devices are used as well, such as near-IR sensors, IR sensors, TOF cameras, etc.

In order to develop high quality depth map for real-time applications, some researchers consider that additional devices need to be used, and these devices must be considered as being part of the depth map generation method.

In order to test these methods, the researchers had to create their own scenes, and due to this reason, they couldn't compare the accuracy of their methods with other available methods [2,3].

This is one of the reasons why we intend to propose a test scene that can be used by any researcher, no matter what type of method they use. In case of the methods that are only implemented in software or hardware, they will be able to download the stereo image pair together with the ground truth from our on-line database. In case of the methods that use additional devices, the researchers will be able to build a similar scene to the one in our database, only by following the specifications of the scene provided by us. The objects used in the scene are common objects that can be purchased no matter where the researchers are geographically located.

The structure of the paper is the following: In section II, a short description of depth map evaluation is provided. In section III the setup of the test scene is presented. We will conclude with the presentation of our progress so far, and our plans for future work

II. DEPTH MAP EVALUATION

Depth maps can be generated using several methods that have been researched in the past decades. Some of these methods include depth from defocus [4], depth from stereo images [1], depth map generation using Time of Flight (ToF) cameras [2], or depth map generation using structured light.

In order to evaluate a depth map, in most of the cases, the resulting depth map, the input image, and the ground truth, which provides pixel level depth information about the image, are needed. In most of the cases, the input images represent complex scenes, which challenge the depth map generation methods in order to differentiate them.

The description of such a scene is described in the following section.

III. TEST SCENE CREATION

A. Selection of the objects

In Scharstein et al [1], the conditions for a scene to be suitable for depth map testing are presented. Based on the Tsukuba stereo image pair, they present the most interesting regions of a test image as being:

1. Specular surfaces, which can cause difficulty in depth computation due to the reflected motion of the light surface.
2. Textureless regions, which are locally ambiguous and they are a challenge for stereo algorithms.
3. Depth discontinuities, which can be seen at the border of all the objects within the test image.
4. Occluded pixels, which can cause algorithms to give incorrect depth results for some objects.

Based on these specifications, it was decided to use the following objects for our proposed test scene:

1. 18% grey background
- The 18% grey background is used in the case of all image quality tests.
2. ISO chart
- The purpose of ISO test chart used is to provide a certain texture to the otherwise gray background.
3. Tennis balls
- The tennis balls have a matte surface, so they don't reflect light, but they are textureless.
4. 0.5l Coca Cola bottle
- The bottle's surface is shiny combined with transparent. It also has a standard color.
5. Macbeth color checker
- The purpose of the color checker is to have a matte and textured object in the test scene.

The work for this project was sponsored by the Irish Research Council for Science, Engineering and Technology (IRCSET) and DigitalOptics Corporation Europe, Ltd.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Creation of the scene

1) Light conditions

The white balance in the Image and Signal processing Pipeline (ISP) is mostly influenced by the temperature of the light. Several algorithms have been developed that are aiming to provide color consistency over a wide range of light temperatures [5]. In order to overcome the possibility of different ISPs generating different colors due to the difference in the light temperature, it was decided to set the light temperature of the test scene to 5500 K.

2) Camera setup

Another important aspect of the proposed test scene is the camera choice and settings. A DSLR camera fitted with a standard 18-55 mm lens was used. The reason is that generally, DSLR cameras are equipped with a high quality sensor, they have high performance ISPs, and give the user the option of full manual settings. We recommend the following settings for the camera: *f* number is 8 in order for the picture to be sharp at any focus distance, *exposure* should be 1/5 because of the small aperture, and the ISO should be set to 200 in order to avoid noisy pictures.

3) Scene setup

The scene is set up in a way to arguably challenge even the most advanced depth map generation methods. Some of these settings include the use of overlayed textureless objects, a number of similar objects placed one behind the other, half-transparent objects, a number of occluded objects, etc. The proposed test scene can be seen in figure 2.

Measurements of the object distance relative to the camera lens are provided for each of the object in the scene. These include vertical distance from the camera, horizontal distance from the camera, angle values relative to the camera's horizontal and vertical axes, light intensity and light temperature values for each of the objects. These measurements provide all the information needed in order to re-create the proposed test scene. The measurements can be seen in figure 2.

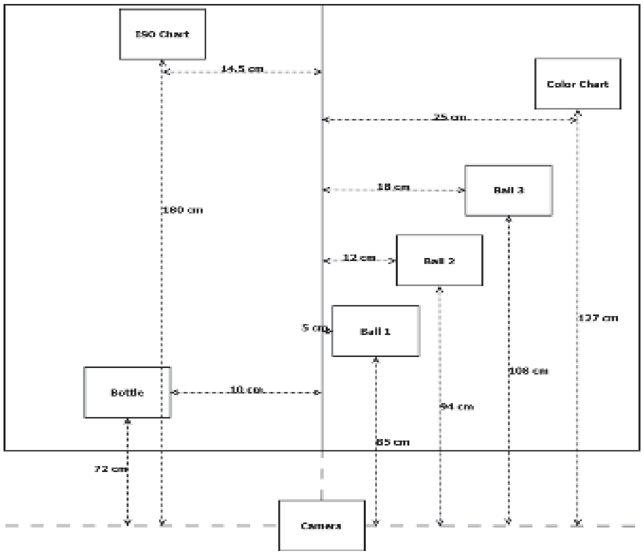


Figure 1. Scene measurements

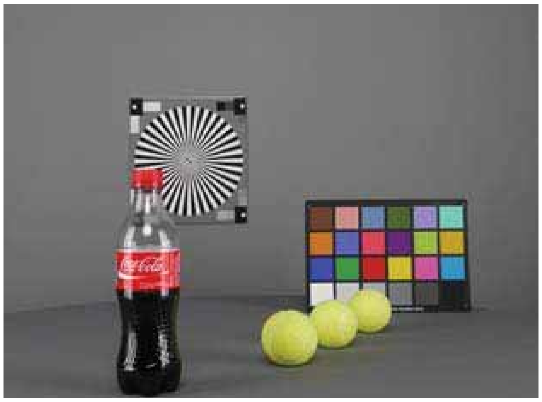


Figure 2. Scene setup

C. Generation of the ground truth

The ground truth of the depth maps can be generated in several ways. Different researchers have provided different ways of ground truth generation [1, 6], but not many have managed to provide a pixel-accurate ground truth.

Recently, a number of papers have been published, that present depth map evaluation methods, that don't need a ground truth. This would be the answer to our ground truth generation problem, and it would also encourage the researchers to use the test scene proposed by us [7, 8].

IV. Conclusions

A test scene for the purpose of depth map evaluation was proposed, which gives the opportunity to all the researchers to compare their work no matter what methods they are using.

All the measurements, together with the proposed test image are provided, and they will also be available online.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *International Journal of Computer Vision*, vol. 47, no.1, pp. 7 – 42, 2002.

[2] J. Zhu, L. Wang, R. Yang, J. E. Davis and Z. Pan, "Reliability fusion of Time-of-Flight depth and stereo for high quality depth maps", *IEEE Transactions on pattern analysis and machine intelligence,* vol.33, no.7, pp 1400 – 1414, 2010.

[3] S-Y. Kim, E-K. Lee and Y-S. Ho, "Generation of ROI emhanced depth maps using stereoscopic cameras and a depth camera", *IEEE Transactions on broadcasting,*vol.54, no.4, p. 732 – 740, 2008.

[4] V. P. Namboodiri, S. Chaudhuri and S. Hadap, "Regularized depth from defocus", *IEEE International conference on image processing,* pp. 1520 – 1523, 2008.

[5] K. Barnard, V. Cardei and B. Funt, "A comparison of computational color constancy algorithms-Part 1: Methodology and experiments with synthesized data", *IEEE Transactions on image processing,* vol. 11, no. 9, pp. 972 – 983, 2002.

[6] C. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection", *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, no. 7, pp. 675 – 684, 1999.

[7] C. T. E. R. Hewage and M. G. Martini, "Reduced-reference quality evaluation for compressed depth maps associated with colour plus depth 3D video", *IEEE International conference on image processing,* pp. 4017 – 4020, 2010.

[8] D. V. S. X. De Silva, W. A. C. Fernando, S. T. Worrall and A. M. Kondoz, "A novel depth map quality metric and its usage in depth map coding", *3DTV conference: The true vision – capture, transmission and display of 3D video,* pp. 1-4, 2011.

Adaptive Local Tone Mapping Based on Retinex for High Dynamic Range Images

Hyunchan Ahn, Byungjik Keum, Daehoon Kim, and Hwang Soo Lee, *Member, IEEE*
Department of Electrical Engineering, KAIST, Daejeon, Korea

***Abstract*--In this paper, we present a new tone mapping technique for high dynamic range images based on the retinex theory. Our algorithm consists of two steps, global adaptation and local adaptation of the human visual system. In the local adaptation process, the Gaussian filter of the retinex algorithms is substituted with a guided filter to reduce halo artifacts. To guarantee good rendition and dynamic range compression, we propose a contrast enhancement factor based on the luminance values of the scene. In addition, an adaptive nonlinearity offset is introduced to deal with the strength of the logarithm function's nonlinearity. Experiments show that our algorithm provides satisfactory results while preserving details and reducing halo artifacts.**

I. INTRODUCTION

The dynamic range of real world scenes is extensively large spanning over four orders from shadows to highlights [1]. Due to its adaptation mechanisms, the human visual system can cope with the real scenes. However, photographs are different from the real scenes, such as daylight outdoor scenes whose dynamic range is vast. The dynamic range of a scene cannot be captured by a conventional camera or a digital camera because of their imperfectness [2]. Although high dynamic range (HDR) images containing the full dynamic range of the real scene can be obtained from differently exposed photographs [3], low dynamic range (LDR) display devices such as ordinary monitors cannot handle the full dynamic range of the scene. The devices can display only two orders of magnitude [1], [4]. Once the HDR images are linearly mapped to the display devices, much information is lost. Thus, the HDR images must be compressed before being mapped to the devices. The mapping techniques from the HDR images to the LDR display devices are called tone mapping or tone reproduction which is related with our work.

There are two categories separating the techniques: first one is the global operator, and the other one is the local operator [2]. The global tone mapping operators apply a single function to all pixels. Drago *et al.* [5] presented logarithmic compression of scene's luminance values with an adaptive logarithm base. In the darkest area, they used 2 as a logarithm base, whereas they used 10 as a logarithm base to compress contrast more in the highest area. Later, Cvetković *et al.* [6] introduced an improved tone mapping function based on splines which enhances the visibility in dark regions while preserving the visibility in bright regions. Furthermore, they combined their algorithm with a multiple-exposure technique to improve SNR in dark regions. These global methods usually require low computational complexity but cannot preserve the details of the scene.

The local tone mapping operators apply different functions to each pixel based on its neighborhood pixels. Reinhard *et al*. [1] developed a new tone mapping operator which uses a photographic experience called Zone System. They introduced an automatic dodging and burning technique to avoid loss of details. Durand and Dorsey [7] introduced a fast bilateral filter to decompose the image into a base layer and a detail layer. They reduced contrast of the base layer, while preserving the detail layer. Fattal *et al*. [8] presented a new method based on gradients. They attenuated magnitudes of the large gradients in a logarithm domain using a gradient attenuation function. After solving a Poisson equation, the tone mapped image was obtained. Later Meylan and Susstrunk [4] proposed a retinex based method to render HDR images. They introduced an adaptive filter and a sigmoid function to solve the drawbacks of the surround-based retinex. These local methods usually perform better than global methods, but there are several drawbacks. One is that the local methods are more complex than global methods. The other is that halo artifacts arise from utilizing neighborhood pixels. Accordingly, along with preserving visual contents and the overall appearance of the original scene, reducing halo artifacts must be considered when designing a tone mapping operator. Especially, it is important to preserve much information from scenes for video surveillance systems.

In this paper, we propose a new local tone mapping method that preserve details and prevent halo artifacts based on the center /surround retinex [9-11].

This paper is organized as follows. Section 2 reviews the center/surround retinex which is the basis for our work. We investigate the characteristics of the center/surround retinex and describe its drawbacks. Our new tone mapping operator is proposed in Section 3, and Section 4 provides experimental results. Finally, Section 5 concludes our work.

II. CENTER/SURROUND RETINEX

The retinex theory was initially defined by Land [12]. It explains how the reliable color information from the world is extracted by the human visual system [4]. Based on the center/surround retinex [9], Jobson *et al.* introduced the single-scale retinex (SSR) [10] and the multiscale retinex (MSR) [11]. In this paper, these are called retinex algorithms. SSR is given by

$$R_i(x,y) = \log I_i(x,y) - \log(F(x,y) * I_i(x,y)), \qquad (1)$$

where *x, y* are the pixel coordinates in the image, $R_i(x,y)$ is the retinex output, $I_i(x,y)$ is the image distribution in the *i*-th spectral band, * denotes the convolution operation, and $F(x,y)$ is the Gaussian surround function,

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$F(x,y) = Ke^{-(x^2+y^2)/c^2}, \tag{2}$$

where c is the Gaussian surround space constant. K is the normalization factor.

A small space constant produces good dynamic range compression but bad color rendition. Conversely, a large constant produces good color rendition but bad dynamic range compression [10], [11]. MSR is described in (3),

$$R_{MSR_i}(x,y) = \sum_{n=1}^{N} \omega_n R_{n_i}(x,y), \tag{3}$$

where N is the number of scales,$R_{n_i}(x,y)$is the i-th component of the n-th scale, $R_{MSR_i}(x,y)$ is the i-th spectral component of the MSR output, and ω_n is the weight associated with the n-th scale.

The purpose of MSR is to reduce halo artifacts around high contrast edges and to keep balance with the dynamic range compression and the color rendition. MSR produces good dynamic range compression, but still suffer from halo artifacts. In addition, SSR with small space constant makes large uniform regions graying out and flat-looking in images. These drawbacks are shown in Fig. 1 and also investigated in previous studies, such as [10], [11], [4].

III. PROPOSED TONE MAPPING BASED ON RETINEX

In Section 2, we described the characteristics and drawbacks of the retinex algorithms. These drawbacks are overcome by introducing a new method. In our algorithm, luminance values are obtained from input HDR images and processed. First, we apply a global tone mapping as a preprocessing. After that, a local tone mapping is applied based on the retinex algorithms. Finally, after normalization, an output image is obtained from the processed luminance values and the input chrominance values.

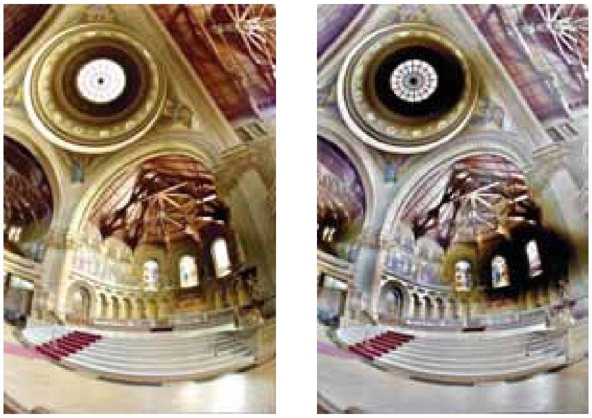
A. Global Adaptation

Global adaptation takes place like an early stage of the human visual system [4]. The human visual system senses brightness as an approximate logarithmic function according to the Weber-Fechner law [5]. To globally compress the dynamic range of a HDR scene, we use the following function in (4) presented in [5].

$$L_g(x,y) = \frac{\log(L_w(x,y)/\overline{L}_w + 1)}{\log(L_{w\max}/\overline{L}_w + 1)}, \tag{4}$$

where $L_g(x,y)$ is the global adaptation output, $L_w(x,y)$ is the input world luminance values, $L_{w\max}$ denotes the maximum luminance value of the input world luminance values, $\overline{L}_w$ and is the log-average luminance [1] and given as

$$\overline{L}_w = \exp\left(\frac{1}{N}\sum_{x,y}\log(\delta + L_w(x,y))\right), \tag{5}$$



(a) (b)

Fig. 1. Results of the retinex algorithms with drawbacks. (a) MSR output. (b) SSR output with small space constant.

where N is the total number of pixels in the image and δ is the small value to avoid the singularity that occurs if black pixels are present in the images.

The input world luminance values and the maximum luminance values are divided by the log-average luminance of the scene. This enables (4) to adapt to each scene. As the log-average luminance converges to the high value, the function converges from the shape of the logarithm function to the linear function. Thus, scenes of the low log-average luminance are boosted more than scenes with high values. As a result, the overall scene luminance values are adequately compressed in accordance with the log-average luminance of the scene.

B. Local Adaptation

Local adaptation based on the retinex theory is applied after the global adaptation process. In (1), the output $R_i(x,y)$ of the input $I_i(x,y)$ which lies near the much brighter pixel values become very dark, and this causes the halo artifacts which make the result looks unnatural. The artifacts can be reduced by introducing an edge-preserving filter. We substitute the guided filter [13] for the Gaussian filter of the retinex algorithms. The guided filter is an edge-preserving filter like the bilateral filter [14] whose weights depend not only on the Euclidean distances but also on the luminance differences. These filters behave similar, but the guided filter has better performance near the edges [13]. Also, its computational complexity is linear-time without approximation and independent of the kernel size [13]. The local adaptation equation can be written as

$$L_l(x,y) = \log L_g(x,y) - \log H_g(x,y), \tag{6}$$

where $L_l(x,y)$ denotes the local adaptation output, and $H_g(x,y)$ is the output of the guided filter applied to $L_g(x,y)$,

$$H_g(x,y) = \frac{1}{|\omega|}\sum_{(\xi_x,\xi_y)\in\omega(x,y)}(a(\xi_x,\xi_y)L_g(x,y) + b(\xi_x,\xi_y)), \tag{7}$$

where ξ_x, ξ_y are the neighborhood pixel coordinates, $\omega(x,y)$ is a local square window of a radius r centered at the pixel (x,y),

| ω | is the number of pixels in $\omega(x,y)$, $a(\xi_x, \xi_y)$ and $b(\xi_x, \xi_y)$ are some linear coefficients,

$$a(\xi_x,\xi_y)=\frac{\mu_2(\xi_x,\xi_y)-\mu^2(\xi_x,\xi_y)}{\sigma^2(\xi_x,\xi_y)+\varepsilon}, \tag{9}$$

$$b(\xi_x,\xi_y)=\mu(\xi_x,\xi_y)-a(\xi_x,\xi_y)\mu(\xi_x,\xi_y), \tag{10}$$

where $\mu(\xi_x, \xi_y)$ and $\sigma_2(\xi_x, \xi_y)$ are the mean and variance of L_g in $\omega(\xi_x, \xi_y)$, $\mu_2(\xi_x, \xi_y)$ is the mean of L^2_g in $\omega(\xi_x, \xi_y)$, and ε is a regularization parameter. The guidance and input image of the guided filter are identical in our algorithm.

After applying the filter, the halo artifacts are significantly reduced, but the output gives unsatisfactory overall appearance due to its low global contrast. Because of the characteristics of the edge-preserving filter, the filter assigns very low weights to the neighborhood pixels which have large differences between their luminance values and the value of the center pixel. As a result, $L_g(x,y)$ and $H_g(x,y)$ values of (6) tend to be analogous, which make no large differences among pixels of the local adaptation output. The result of (6) gives flat-looking appearance and loses its original luminance distribution of the input image. This analysis leads us to introduce new methods.

To prevent the flat-looking appearance caused by the filter and improve the performance of our method, we introduce two important factors. One, the contrast enhancement factor is given by

$$\alpha(x,y)=1+\eta\frac{L_g(x,y)}{L_{g\max}}, \tag{11}$$

where η denotes the contrast control parameter, and $L_{g\max}$ is the maximum luminance value of the global adaptation output.

The other, the adaptive nonlinearity offset which varies in accordance with the scene contents can be written as

$$\beta=\lambda\overline{L}_g, \tag{12}$$

where λ is the nonlinearity control parameter, and $\overline{L}_g$ is the log-average luminance of the global adaptation output.

By integrating these factors into (6), the final local adaptation equation is established as follows:

$$L_{out}(x,y)=\alpha(x,y)\log\left(\frac{L_g(x,y)}{H_g(x,y)}+\beta\right), \tag{13}$$

where $L_{out}(x,y)$ is the final local adaptation output.

We introduce the contrast enhancement factor $\alpha(x,y)$ to achieve satisfactory overall appearance of the rendered image. As mentioned above, using the guided filter causes the flat-looking appearance. The global adaptation output $L_g(x,y)$ controls the contrast enhancement factor of each pixel. Originally dark pixels make the contrast enhancement factor low and bright pixels make the factor high. Therefore, the local adaptation output has more natural appearance than before by considering the globally compressed scene's luminance values.

TABLE Ⅰ
The parameters used for our experiments.

Parameters	Value
r	10
ε	0.01
η	36
λ	10

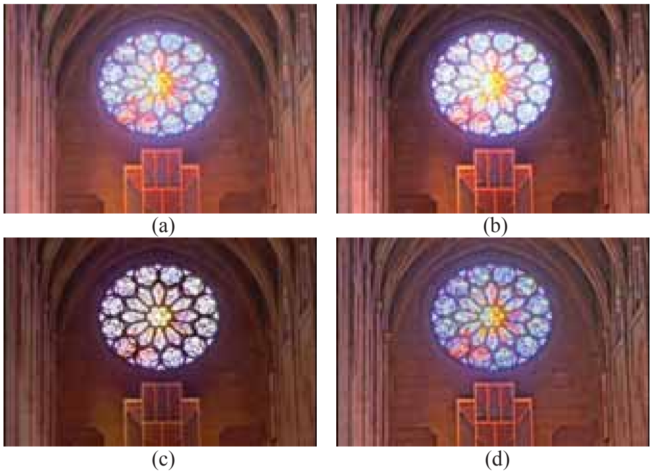

(a) (b) (c) (d)

Fig. 3. Tone mapping results of Rosette. (a) Adaptive Logarithmic. (b) Photographic. (c) Retinex based adaptive filter. (d) Proposed.

In order to control the nonlinearity of the logarithm function according to the scene, we propose the adaptive nonlinearity offset β. The logarithm function is a nonlinear function whose gradient is gradually decreasing. The log-average luminance of the global adaptation output controls the strength of the nonlinearity by changing the starting point of the logarithm in (13). The low log-average luminance of the global adaptation output makes the logarithm function start from a large gradient. Then, the logarithm curve increases the overall luminance values more than the high log-average luminance case. This ensures proper mapping of the local adaptation output based on the scene contents.

After the local adaptation, the processed luminance values are rescaled from 0 to 1. Finally, the tone mapped image is obtained from the luminance values of the local adaptation output and the input HDR image.

IV. EXPERIMENTAL RESULTS

A variety of HDR images are tested in our experiments and the following luminance value is used: $L = 0.299R + 0.587G + 0.114B$. The parameters used for our experiments are shown in Table Ⅰ and the default parameters work well for various HDR images. The values of the radius r and regularization parameter ε of the guided filter keep balance with reducing halo artifacts and preserving the local contrast, the value of the nonlinearity control parameter λ ensures appropriate consideration to the contents of the scene, and the value of the contrast control parameter η provides proper overall contrast. As λ and η become larger, the overall luminance values of the output become darker and the global contrast of the output increases respectively.

The comparison with other tone mapping operators which are the adaptive logarithmic method [5], the photographic

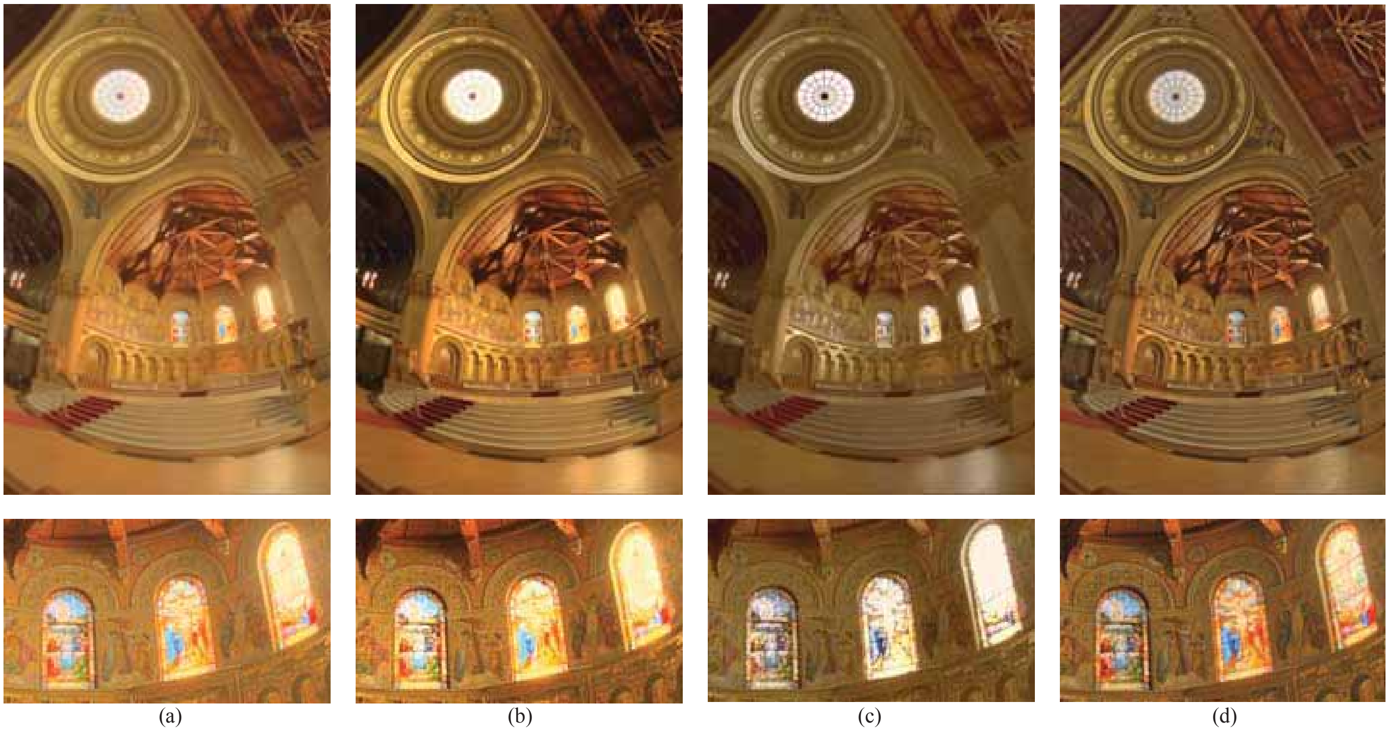


Fig. 4. Tone mapping results of Stanford Memorial Church. (a) Adaptive Logarithmic. (b) Photographic. (c) Retinex based adaptive filter. (d) Proposed.

method [1], the retinex based adaptive method [4], and the proposed method is presented in Fig. 3 and Fig. 4. All the methods are tested with default parameters. Since it is difficult to evaluate the objective performance of the methods, we choose psychophysical experimentation to evaluate the results. First three methods cannot preserve details well in bright region, but our method not only preserves a lot of visual contents but also shows good overall appearance which is competitive or better than other methods. The enlarged images in Fig. 4 show that our method displays the clearest details of all the available methods. Besides, it is another advantageous of our method that users can control the global contrast depending on its application.

Excluding the guided filtering process, our method takes 26ms for 512*768 pixels on 2.40 GHz Core 2 Quad with un-optimized code. The running time of the guided filtering which is faster than bilateral filtering is reported in [13].

V. CONCLUSIONS

With the increasing use of HDR images, the tone mapping techniques have been widely studied. We propose a local tone mapping algorithm based on the retinex theory to process the HDR images. Instead of using the Gaussian filter of the retinex algorithms, we adopt a guided filter to reduce the halo artifacts. We found that the guided filter does not take full advantage of neighborhood pixels, which causes the flat-looking appearance. To prevent this drawback, we introduce the contrast enhancement factor. Furthermore, we handle the logarithm function's nonlinearity by adding the adaptive nonlinearity offset. With these factors, we simultaneously achieve a good dynamic range compression and good overall appearance. The experimental results demonstrate that our method yields satisfactory rendering of HDR images and preserves much details, which will be beneficial for future video surveillance systems and consumer digital cameras.

REFERENCES

[1] E. Reinhard, M. Stark, P. Shirley, J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graphics*, vol. 21, no. 3, pp. 267–276, July. 2002.

[2] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, "High Dynamic Range Imaging," *Morgan Kaufmann Publishers*, 2005.

[3] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. SIGGRAPH 97*, Aug. 1997.

[4] L. Meylan and S. Susstrunk, "High dynamic range image rendering with a retinex-based adaptive filter," *IEEE Trans. Image Processing*, vol. 15, no. 9, Sept. 2006.

[5] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive Logarithmic Mapping for Displaying High contrast Scenes," *Computer Graphics Forum*, vol. 22, no. 3, pp. 419-419, Sept. 2003.

[6] S. Cvetković, J. Klijn, and P. H. N. de With, "Tone-mapping functions and multiple-exposure techniques for high dynamic-range images," *IEEE Trans. Consumer Electronics*, vol. 54, no. 2, pp. 904–911, 2008.

[7] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic- range images," *ACM Trans. Graphics*, vol. 21, no. 3, pp. 257–266, July. 2002.

[8] R. Fattal, D. Lischinski, and M. Werman, "Gradient Domain High Dynamic Range Compression," *ACM Trans. Graphics*, vol. 21, no. 3, pp. 249–256, July. 2002.

[9] E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," in *Proc. Nat. Acad. Sci.*, vol. 83, pp. 3078–3080, 1986.

[10] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Processing*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[11] D. Jobson, Z. Rahman, and G. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Processing*, vol. 6, no. 7, July. 1997.

[12] E. Land and J. McCann, "Lightness and Retinex theory," *J. Opt. Soc.Amer.*, vol. 61, no. 1, pp. 1–11, Jan. 1971.

[13] K. He, J. Sun, and X. Tang. "Guided image filtering," in *Proc. European Conf. Computer Vision,* pp. 1-14, 2010.

[14] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Computer Vision*, Jan.1998.

Implementation of the Raster Pipeline over the 3D Geometry Pipeline: a Point-Set Approach

Nakhoon Baek, *Member, IEEE*
Kyungpook National University, Daegu 702-701, Korea

Abstract-- **We implemented the raster graphics features on the typical mobile graphics devices, which only support 3D geometric output primitives. Our approach is based on the point sets, rather than traditionally-recommended texture maps.**

I. Introduction

Current mobile graphics standards are targeted at the mobile phones and embedded systems with relatively low-tier hardware. Thus, they usually remove unnecessary and/or rarely used features. The raster pipelines are often intensively removed, since their usages are expected to have decreased [1].

As typical examples, the standard specifications of *OpenGL ES* (embedded systems) version 1.0, 1.1, and 2.0 lacks any raster graphics pipelines and/or their corresponding API (application program interface) functions, even for handling *pixmaps* (true color images) and *bitmaps* (black-white images) [2]. In contrast, the original *OpenGL* has plenty of raster operations [3].

At the early design stages of the mobile graphics standards, they naturally expected to indirectly emulate the raster operations through *texture mapping* features. In contrast, various OpenGL application programs use image displays (with pixmaps) and character font displays (with bitmaps), much more frequently than expected. Additionally, the texture mapping techniques are tiresome to set up the 3D objects and their texture mappings, even for simple 2D image displays.

In this paper, we show our new raster pipeline implementation scheme over the mobile graphics standards without raster operation supports. Generally, the texture mapping techniques are often referred to be suitable for emulating raster operations [4]. However, at least to the best of our knowledge, there has been no detailed description or literature for the implementation details of these emulated raster operations. Remarkably, we also found that there are some technical difficulties for implementing raster operations with texture maps, against the traditional expectations on them. Considering pixmaps and bitmaps as sets of points, we present another way of implementing the raster operations over the 3D geometry processing pipeline.

II. Problem Definition

Typical raster pipelines perform the direct display of 2D color-value arrays representing color images or character fonts. In its simplest implementations, a 2D image will be shown on the screen, just as is. In most of widely used 3D graphics libraries, their implementations are more enhanced, for seamless uses with the main-stream 3D geometric objects. Based on the most widely used OpenGL specifications [3], we need to provide at least the following raster functions:

RasterPos(float *x*, *y*, *z***)**: This function specifies the *current raster position*, where bitmaps/pixmaps are located. The given 3D coordinates are processed with the current 3D geometry pipeline settings, to easily specify the raster position with respect to the target 3D objects, as shown in Figure 1.(a).

Color(float *r*, *g*, *b*, *a***)**: It specifies the output color for the bitmap display, which is stored as the *current raster color*.

DrawPixels(sizei *width, height,* **enum** *format, type,* **void*** *data***)**: It directly displays the given pixmap, which is a two-dimensional color array, sized *width* × *height*, and stored at *data*, at the current raster position. The *format* and *type* specify the internal data representations of pixels values, which is typically RGB (red, green, blue) or RGBA (red, green, blue, alpha) tuples.

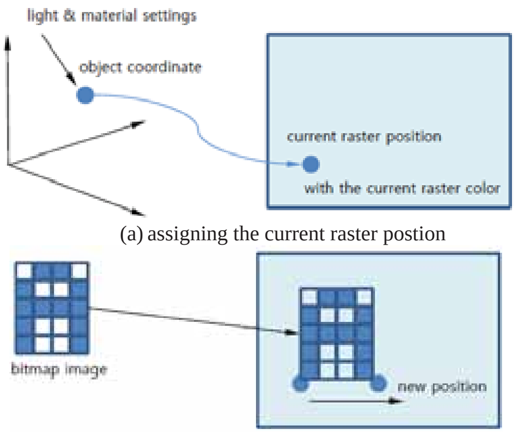
Bitmap(sizei *width, height,* **float** *xorg, yorg, xinc, yinc,* **ubyte*** *data***)**: It displays the given bitmap, which is a two-dimensional array with 0 or 1 values, sized *width* × *height*, and stored at *data*, with the foreground color of current raster color. The bitmap array origin is shifted with (*xorg*, *yorg*). Additionally, the current raster position is updated with (*xinc*, *yinc*), to be ready for the next output. Notice that bit 0's produce no output, as transparent pixels.

To emulate these raster operations, they say we can apply texture mapping of the pixmap/bitmap image onto a rectangle perpendicular to the scene camera. However, we found that this texture mapping method has drawbacks as follows:

- **Texture sizes are limited**. In typical mobile graphics pipelines, the texture mapping features are minimized to be more cost-effective, and it is general to support textures with only power-of-two widths and heights. To show the generally-sized rectangular images, we should use other graphics techniques such as *stencil masking*.
- **Texture mappings are heavy operations**. In most mobile graphics systems, their texture mapping capabilities are tightly limited. Thus, applying texture mapping techniques to relatively simple bitmap/pixmap displays, we can easily meet rapid performance failures.
- **Bitmaps should be transparent**: Bit 0's in a bitmap represents transparent pixels. In its texture mapping, we

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program (Immersive Game Contents CT Co-Research Center).

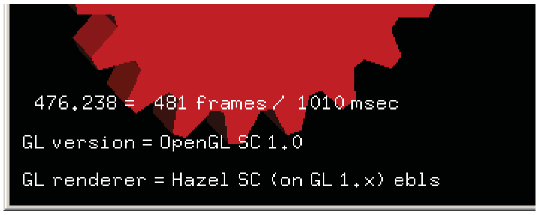
978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a) assigning the current raster postion

(b) display of a pixmap/bitmap image

Figure 1. Raster pipeline operations.



```
476.238 = 481 frames / 1010 msec
GL version = OpenGL SC 1.0
GL renderer = Hazel SC (on GL 1.x) ebls
```

Figure 2. Our system generates the bitmap characters.

need to generate a wholly new texture image in which those pixels have *alpha* values of 0. Masking out those pixels with stencil buffer is another way of implementing it.

From these reasons, the emulation method with texture mapping needs extra graphics techniques such as *alpha blending* or *stencil masking*. This extra requirement may seriously limit user's choice for other special effects. Furthermore, most pixmap/bitmap images need to be resized to power-of-two widths and/or heights.

To avoid these difficulties, we use another way of generating a set of points corresponding to the pixels in the pixmap/bitmap images. We treat the given 2D image as a set of pixels, and uses 3D graphics primitives to present those sets of points to the corresponding window positions.

In this case, we can handle any width and height sizes, and do not need *alpha blending* or *stencil masking* techniques any more. Even on low-tier systems without any alpha-blending or stencil masking facilities, our method can work.

Our method has a drawback. For an image pixel, we should specify its (*x*,*y*,*z*) coordinates for the output function, which will increase data transfers between CPU and graphics card. For the bitmap images, the transparent pixels do not need these data transfer, and thus the total overhead is not so significant.

III. Implementation Strategy

Based on the required features presented in the previous section, our implementation strategies are shown in this section.

A. Algorithm for Pixmap Display

step 1. Save the state variables: Save all the state values of the coordinate transformation pipeline, to later use the transformation features. Turn off all the texture mapping, light-and-material features, to by-pass the color values directly to the pixels.

step 2. Output: Set all the transformation matrices to be identity, to output the window coordinates as is. Providing the pixel coordinates (*x*,*y*,*z*) into the vertex array, and RGBA color values into the color array, use output primitive functions to draw the image.

step 3. Restore: Restore all the saved variables.

B. Algorithm for Bitmap Display

step 1. Save the state variables.

step 2a. Output: We need some extra works to get transparent pixels. Provide the current raster color as the foreground RGBA color values. Draw the given bitmap with the output primitive function, through providing the point coordinates for the bit 1's, while ignoring bit 0's.

step 3. Restore.

step 4. Extra post-processing: The current raster position is moved by (*xinc*, *yinc*) values.

IV. Implementation Results

We implemented our raster pipeline features, based on the algorithms shown in Section 3. Figure 2 shows the examples of bitmap characters produced by our implementation. To approve its correctness, we used a set of tests from the official standard conformance test suites [5]. Our implementation finally passed all these test routines.

V. Conclusion

In light-weight graphics systems typically for mobile phones and embedded systems, they removed traditional raster pipelines. However, we still need the raster handling features for displaying bitmap-based fonts and pixmap images. Though it was known that texture mappings may be suitable for the emulation of these raster features, we found that they need some extra features such as alpha-blending and/or stencil techniques. To avoid these difficulties, we presented another way of generating a set of points and using the traditional 3D geometry pipeline, to display them. We presented detailed algorithms and implementation results.

References

[1] K. Pulli, "New APIs for mobile graphics," In *Proc. SPIE Multimedia on Mobile Devices II*, 2006.

[2] D. Blythe, *OpenGL ES Common/Common-Lite Profile Specification*, Khronos Group, 2007.

[3] M. Segal and K. Akeley, *The OpenGL Graphics System: A Specification*, Version 2.1, Dec 2006.

[4] T. Olson, *OpenGL ES Extension No. 7: OES_draw_texture*, Khronos Group, 2004.

[5] Khronos Group, http://www.khronos.org/opengles/sc/.

Single Reference Super-Resolution Using Inter-Subband Correlation of Directional Edge

Oh-Jin Kwon[1], Eun-Hee Lee[1], Hee-Suk Pang[1], and Youngseop Kim[2], *Member, IEEE*
[1]Department of Electronics Engineering, Sejong University, Seoul, Korea
[2]Department of Electrical and Electronics Engineering, Dankook University, Yongin, Korea

***Abstract*--We propose an efficient single reference super-resolution system based on the discrete wavelet transform. We assume that the low-resolution image to be enhanced is the low-frequency subband of the high-resolution image to be reconstructed. We design a support vector machine synthesizing the high-frequency subband based on the inter-subband correlation of the directional edges. Experimental results of sample images show that the proposed system offers improvements in terms of both measured distortion and subjective appearance.**

I. INTRODUCTION

Super-resolution (SR) is the technology enhancing the image resolution of a low-resolution (LR) observation to obtain a high-resolution (HR) image. Based on the number of available LR images, SR algorithms can be classified into two types: single reference SR (SRSR) and multiple reference SR (MRSR). MRSR algorithms perform SR by assuming that a set of LR images obtained from different viewpoints is available. However, in SRSR, only one image is available. Traditionally, image interpolation techniques such as nearest-neighbor interpolation, bilinear interpolation, and bicubic interpolation have been used. These techniques are known to work well in smooth regions but tend to blur edges and sharp details in the images [1]-[4]. Recently, many researchers have investigated SRSR techniques based on the discrete wavelet transform (DWT), which have proven to be superior to interpolation-based techniques.

This paper proposes an efficient SRSR system based on the DWT. As in previous works [1]-[4], we also assume that the LR image to be enhanced is the low-frequency subband of the HR image to be reconstructed. Our contribution is that we introduce the importance of edge directionality for exploiting the inter-subband correlation in the DWT domain. We have found that relative distribution of large coefficients in the high-frequency subbands is shown to be closely related to the edge strength and direction at a given position. Motivated by the success of the learning-based approach for SRSR [4], we design a support vector machine (SVM) performing SR based on the inter-subband correlation of the directional edges.

II. PROPOSED SYSTEM

The DWT decomposes an input image into multiple subbands: LL, LH, HL, and HH. Multi-level decomposition is achieved by obtaining a higher level decomposition from recursively applying the transform on the LL band of the lower level. In this paper, we denote the *k*'th row and the *l*'th column DWT coefficient of the LL, LH, HL, and HH bands of decomposition level s by $x_{LL}^{(s)}(k,l)$, $x_{LH}^{(s)}(k,l)$, $x_{HL}^{(s)}(k,l)$, $x_{HH}^{(s)}(k,l)$, respectively. Using this notation, we may formulate that the design problem for the SRSR system is to reconstruct the high-frequency subbands of HR image corresponding to $x_{LH}^{(1)}(k,l)$, $x_{HL}^{(1)}(k,l)$, and $x_{HH}^{(1)}(k,l)$, by assuming that the given LR image is $x_{LL}^{(1)}(k,l)$ which is equivalently a set of $x_{LL}^{(2)}(m,n)$, $x_{LH}^{(2)}(m,n)$, $x_{HL}^{(2)}(m,n)$, and $x_{HH}^{(2)}(m,n)$.

Our interpretations and observations on DWT coefficients useful for the SRSR system design can be summarized as follows.

1) The energy of the DWT coefficients in the high-frequency subband is concentrated in the edge region and the DWT coefficients are large in magnitude in this region. The magnitude of the DWT coefficients tends to decay across scales and the inter-subband correlation of small DWT coefficients is negligible [1]. Therefore, it is sufficient for the SRSR to reconstruct the DWT coefficients only in the edge region of the high-frequency subband of the HR image.
2) The relative distribution of the large coefficients in the LH, HL, and HH subband is closely related to the edge strength and direction at a given position. Therefore, reconstructing the high-frequency DWT coefficients based on the edge strength and direction information at a given position is highly recommended.
3) The inter-subband correlation between the DWT coefficients in the mother and child bands is almost negligible when the distance is greater than 2 pixels in the mother band [2]-[4]. Therefore, it is sufficient for the SRSR to reconstruct the DWT coefficient in the child band by only using its corresponding coefficient and the 8 neighboring coefficients in the mother band.

This then allows an SRSR system for reconstructing an image that is twice the size of the original image to be proposed. The basic structure of the system is described as follows.

1) Assume that the given LR image to be enhanced is the low-frequency subband of the HR image to be reconstructed and denote the LR image as $x_{LL}^{(1)}(k,l)$.
2) Perform one-level DWT on $x_{LL}^{(1)}(k,l)$ and obtain the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

TABLE I
PSNR COMPARISONS WITH OTHER TECHNIQUES

Techniques	Lena	Elaine	Baboon	Peppers
Bilinear	30.84	31.10	22.44	29.35
Bicubic	31.15	31.30	22.64	29.41
Wavelet-0	35.92	33.45	24.08	32.27
Temizel and Vlachos [2]	35.39	33.40	24.52	34.46
Proposed	37.23	33.63	26.04	34.71

DWT coefficients. Denote the resulting coefficients in the LH, HL, and HH bands by $x_{LH}^{(2)}(m,n)$, $x_{HL}^{(2)}(m,n)$, and $x_{HH}^{(2)}(m,n)$, respectively.

3) Perform edge detection on $x_{LL}^{(1)}(k,l)$ and denote the edge strength and direction by $|e(k,l)|$ and $\angle e(k,l)$, respectively.

4) Generate the high-frequency subbands of the HR image, $x_{LH}^{(1)}(k,l)$, $x_{HL}^{(1)}(k,l)$, and $x_{HH}^{(1)}(k,l)$, by using the SVM whose input vector is { $|e(k,l)|$, $\angle e(k,l)$, $x_{LH}^{(2)}(m,n)$, $x_{HL}^{(2)}(m,n)$, $x_{HH}^{(2)}(m,n)$; $\lceil k/2 \rceil - 1 \le m \le \lceil k/2 \rceil + 1$, $\lceil l/2 \rceil - 1 \le n \le \lceil l/2 \rceil + 1$ }, where $\lceil z \rceil$ is the maximum integer value not greater than *z*.

5) Perform inverse DWT on { $x_{LL}^{(1)}(k,l)$, $x_{LH}^{(1)}(k,l)$, $x_{HL}^{(1)}(k,l)$, and $x_{HH}^{(1)}(k,l)$ } to obtain the SR image.

III. EXPERIMENTAL RESULTS AND CONCLUSIONS

In our experiments, we have used the well-known Daubechies 9/7 analysis-synthesis filters for the implementation of the DWT. For the implementation of edge detection, we have used Sobel's edge detector. In order to train our SVM to be least dependent on selected image samples, we have randomly collected 200 images with sizes larger than 512×512. These sample images include a variety of indoor and outdoor shots of buildings, streets, people, faces, animals, landscapes, etc. taken from cellular phones, digital cameras or camcorders, HDTV, and CCTV. For simple implementation and fast processing, we have quantized the input and output vectors of the SVM: $|e(k,l)|$, $x_{LH}^{(2)}(m,n)$, $x_{HL}^{(2)}(m,n)$, $x_{HH}^{(2)}(m,n)$, $x_{LH}^{(1)}(k,l)$, $x_{HL}^{(1)}(k,l)$, and $x_{HH}^{(1)}(k,l)$, to the integer values. We have also quantized the edge direction $\angle e(k,l)$ to 36 levels. No detectable degradation of the performance was observed as a result of this quantization.

For the performance tests, we have used four well-known test images: Lena, Elaine, Baboon, and Peppers, which were not included in the sample set for the SVM training. All the images are 512×512. For objective tests, we have performed one-level DWT on the test images and assumed the LL band images of 256×256 size as the given LR images for a 2× resolution enhancement. We have used the original images as the ground truth and compared the performance in terms of the peak signal-to-noise ratio (PSNR) metric. The PSNR results are listed in Table I. We compare our results with two popular interpolation techniques: bilinear interpolation and bicubic interpolation. For comparison purposes only, the PSNR results obtained by Temizel and Vlachos [2] that are, to the authors' knowledge, regarded as one of the best set of results among the DWT based SR systems are also listed. In Table I, 'Wavelet-0' refers to the basic reconstruction method where unknown high-frequency subbands: $x_{LH}^{(1)}(k,l)$, $x_{HL}^{(1)}(k,l)$, and $x_{HH}^{(1)}(k,l)$, are estimated as zeros. The results in Table I show that the proposed SRSR system gives an improved metric compared with the other techniques for all test images.

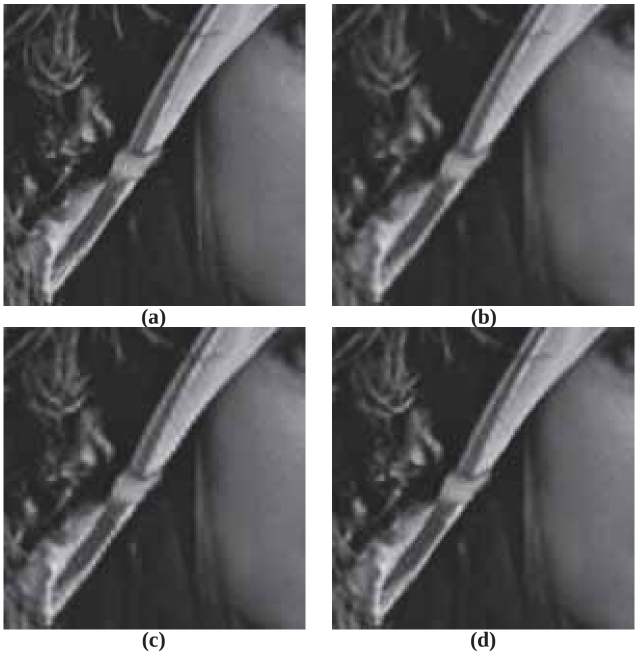

(a) (b) (c) (d)

Fig. 1. Enlarged part of reconstructed HR image: (a) the original image and the image reconstructed by (b) bicubic interpolation, (c) 'Wavelet-0', and (d) the proposed technique.

For subjective tests, we present the experimental results of the Lena image. Results for the other images were similar. For the purpose of better display, we show the enlarged part of reconstructed HR image in Fig. 1. We may easily notice that the proposed system improves the blurry and ringing distortions seen in the images reconstructed by the conventional bicubic interpolation method and the 'Wavelet-0' method.

REFERENCES

[1] W. K. Carey, D. B. Chuang, and S. S. Hemami, "Regularity-preserving image interpolation," *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1293-1297, Sep. 1999.

[2] A. Temizel and T. Vlachos, "Wavelet domain image resolution enhancement", in *Proc. IEE Vis. Image Signal Processing*, vol. 153, no. 1, pp. 25-30, Feb. 2006.

[3] S. Kim, W. Kang, E. Lee, and J. Paik, "Wavelet-domain color image enhancement using filtered directional bases and frequency-adaptive shrinkage, " *IEEE Trans. Consumer Electron.*, vol. 56, no. 2, pp. 1063-1070, May 2010.

[4] P. P. Gajjar and M. V. Joshi, "New learning based super-resolution: use of DWT and IGMRF prior," *IEEE Trans. Image Processing,* vol. 19, no. 5, pp. 1201-1213, May 2010.

Super-resolution From Digital Cinema to Ultra High Definition Television Using Image Registration of Wavelet Multi-scale Components

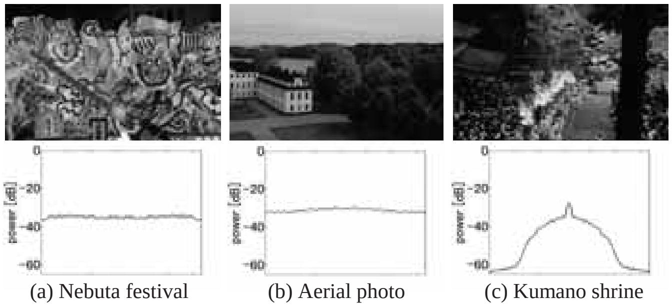
Yasutaka MATSUO[1,2], *Member, IEEE*, Shinya IWASAKI[1], Yuta YAMAMURA[1], and Jiro KATTO[1], *Member, IEEE*
[1]Waseda University, Tokyo, Japan, and [2]Japan Broadcasting Corporation (NHK), Tokyo, Japan

***Abstract*—Quality of image super-resolution (SR) degrades by noise component. However it should not be eliminated because it is important for high definition impression for digital cinema. Therefore we propose an image SR method by synthesis of resolution-enhanced signal and noise components respectively after dividing an original image into signal and noise components. The signal component is resolution-enhanced using image registration between the signal component and its wavelet multi-scale components with resolution-enhanced parameter optimization.**

I. INTRODUCTION

We study an appropriate image SR technique from digital cinema with 4K horizontal pixels [1] to future TV broadcast with 8K horizontal pixels [2]. Digital cinema image has the film grain noise or the thermal noise due to CMOS sensor characteristics. These noises degrade image quality for SR. However, these noises should not be eliminated because they are important for high definition impression as some researches had referred to [3]. As the state-of-the-art image SR method, wavelet SR [4], total variation interpolation [5], and multi-frame registration [6] are known. Previous wavelet SR and total variation interpolation methods cannot generate correct resolution-enhanced components because they generate resolution-enhanced components by applying low-pass oriented linear or nonlinear filtering in a single frame. Multi-frame registration methods can generate correct resolution-enhanced components if registration is precisely done with high accuracy. However noises degrade SR image quality because they impair the registration accuracy. In addition, digital cinema image has low inter-frame correlation because it has large motion blur by a low frame rate of 24 fps against a high spatial resolution of 4K horizontal pixels.

In this paper, we analyze power spectrum of noise component of digital cinema images in Sec. 2. Based on this analysis, we propose a novel image SR method by synthesizing resolution-enhanced signal and noise components respectively after dividing an original image into signal and noise components in Sec. 3. The signal component is resolution-enhanced using image registration between the signal and its wavelet multi-scale components on the assumption that a high resolution image has self-similarity in a single frame. Then resolution-enhanced parameters are optimized by minimizing differences between the signal component and the resolution-reduced signal components for SR image based on the Double Interpolation PSNR (DI PSNR) [7]. Experimental results of the proposed method are shown in Sec. 4. Finally, the conclusion of this paper is described in Sec. 5.



(a) Nebuta festival (b) Aerial photo (c) Kumano shrine

Fig. 1 Three monochrome test sequences and its noise spectrums.
(Horizontal axis is frequency. DC component is center of horizontal axis.)

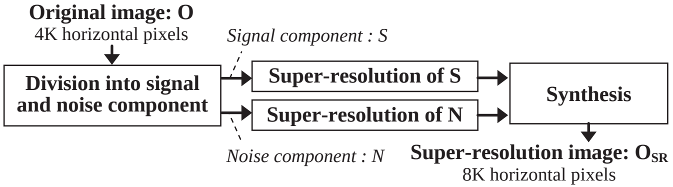


Fig. 2. Overview diagram of our proposed method.

II. SPECTRUM ANALYSIS OF DIGITAL CINEMA NOISE

We carried out spectrum analysis of digital cinema noises by using 128 points one-dimensional discrete Fourier transform (1D DFT) with the hamming window and 64 pixel shifts, which is applied to a cutoff image of flat areas. Fig. 1 shows three monochrome test sequences and their noise spectrum analysis results. Fig. 1 (a) is a shot of a CMOS sensor camera with 4K horizontal pixels. Figs. 1 (b), (c) are shots with 4K horizontal pixels achieved by digitizing 65 mm film and high sensitivity 65mm film, respectively. We can notice that noise frequency spectrums of Figs. 1 (a), (b) are nearly white but that of Fig. 1 (c) is not white. In this case, coarse granularity of the high sensitivity film is thought to be the cause of the non-whiteness. In general, low-frequency power spectrum of signal components is higher than high-frequency ones. Therefore, the more an image has white noises, the more conventional SR methods degrade their image quality because signal components are buried in noise components in high-frequency band. Therefore, we propose a SR method which divides an original image into signal and noise components, to each of which SR is applied .

III. PROPOSED METHOD

Fig. 2 is an overview diagram of our proposed method. Hereinafter, details of block diagram in Fig. 2 is explained.

A. Division into signal and noise components

In this block, white noises are divided by using spatial wavelet decomposition. First, an original image is decomposed by 1-level spatial wavelet decomposition without

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

decimation filtering. And then, white noise component N is extracted by wavelet reconstruction by using isolated elements in spatial high frequency HH^1 band because it has spatial low correlation as white noise. Finally, signal component S is extracted by wavelet reconstruction by using residual elements except isolated elements in HH^1 band.

B. Super-resolution of signal component

Fig. 3 depicts details of this block. The “n-level spatial wavelet decomposition of S without decimation” block extracts spatial low or high frequency components LL^n, LH^n, HL^n, and HH^n. In “Registration of wavelet multi-scale component” block, 1-level spatial high frequency components LH^1, HL^1, and HH^1 are set to resolution-enhanced components sLH, sHL, and sHH of S at first. Next, LL^n is divided 4x4 size blocks. And then, motion vector of each blocks are calculated to S by the full search block matching with the SSD parabola fitting. And then, sLH, sHL, and sHH are reconstructed by the MAP (Maximum A Posteriori) method after registration from LH^n, HL^n, and HH^n to sLH, sHL, and sHH using motion vectors of each block in LL^n. The “Bilateral filtering” block convolves sLH, sHL, and sHH using Gaussian function with standard deviations $\{\sigma_{LH}, \sigma_{HL}, \sigma_{HH}\}$ and gains $\{\gamma_{LH}, \gamma_{HL}, \gamma_{HH}\}$. The “Wavelet reconstruction” block extracts resolution-enhanced signal component S_{SR} using S, sLH, sHL, and sHH. The “Parameter optimization” block optimizes the $\{\sigma_{LH}, \sigma_{HL}, \sigma_{HH}\}$ and $\{\gamma_{LH}, \gamma_{HL}, \gamma_{HH}\}$ by calculating the smallest difference of DI PSNR. DI PSNR calculates by sum of differences between S and resolution-reduced signal component of S_{SR} which are extracted by applying 4:1 pixel resampling with 1 pixel phase shift for horizontal, vertical, and diagonal directions respectively. The output S_{SR} is resolution-enhanced signal component with optimized parameter.

C. Super-resolution of noise component and synthesis of resolution-enhanced signal and noise components

Resolution-enhanced noise component N_{SR} is generated by 1-level spatial wavelet reconstruction using spatial low frequency component of N and high frequency component of 0. And then, SR image O_{SR} is output by $S_{SR}+N_{SR}$.

IV. EXPERIMENTS

Test sequences are 2K horizontal pixels which are reduced-size 4K horizontal pixels of Figs. 1. The low-pass filter uses a 51-tap linear filter. Test sequences are resolution-enhanced to 4K horizontal pixels. The conventional and state-of-the-art SR methods uses below.

(A) Bi-cubic interpolation
(B) Total variation interpolation [5]
(C) Wavelet domain zero padding (WZP)
(D) WZP and cycle-spinning and edge rectification [4]
(E) Multi-frame registration [6]
(F) The proposed method

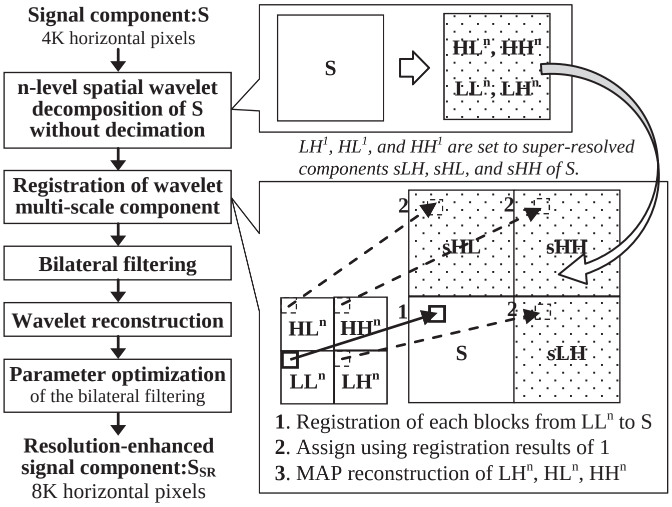

Signal component:S
4K horizontal pixels
n-level spatial wavelet decomposition of S without decimation
Registration of wavelet multi-scale component
Bilateral filtering
Wavelet reconstruction
Parameter optimization
of the bilateral filtering
Resolution-enhanced signal component:S_{SR}
8K horizontal pixels
S
HL^n, HH^n
LL^n, LH^n
LH^1, HL^1, and HH^1 are set to super-resolved components sLH, sHL, and sHH of S.
2
2
2
1
sHL
sHH
S
sLH
HL^n
HH^n
LL^n
LH^n
1. Registration of each blocks from LL^n to S
2. Assign using registration results of 1
3. MAP reconstruction of LH^n, HL^n, HH^n

Fig. 3. Details of super-resolution of signal component.

TABLE I
PSNR RESULTS (SUPER-RESOLUTION FROM 2K TO 4K HORIZONTAL PIXELS)

Sequences in Fig. 1	Image super-resolution methods in Sec. IV (A)	(B)	(C)	(D)	(E)	(F)
(a) Nebuta fest.	20.33	20.61	20.45	20.73	20.41	21.54
(b) Aerial photo	31.35	31.46	31.13	31.36	31.35	32.34
(c) Kumano sh.	34.69	37.39	37.01	37.50	37.82	38.13

(dB)

In (F), the number of taps in Gaussian filtering is five, standard deviations are set from 0.1 to 5.1 with 1.0 steps, and gains are set from 0.1 to 2.1 with 0.5 steps by exploratory experiment. The PSNR results are shown in Table I. We can confirm that PSNR value of the proposed method (F) is higher than any other methods.

V. CONCLUSION

We have proposed an image SR method by dividing an original image into signal and noise components in order to eliminate the effect of noises caused by conventional SR. Signal components are resolution-enhanced by image registration which minimizes differences between signal components and their wavelet multi-scale components with parameter optimization. Experimental results showed that the proposed method has an objectively better PSNR measurement than the conventional and state-of-the-art image SR methods.

REFERENCES

[1] Digital Cinema Initiatives, LLC: “Digital cinema system specification, version 1.2,” Mar. 2008.

[2] Y. Shishikui, Y. Fujita, and K. Kubota, “Super hi-vision - the star of the show!,” *EBU Technical Review*, pp. 4-16, Jan. 2009.

[3] M. Schlockrman, S. Wittmann, T. Wedi, and S. Kadono, “Filmgrain coding in H.264/AVC,” *JVT of ISO/IEC MPEG & ITU-T VCEG*, JVT-I034, Sep. 2003.

[4] T. Temizel and T. Vlachos, “Image resolution upscaling in the wavelet domain using directional cycle spanning,” *Journal of Electronics Imaging*, vol. 14, no. 4, 2005.

[5] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Total variation super resolution using a variational approach,” *Proceedings of IEEE ICIP*, pp.641-644, 2008.

[6] D. Capel: “Image mosaicing and super-resolution,” *Springer*, 2004.

[7] T. Saito, K. Ishikawa, and T. Komatsu, “Superresolution interpolation with a quasi blur-hypothesis,” *Proceeding of IEEE ICIP*, pp.1169-1173, 2011.

Catadioptric All-Focused Imaging Method Based on Coded Aperture

Yongle Li, Yu Liu, and Maojun Zhang
College of Information System and Management
National University of Defense Technology, Changsha, China

***Abstract*--The problem of catadioptric imaging defocus blur, which is caused by lens aperture and mirror curvature, becomes more severe when high resolution sensors and large apertures are applied. In order to overcome this problem, this paper proposes a simple modification to a conventional catadioptric system for the restoration of an all-focused catadioptric omni-directional image. The modification is designed by inserting a patterned occluder within the aperture of the camera lens, creating a coded aperture. Comparing to the conventional aperture, the coded aperture techniques identify the blur scale easier and more accurate. The restored all-focused image eliminates the defocus blur and shows the efficiency of the approach.**

I. INTRODUCTION

Catadioptric imaging systems consisting of conventional cameras and curved mirrors for capturing 360^o field of view owing to the advantage of one-shot seamless panoramic imaging, are widely used in many vision applications, such as robot navigation, surveillance, video conferencing. Most of the prior work in this field has focused either on aspects of mirror design, resolution enhancement, or stereo vision etc. In contrast, little attention has been paid to the aspect of obtaining sharp/all-focused images. Nevertheless, the problem of catadioptric imaging defocus blur, which is caused by lens aperture and mirror curvature, becomes more severe when high resolution sensors and large apertures are applied. S. Baker et al. [1] proposed detailed analysis of the defocus blur caused by the use of a curved mirror [2] in a catadioptric sensor. R. Swaminathan [3] presented a framework to derive the optimal focal plane at which to focus the lens for sharp image formation. Sujit Kuthirummal [4] employed iterative Richardson Lucy algorithm with a piece-wise constant interpolation of the blur kernels and total variation prior for minimizing artifacts. Weiming Li et al. [5] stated that the shapes of the best focused image regions in multi-focal omni-directional images can be modeled by a series of neighboring concentric annuluses based on an analysis on the spatial distribution property of virtual features. This paper draws inspiration from coded aperture imaging [6, 7] which employs deblurring algorithm to obtain the all-focused image. The main contribution of this paper can be considered as the introduction of the coded aperture techniques to the catadioptric imaging system for obtaining all-focused omni-directional images. The principle of our approach is to measure the effect of defocus so that we can estimate the amount of defocus easily, and create all-focused images.

II. ANALYSIS OF CATADIOPTRIC IMAGING DEFOCUS

Fig. 1 shows an illustration of cartesian frame *ROZ* where *o* is the location of the effective viewpoint which also is the origin and one focus of hyperboloid mirror, p_0 is the location of the effective pinhole which also is the other focus of hyperboloid. Suppose that one ray of light from *w* which is reflected by the mirror at m_1 passes through the point p_1 on the lens, and finally arrives at the image plane at the point w_{i1}. Similarly, another ray of light from *w* which is reflected by the mirror at m_2 passes through the point p_2 on the lens, and finally arrives at the image plane at the point w_{i2}. In general, these two rays of light will not be imaged at the same point on the image plane as the principal ray. Then the defocus blur is occurred [1]. The defocus simulation results and the real catadioptric imaging defocus blur obtained by the system equipment utilized in this work are shown in Fig. 2 and Fig. 3 respectively. Obviously, the omni-directional image is focused in the center of the image and blurred in the periphery, or blurred in the center of the image and focused in the periphery.

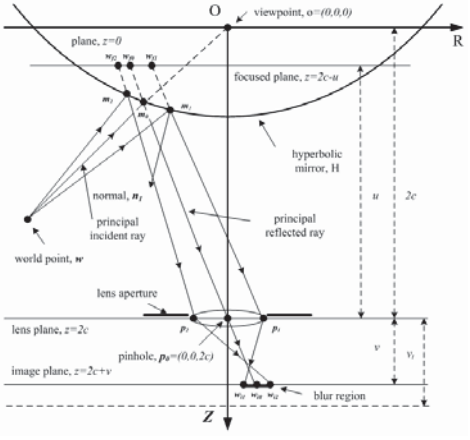


Fig. 1. Defocus blur analysis of catadioptric imaging system

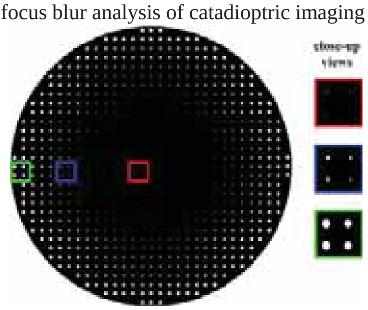


Fig. 2. Simulation results of catadioptric imaging defocus in ZEMAX.

This research was partially supported by National Natural Science Foundation of China under project No.61175006, No.61175015, No.61271438, and No.61275016; iCORE (Alberta Innovates) Alberta, and NSERC Canada.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

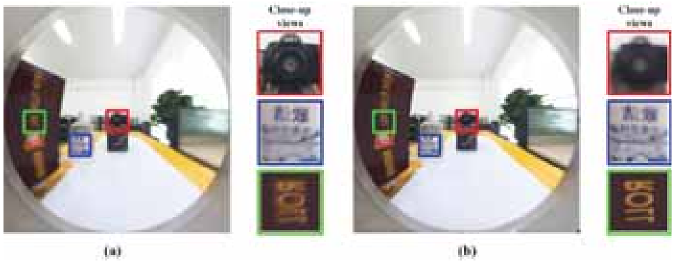


Fig. 3. Actual defocus blur images of catadioptric imaging system. (a)the center of image is sharp and periphery is blurry; (b) the center of image is blurry and periphery is sharp.

III. CATADIOPTRIC ALL-FOCUSED IMAGING METHOD

A. Best Focused plane for the lens

Ideally, in order to minimize the loss of high frequencies one would like to focus the lens such that all PSFs are as compact as possible, so that deconvolution has to do the least work. This would imply that it would be best for the lens to be focused somewhere in-between focusing on reflections in the center and reflections at the edges of a captured image. Naturally, to save more high frequencies information in the omni-directional images, R. Swaminathan's method [3] was applied to obtain the best focused plane for the lens in our framework.

B. Concentric Annuluses Division

The blur scale of the omni-directional image strongly related to the angle. If the incidence angles of the light rays are different, they have different blur scales. On the contrary, the same incidence angles have the same blur scale for the light rays, and the same incidence angles of light rays form an annulus [5]. So we divide the image into annuluses for restoration. When the division is finished, we can do the deconvolution deblurring for each annulus image individually.

C. Blur Scale Identification Using Coded Aperture

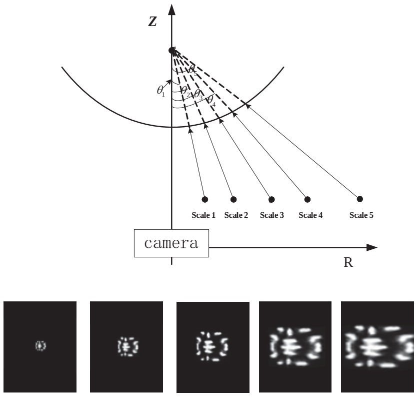


Fig. 4. Sketch map of blur scale identification in ZEMAX. (a) shows 5 different incident angles and 5 different scales; (b) shows the 5 different PSFs capture by the camera.

The probability model introduced in A. Levin et al. [6] allows us to detect the correct blur scale for each annulus. The correct scale should, in theory, be given by the model and suggesting the most likely explanation: The scale computed from the probability model is the blur scale of the corresponding annulus. Fig. 4 shows the sketch map of blur scale identification in ZEMAX.

D. Deconvolution Deblurring

After the correct blur scale of an observed annulus image is identified, the next step is to remove the blur, and to obtain the all-focused image. This task is also known as deblurring or deconvolution. Fig. 5 shows the all-focused image computed from PSFs using the method K. Dabov et al. [8], in which the entire scene appears sharp and well focused.

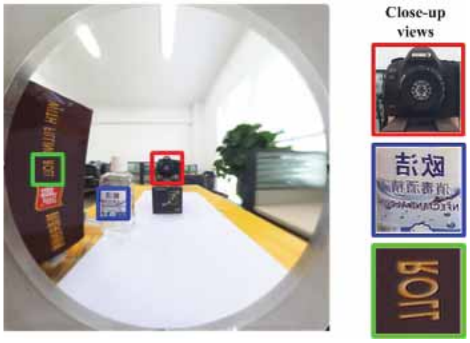


Fig. 5. All-focused image

IV. CONCLUSION

In this work, a catadioptric all-focused imaging method is proposed. We have analyzed the defocus blur, and applied coded aperture technique to solve the catadioptric defocus problem inspired by the idea of computational photography. This approach do not require complex experimental appliance, however, it produces good results. The 360-degree panoramic imaging device can be used in a remotely-piloted Unmanned Surface Vehicle (USV) application [9]. Since this work only focuses on the method of the restoration of defocus blurry image, the strategy of image division and the annulus image stitching will be considered as our future work.

REFERENCE

[1] S. Baker, S. K. Nayar, "A Theory of Single-Viewpoint Catadioptric Image Formation," Int. J. Comput. Vision, 35(2):1–22, 1999.

[2] E. Hecht, "Optics," Addison-Wesley. 4th edition, 2003.

[3] R. Swaminathan, "Focus in Catadioptric Imaging Systems," in *Proceedings of International Conference on Computer Vision*, 2007.

[4] S. Kuthirummal, "Flexible Imaging for Capturing Depth and Controlling Field of View and Depth of Field," America: Columbia University, 83-101, 2009.

[5] W. Li, Y. F. Li, Y. Wu, "A Model Based Method for Overall Well Focused Catadioptric Image Acquisition with Multi-focal Images," in *Proceedings of International Conference on Computer Analysis of Images and Patterns*,, 460-467, 2009.

[6] A. Levin, R. Fergus, F. Durand, W. T. Freeman, "Image and Depth from a Conventional Camera with a Coded Aperture," ACM SIGGRAPH, Aug. 2007.

[7] C. Zhou, S. Lin, S. K. Nayar, "Coded Aperture pairs for Depth from Defocus and Defocus Deblur," Int. J. Comput. Vision, 93:53–72, 2011.

[8] K.Dabov, A.Foi, V.Katkovnik, and K. Egiazarian, "Image restoration by sparse 3D transform domain collaborative filtering," in *Proceedings of the International Society for Optical Engineering*, SPIE Electronic Imaging, 2008.

[9] http://www.remotereality.com/

Resource-usage modes detection for run-time resource prediction of video components

Ionut David, Martijn M.H.P. van den Heuvel, Rudolf H. Mak and Johan J. Lukkien

Department of Mathematics and Computer Science, Eindhoven University of Technology,
P.O. Box 513, 5600 MB, Eindhoven, the Netherlands

Abstract— **Consumer electronic products become increasingly more open and flexible, which is achieved by scalable and re-usable software components. Important features in many of those products are provided via video components. Since these components have fluctuating resource usage, run-time resource-prediction strategies are required to enable cost-effective resource allocations. In this paper, we propose a novel method to identify *resource-usage modes* for video components.**

I. Introduction

Video components typically impose real-time constraints with highly transient variations in the processing of their stream rendering. To enable cost-effective video processing, scalable video components have been conceived [2], [5], which allow trading resource usage against output quality at the level of individual frames. Because a quality level (mode) can only be changed at the level of individual frames, a frame is entirely processed in a particular mode, for which processing resources may or may not be sufficient. In the latter case, for cost-effectiveness reasons, a work-preserving approach is often taken in which the processing of the current frame is completed and a next frame is skipped [5]. Allocating a static amount of processing resources to a video component is therefore unsuitable, because it leads either to frame misses or to an over-provisioning of resources.

The fluctuating resource requirements of video components make it difficult to allocate resources efficiently. This problem has received much attention; for example, see [2] and [5]. The key technology to enable dynamic resource allocation is predicting the resource usage of a component. In this paper, we present a method based on a new model of resource states [3]. With these states we construct a utilization profile of each executing video component by run-time monitoring. We observe that these utilization profiles typically have clustered states, which can be used to define resource-usage modes.

Our main contribution is a method to detect *resource-usage modes* suitable for run-time resource prediction. First, we define and identify experimentally a set of *resource-usage modes* of two example video components. Secondly, we introduce a method to detect these resource-usage modes using a peak-finding mechanism. Finally, we evaluate the accuracy of our detection process.

II. Related work

According to a classification by Koziolek [1], we follow a stochastic approach to performance evaluation of components. Our method for monitoring and prediction enables adaptive (mode-based) resource allocations, e.g., as by Wüst et al. [5].

Similar to [2], we monitor the processor utilization of video applications with instrumented code. The difference with our work is that they evaluated a set of proprietary video algorithms with a novel, flexible abortion strategy whereas we evaluate the open-source and widely-used OpenCV library [4] with work-conserving video components.

III. From resource-usage states to modes

In previous work [3], we define a state-based resource model that captures the processor utilization of a video component in the following way. The utilization range [0,1] is divided into n equal-sized sub-ranges, to which we refer as (utilization) *states*. For any utilization value $u_i \in [0,1]$, its state $\sigma(u_i)$ is defined as the rank of the utilization interval to which it belongs, i.e., $\sigma(0)$=1 and $\sigma(u_i) = \lceil nu_i \rceil$, for $u_i \neq 0$. Furthermore, we define the (normalized) state frequency $f_s = \#\{s \mid \sigma(u_i) = s \wedge 0 \leq i < N\}/N$, where N is the total number of monitored video frames. Frequency f_s represents the fraction of video frames whose processing require a utilization in the range of state s.

In this work we show that, in practice, videos often exhibit a state-frequency distribution in which a few clusters of densely populated states can be discerned. This observation leads us to define so called *resource-usage modes.* Such a mode is characterized by two parameters: λ expressing a predefined threshold frequency and μ representing the number of the states contained in the mode. More precisely, a (λ, μ)-mode is defined by at most μ consecutive states satisfying $f_s > \lambda$. Since there are far fewer modes than states, and since sojourn times are longer for modes than for states, prediction of modes is expected to be more accurate than prediction of states. Moreover, if modes are pronounced (large λ, small μ), mode-based instead of state-based prediction is expected to yield similar resource usage.

IV. Experiments and results

In this section, we first experimentally confirm the presence of resource-usage modes. Next, we describe a method to detect those modes. Finally, we evaluate the detection method.

A. State-based resource monitoring

To demonstrate state-based resource monitoring, we use the OpenCV video player [4]. This video player performs two activities: decoding a video stream and displaying it. We disable frame buffering and hardware acceleration, so that the video player executes an alternating sequence of decoding and displaying steps on a frame by frame basis.

Our experiments are conducted on 10 movies with various content and with different encodings, i.e., xvid, mpeg2, mpeg4.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

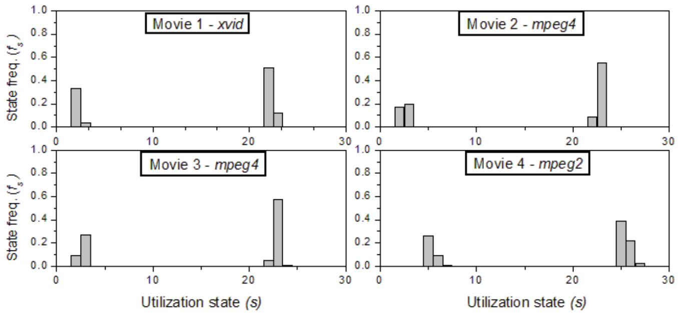


Fig. 1 Resource-usage modes of a video player.

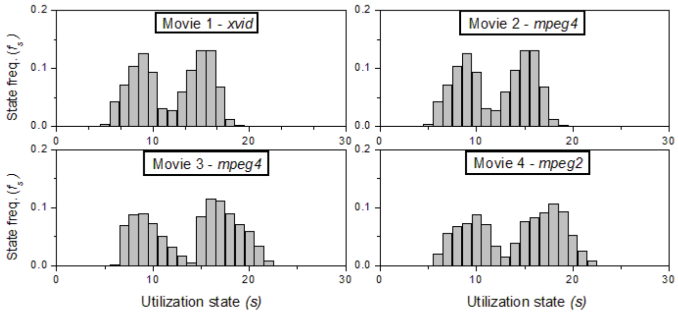


Fig. 2 Resource-usage modes of a motion-detection component.

Each movie has 97,500 frames and plays for approximately 1 hour at 25 frames per second (fps). For each of these movies, we monitor the CPU utilization of two video components on a frame-by-frame basis, distinguishing n=100 states. First, we use a simple video player. Subsequently, the same player is complemented with a motion-detection component.

Figure 1 and Figure 2 present the partial results (only 4 out of 10 movies shown) of these two experiments. In all histograms, one can recognize two peaks, each coinciding with our notion of a resource-usage mode. The remainder of our experiments produced a similar pattern.

B. *Detecting resource-usage modes from states*

We observe that, in our experiments, resource usage modes show up as peaks in the state-frequency curve. Mode detection therefore becomes a peak-detection process.

For example, consider the curve of Movie 1 in Figure 1. A standard local-maxima algorithm detects two peaks, viz., at state 2 and state 22. These two peaks indicate the presence of two resource-usage modes. As a result, a predictor has to transit between these two modes at runtime.

C. *Evaluating the accuracy of mode detection*

Finally, we measure the accuracy of our peak-finding mechanism using the statistical metrics *Precision, Recall* (for incorrectly and correctly detected modes respectively) and the efficiency of using those modes for resource allocation. Figure 3 presents the results of the peak-detection accuracy for the total input set of 10 different movies. It also shows the *F-measure* (harmonic mean of Precision and Recall).

During peak detection, the width μ of the peaks is ignored by the applied detection algorithm based on local maxima. Figure 3 shows the effect of varying the threshold, λ. On the one hand, choosing a large λ leads to a small number of accurately detected modes. This is desired for fast prediction at run-time, but a small number of modes implies a large state space contained in a single mode and, hence, a waste of resources. On the other hand, choosing a small λ may lead to a large number of detected modes, thereby reducing waste, but it also complicates their prediction (see the reduced precision of the motion-detection component).

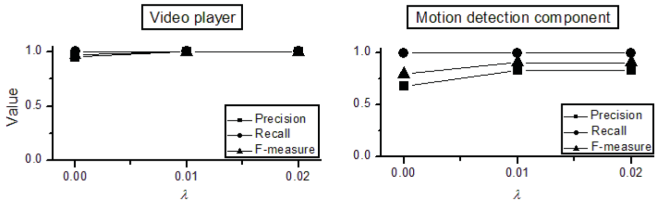


Fig. 3 Peak-detection accuracy measured over 10 example movies.

Looking at a single component distinguishing M modes, it is desirable that the majority of the processed frames is covered by the utilization states defining those M modes, i.e., $\sum_{1\leq x\leq M}\lambda_x\mu_x \geq 1-\epsilon$ where ϵ should ideally be close to zero. When this criterion is met, resource-usage prediction based on modes rather than prediction based on more volatile states is expected to yield both

(i) little waste of allocated resources and

(ii) a small number of skipped frames.

Considering the detected modes in Figure 3, we assumed that each video transits between at most $M = 2$ modes. We can control the number of states per mode, for example, by selecting $\mu = 5$ (for $\lambda = 0.02$), where μ bounds variations in utilization to 5% within a single mode. The video player processes 100% of the frames in the detected modes and the motion detector processes 85±6% (with a 95% confidence interval) of the frames in detected modes.

Although with a simple local-maxima algorithm we obtain already a high coverage of utilization states inside modes, mode detection may be more complex, in both dimensions λ and μ, for different video-component types.

V. Conclusions

This paper presents a method for monitoring and detecting resource-usage modes of video components. Modes are detected as peaks, i.e. densely populated clusters of states, in the state frequency profile of a video component. Typically, these profiles exhibit a small number of modes compared to the number of states. Hence, modes present a promising solution for run-time resource prediction in future consumer electronics.

References

[1] H. Koziolek, “Performance evaluation of component-based software systems: A survey”, *Journal of Performance Evaluation*, vol. 67, pp. 634-658, July 2009

[2] M.M.H.P. van den Heuvel, R.J. Bril, S. Schiemenz, and C. Hentschel, “Dynamic resource allocation for real-time priority processing applications”, *IEEE Transactions on Consumer Electronics (TCE)*, vol. 56, issue 2, pp. 879–887, May 2010

[3] I. David, R.H. Mak, J.J. Lukkien, “Resource-usage modeling for runtime resource management of component-based applications,” *Int. Conf. on Advances in Computer Science and Application*, June 2012

[4] OpenCV framework website. http://opencv.willow.garage.com/

[5] C. C. Wüst, L. Steffens, W. F. Verhaegh, R. J. Bril, and C. Hentschel, “QoS control strategies for high-quality video processing,” *Real-Time Systems Journal,* vol. 30, no. 1-2, pp. 7–29, 2005

Background Subtraction Using Edge Cues and Color Difference for Stabilized CMOS Images

Juhan Bae, Youngbae Hwang and Byeongho Choi
Multimedia IP Center, Korea Electronics and Technology Institute
Email: jhbae@keti.re.kr

***Abstract*—Video stabilization followed by background subtraction using CMOS sensor shows erroneous foregrounds. In this paper, we present a background subtraction method using edge cues and color difference to deal with preprocessed stabilized images. We use a non-parametric background model with edge cues based on the observation that interpolation errors are larger in textured regions than in homogeneous regions. By additional CIELab difference model, we show that misclassifications are dramatically decreased than conventional background subtraction methods.**

I. Introduction

Cameras using CMOS sensors often can exhibit unexpected skew and wobble due to unwanted shaking. Therefore, video stabilization is conditionally followed by background subtraction in practical surveillance systems to decrease false alarms by excluding the effect of environmental changes. They, however, still suffer from erroneous detection caused by image stabilization.

There have been various previous works that endeavor to obtain accurate foregrounds with suppression of false classification. Friedman and Russel [1] and Stauffer and Grimson [2] proposed Mixture of Gaussians that can be adapted for complex scenes such as streaming waters and waving trees. Elgammal et. al. [3] proposed a non-parametric approach to deal with uncertainties in the correct manner. These approaches, however, do not consider that the results of video stabilization can be degraded by perspective distortion, additional noises and rolling shutter effects.

In this paper, we propose a color difference background model with edge cues to reduce errors caused by stabilized CMOS images. Inhomogeneous regions such as edges are more sensitive to video stabilization than homogeneous regions. Background model with edge cues based on non-parametric approach is considered to minimize false alarms. The edge cues based on the non-parametric approach can minimize additional false alarms caused by image stabilization. In addition, we take into account CIELAB color difference that is known as close to human visual system.

II. Background Subtraction

Elgammal et. al. [3] proposed a non-parametric approach to model multi-modal probability density function. Let $I_1, I_2, \cdots, I_N$ be the recent intensity value of video sequences, K is kernel and N is the number of previous frames to estimate $P(I_t)$. If we use RGB color images and choose kernel estimator function K to be a Normal function $N(0, \sum)$, where $\sum$ represents the kernel function bandwidth, then the density estimation can be defined as

$$P(I_t) = \frac{1}{N} \sum_{i=t-N}^{t-1} \prod_{j=i}^{3} \frac{1}{\sqrt{2\pi(\sigma_j)^2}} e^{-\frac{1}{2}\frac{(I_{t,j}-I_{t,j})^2}{(\sigma_j)^2}} \tag{1}$$

when we assume that color channels are independent with a different kernel bandwidth σ_j^2 and the pair (I_i, I_{i+1}) is local-in-time.

We assume that video stabilization is followed by background subtraction. We propose a background model with edge cues of stabilized image sequences in order to suppress incorrect detection. The background model of edge cues is built based on a non-parametric approach [3]. If we define that $E_1, E_2, \cdots, E_N$ are the recent Sobel edge magnitude value of current stabilized image sequences, the density estimation of edge magnitude can be described to

$$P(E_t) = \frac{1}{N} \sum_{i=t-N}^{t-1} \frac{1}{\sqrt{2\pi(\sigma_E)^2}} e^{-\frac{1}{2}\frac{(E_t-E_i)^2}{(\sigma_E)^2}} \tag{2}$$

A pixel is labeled as edge foreground when $P(E_t)$ is smaller than a pre-defined threshold T_E.

Each RGB channel value is the sensor response of an image and has different noise characteristics. Estimation of each color channel bandwidth may be inaccurate and Hwang [4] shows that CIELAB color space has less noise effects than other color spaces. If $D_1, D_2, \cdots, D_N$ are defined as the recent CIELAB color difference between the key frame and stabilized image sequence, the density estimation of color difference can be also described as

$$P(D_t) = \frac{1}{N} \sum_{i=t-N}^{t-1} \frac{1}{\sqrt{2\pi(\sigma_D)^2}} e^{-\frac{1}{2}\frac{(D_t-D_i)^2}{(\sigma_D)^2}} \tag{3}$$

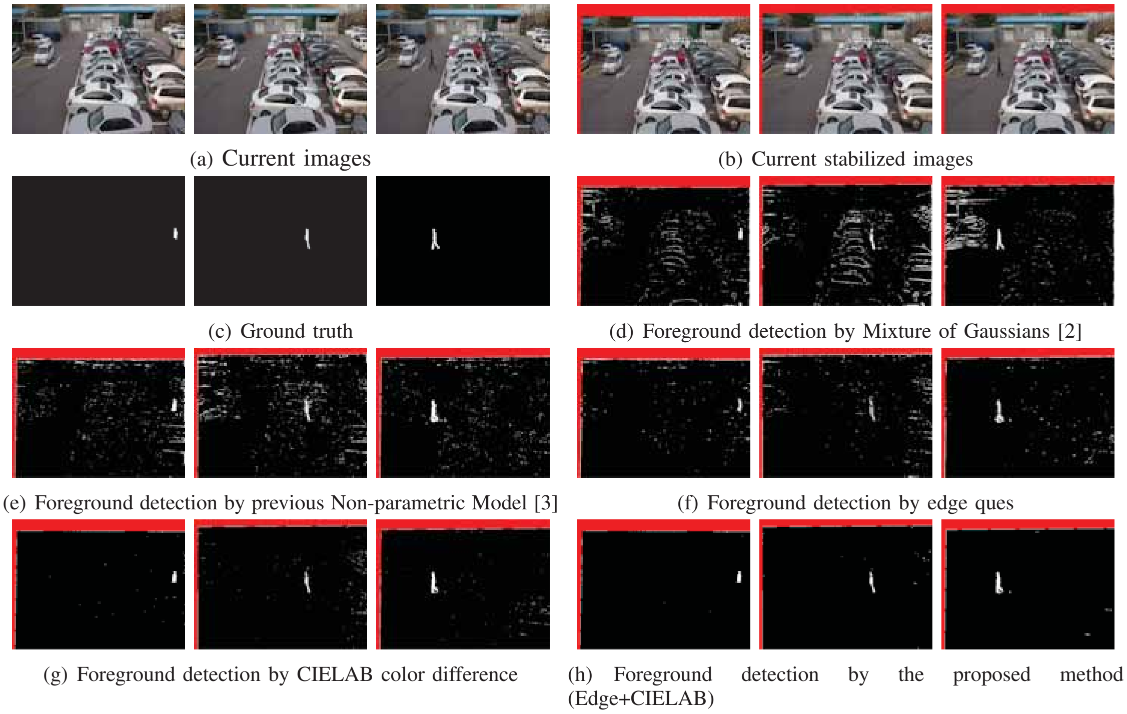
A pixel labeled as color difference foreground when $P(D_t)$ is smaller than pre-defined threshold T_D as well.

Finally a pixel is considered as part of the foreground only if $(P(E_t) > T_E) \wedge (P(D_t) < T_D)$. True positives will be decreased by taking intersection in the edge cues model. To cope with this, all pixels that are adjacent to the pixel detected by each short term model [3] are included in the final results.

III. EXPERIMENTAL RESULTS

We apply our algorithm to four outdoor video sequences (3 minutes each) that contain perspective distortion, small amount of rolling shutter effects using a Olympus PEN E-P3 camera with 640x480 resolutions. For the non-parametric

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a) Current images

(b) Current stabilized images

(c) Ground truth

(d) Foreground detection by Mixture of Gaussians [2]

(e) Foreground detection by previous Non-parametric Model [3]

(f) Foreground detection by edge ques

(g) Foreground detection by CIELAB color difference

(h) Foreground detection by the proposed method (Edge+CIELAB)

Fig. 1. Background subtraction results

background model, static 50 frames from hand-rolled camera were recorded without any moving object to represent the initial background model. In conventional background update method [3], the short-term model contains the most recent 50 background samples while long-term model contains 50 samples taken over a 1000 frame time window. The key frame is set to as the first frame without moving objects and video stabilization is performed by the warping current frame to the key frame based on the estimated projective matrix.

In Fig. 1(d) and Fig. 1(e), we show that video stabilization severely affects on wrong foreground parts compared to ground truth in Fig. 1(c). Edge cue model in Fig. 1(f) reduces undesirable foregrounds around edges effectively even though false negatives are slightly increased around object edges. CIELAB color difference model shows better suppression for improper foreground parts in Fig. 1(g). Moreover, the proposed method in Fig. 1(h) by color difference with edge cues effectively reduces unwanted foregrounds around edges and also decreases total error down to 31% of conventional non-parametric method [3] as shown in Table. I since color difference model works as complement of edge model.

IV. Conclusion

We present a background subtraction method for stabilized CMOS images. Video stabilization brings about erroneous foreground parts in conventional background subtraction. Edge cue model is considered for non-homogeneous regions and color difference model is applied for different color channel noise characteristics respectively. Consequently, the combination of edge cues and CIELAB color difference is effective for the elimination of unwanted foregrounds for stabilized CMOS images.

Method	False negatives	False positives	Total error
Mixture of Gaussian[2]	30/376 (7.98%)	2122/287465 (0.74%)	0.75%
Non-Parametric[3]	31/376 (8.24%)	1186/287465 (0.41%)	0.42%
Edge Cues	98/376 (26.06%)	908/287465 (0.32%)	0.35%
CIELAB Difference	10/376 (2.66%)	630/287465 (0.22%)	**0.22%**
Proposed (Edge+CIELAB)	9/376 (2.39%)	356/287465 (0.12%)	**0.13%**

TABLE I
The average ratio of false classification

Acknowledgment

This work was supported by the IT R&D program of MKE/KEIT. [10040246, Development of Robot Vision SoC/Module for acquiring 3D depth information and recognizing objects/faces]

References

[1] N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," in *Annual Conference on Uncertainty in Artificial Intelligence*, 1997, pp. 175–181.

[2] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 747–757, 2000.

[3] A. Elgammal, D. Harwood, and L. Davis, "Non-parameteric model for background subtraction," in *European Conf. on Computer Vision*, 2000, pp. 751–767.

[4] Y. Hwang, J.-S. Kim, and I.-S. Kweon, "Change detection using a statistical model in an optimally selected color space," *Computer Vision and Image Understanding*, vol. 112, no. 3, pp. 231–242, 2008.

Seamlessly Expanded Natural Viewing Area of Stereoscopic 3D Display System

Ungyeon Yang[1]†, Namkyu Kim[2], Jinseok Seo[2], and Ki-Hong Kim[1]
[1] *Electronics and Telecommunications Research Institute, Daejeon, Korea*
[2] *Dong-Eui University, Busan, Korea*

Abstract —*Common 3D Stereoscopic displays have a convergence and accommodation (CA) conflict phenomena that interfere with the human's sense of visual system. These characteristics can cause from visual fatigues to safety problems when watching a 3D video, and thus human factors have emerged as a major research topic. Therefore, we propose a technique that can naturally expand the sense of 3D viewing space, and present the current status of our study.*

Index Terms — 3D Stereoscopic Display, Human Factors, Comfortable Viewing Zone, Mixed Display Platform

I. Introduction

Although three-dimensional (3D) technology has recently been the subject of public attention, negative human factors are gaining attention with questions regarding the safety of watching 3D content. As shown in Fig.1, when we want to experience a positive and a negative depth effect, the current 3D display technology has a fundamental disadvantage of a mismatch of the convergence and accommodation (CA) between two eyes. In previous study [2], we proposed a layered multiple display (LMD) architecture that fuses the visualization areas from homogeneous and heterogeneous displays into one connected space.

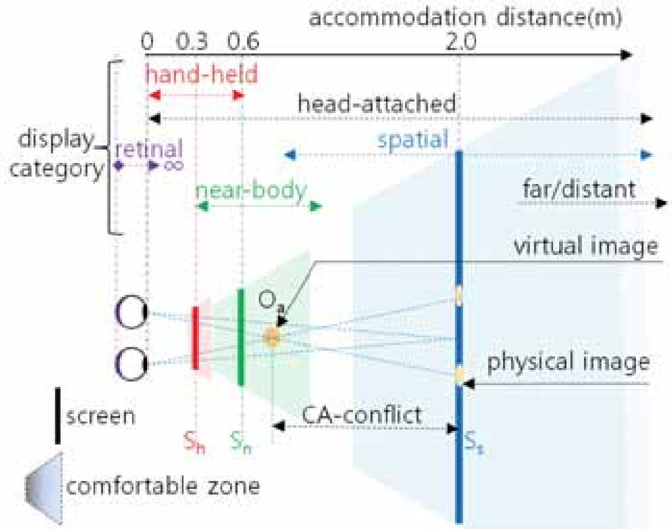


Fig. 1. Display continuum and example of CA-conflict

In this study, we suggest the use of expanded 3D (E3D) technique, which aims to develop a new platform that can naturally visualize and interact with virtual 3D objects around user's near-body space. Using a multiple-display platform, our technique complements a single 3D display, which has a limited comfortable 3D viewing zone [1].

II. Expanded 3D Display Platform

We can classify various displays as shown in Fig. 1 based on the screen distance from the user, and can also analogize the comfortable zone (CZ) of each display. By connecting the comfortable zone of each display, we can split and control the area of visualization space to preserve the natural 3D stereoscopic view around the user's performance area. The extreme effect of depth perception, which is impossible with a single device, and the expression of objects located outside the single display's view frustum can then be embodied. If the virtual object, O_a, in Fig. 1 is rendered with a spatial display, S_s users will experience excessive discomfort from a too significant CA-conflict. Therefore, in this case, if it is output in the CZ of the near-body display, S_n through a seamless transition, users will be able to experience a more natural 3D viewing.

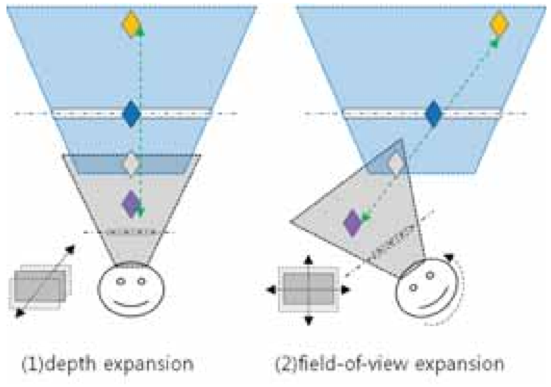


Fig. 2. Expanding natural viewing area

Figure 2 shows the generalized 3D space expansion model of the E3D platform.

III. Experimental Implementation and Discussions

Figure 3 shows an overview of the pilot system of the proposed idea to verify its feasibility. In the left side of the room is a virtual golf system, if a player stands in front of the hole cup, the front screen must output an excessive negative

This work was supported by the Ministry of Culture, Sports and Tourism (MCST) and Korea Evaluation Institute of Industrial Technology (KEIT) (10039923, Development of Live4D contents platform technology based on expansion of realistic experiential space)

† e-mail: uyyang@etri.re.kr

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

parallax or the virtual image of the hole cup may be out of the front screen's view frustum. We can then apply an eye glasses type display (EGD) for naturally showing what is in front of the user's feet. In the right side of the room is a mixed display environment with homogeneous and heterogeneous devices with multimodal interfaces and autostereoscopic displays [3]. A testing contents, a curling match in winter that three users can take part in and experience cooperative work and content-sharing scenarios.

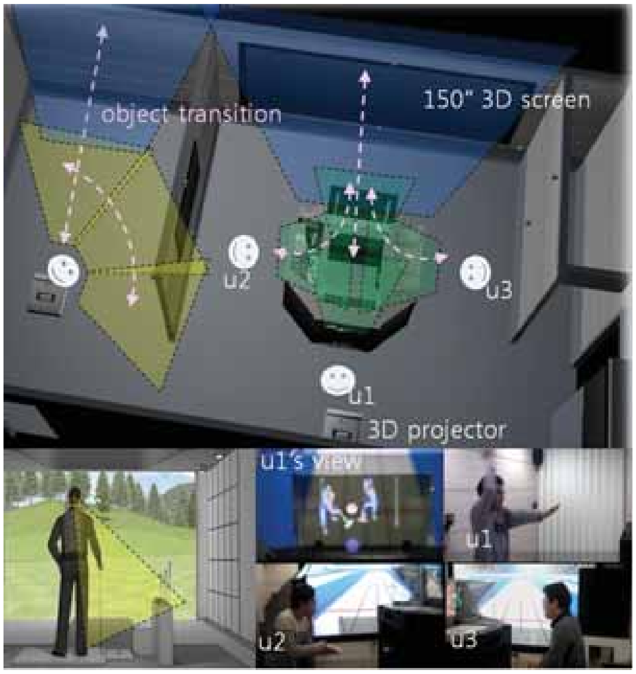


Fig. 3. The pilot testing contents of the E3D platform

To verify the feasibility of our idea, we implemented a virtual soccer stadium, as shown in Fig. 4. At the beginning of the experiment, the ball was displayed only on the 3D TV, and the display device for the ball switched to an HMD at a specific distance. The questionnaire included one question for each type of content, which asked the subjects to rate the unnaturalness of the perception. The 27 subjects were college students in their 20s. The ANOVA test showed that the unnaturalness of each of the five contents were significantly different (F = 3.732, $p < 0.007$), and the Tukey test results differed between group A (-2m, -1m, 0m) and group B (-2m, -1m, 1m, 2m). In particular, we ascertained that the subjects felt the unnaturalness in the contents that switched devices at 1 and 2 m distances from the zero parallax planes much more than at a 0 m distance. We can say that the results showed the feasibility of our proposed technology and the necessity for the method to optimize the perception in an expanded 3D display platform with mixed devices.

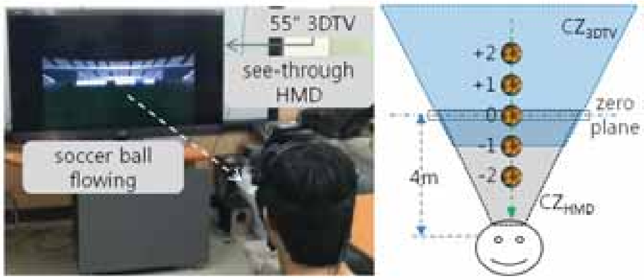


Fig. 4. Overview of the pilot experimental environment

The E3D system environment consists of various 3D stereoscopic display devices, and their 3D stereo contents should be systematically connected for a visual expansion. To give more natural stereoscopic perception to the viewers, calibration processes between the 3D display devices and viewers are required. The toe-in method is suitable to for the stereo rendering of near-focused objects, and an off-axis method is applied to far and wide space stereo rendering. Taking advantage of each method, the toe-in method is used for the EGD, and the off-axis method is for the 3DTV. Using the use of these heterogeneous rendering methods will be occurred in incur a matching error between the virtual objects on the EGD and 3DTV. To minimize this error, our E3D system has to consider calibration processes and its implementation. Currently, the conventional calibration method of augmented reality, SPAAM [4], is used for matching virtual objects on EGD and 3DTV. In the future, considering viewer's eye conversion capability and heterogeneous rendering methods, a calibration method will be proposed.

IV. Conclusion

In this paper, we proposed a heterogeneous 3D display mixing platform to solve human eye's conversion and accommodation conflict problem of popular 3D displays, and checked the possibility of its realization in the pilot test. We have a plan to develop a new wearable display device, which meets the proposed concept for expanding 3D viewing area. Also, to observe and reduce human factor issues in 3D stereoscopic viewing, various subject-group testing will be performed in our mixed stereoscopic display system.

References

[1] T. Shibata, J.H. Kim, D.M. Hoffman, and M.S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays", Journal of Vision, Vol.11, No.8, P1-29, 2011.

[2] G.A. Lee, U.Y. Yang, and W.H. Son, “Layered Multiple Displays for Immersive and Interactive Digital Contents”, ICEC, LNCS 4161, 2006.

[3] H.M. Kim, G.A. Lee, U.Y. Yang, T.J. Kwak, and K.H. Kim, “Dual Autostereoscopic Display Platform for Multi-user Collaboration with Natural Interaction” ETRI Journal, Vol. 34, No.3, P466-469,2012.

[4] M. Tuceryan, Y. Genc and N. Navab, “Single-Point Active Alignment Method (SPAAM) for Optical See-Through HMD Calibration for Augmented Reality”, Presence, MIT Press, Vol. 11, No. 3, P259-276, 2002.

Reducing 3D Visual Fatigue based on Salient Color Model

Ji Young Hong, Yang Ho Cho, Ho Young Lee, Du Sik Park, and Chang Yeong Kim
Advance Media Lab.
Samsung Advanced Institute of Technology

Abstract—**In this study, a 3D image processing method of improving the visual sense of depth in line with the human visual perception characteristics was developed. Some of the visual processes that occur in human visual perception were reproduced to estimate the interest area that would correspond to the visual selective attention through the salient color model, the depth was translated on the basis of the interest area, and the image quality was enhanced in tune with the 3D image characteristics. In this way, the perceptive sense of depth can be improved and fatigue can be decreased when watching a 3D display.**

I. Introduction

This visual fatigue, which is a major issue with 3D displays, is closely related to the sense of depth. The greater the sense of depth is, the more visual fatigue occurs [1], [2], [3]. Although many methods of addressing visual fatigue in 3D displays and improving satisfaction with the sense of depth have been proposed, no methods of simultaneously reducing visual fatigue and improving satisfaction with the sense of depth have been proposed.

In this study, the colors, shapes, motions, and depth that are used in the visual process in the brain were assumed as the basis for theoretical modeling that corresponds to selective visual attention [4]. Among them, color information, which has great influence on human visual perception in static images, and sense of depth information, which is a differentiator of 3D displays, were selected as stimulants. An algorithm for improving the perceptive sense of depth and reducing visual fatigue based on salient color model was developed to reproduce images that give a sense of depth and a sense of three dimensions while reducing visual fatigue.

II. The Algorithm

The designed technical algorithm for the reduction of visual fatigue and the enhancement of the sense of three dimensions is described as follows. A visual attention map that is adaptive to inputted 3D images was created on a 3D display on the basis of salient color model to certain human visual perception characteristics. The depth was readjusted on the basis of the interest area that was predicted with the visual attention map to improve or maintain the sense of three dimensions while reducing visual fatigue.

The technology for reducing visual fatigue and enhancing the sense of three dimensions that was developed in this study largely consists of the construction of a visual attention map, depth readjustment, and the creation of a 3D image with the application of the optimum depth. Among these components, the construction of a visual attention map divides the area into 1 to N using the depth information that represents the sense of depth when the depth information, which represents the color information and depth of the 3D images, is inputted. The weight value of the pixel depth is applied differently to each area that has been divided by depth, and is used to complete the visual attention map that corresponds to the selection visual attention.

For color information, R, G, and B are converted, based on the pixel value of the input image, into the J (lightness), C (chroma), and H (hue) of CIECAM02 and Munsell color wheel. The pixels that have the specified color property values are given a weight because they are regarded as having selective visual attention. For example, there are colors that catch the viewers eye very easily, which are referred to as accent colors [5]. It is determined if the value that corresponds to H is an accent color, and a weight is applied to the value that is found to be accent color. As with H, different weights are applied to the J and C values that corresponds to the specified thresholds to which selective visual attention is applied. Through this method, the pixels that correspond to each characteristic that is given selective visual attention have greater weights than other pixels, and the values calculated for each pixel are summed up in the Visual Attention Map Calculation, after which the visual attention map is constructed.

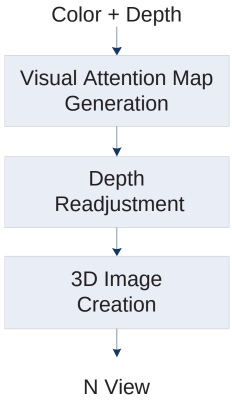


Fig. 1: Flowchart of the algorithm.

$$VAM = \frac{w_1 x_1 + w_2 x_2 + ... + w_n x_n}{w_1 + w_2 + ... + w_n} \tag{1}$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

The Depth Readjustment plays the role of deciding on the representative value so that depth can be readjusted using the depth of the completed visual attention map and the depth of the input image in the Interest Area Construction. There are many methods of setting the representative value for readjustment, but in this study, the depth was divided into sections and the value of the depth section to which the greatest number of pixels that correspond to the visual attention map belongs and the representative value were set. That is, the representative depth value was determined on the basis of the depth area to which the greatest number of pixels that corresponded to the visual attention map belonged. The depth representative value is used to move the depth of all the inputted 3D images (Fig. 2). In other words, the difference in the depth of each pixel that corresponds to the interest area or the non-interest area from the representative depth value is used to change the depth value to that of the corresponding pixel. The final depth value is applied to each pixel, the pixels that correspond to the interest area are positioned on the display, and the other pixels are moved on the basis of the representative depth value in the interest area.

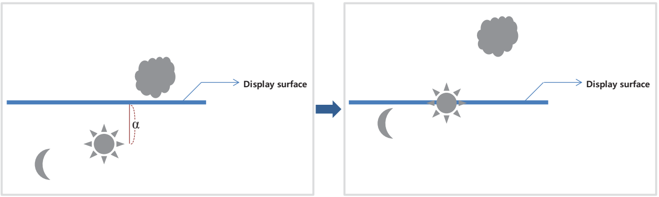


Fig. 2: Before (left side) and after (right side) the depth adjustment in the 3D Image Creation.

The 3D Image Creation plays the role of rendering again the 3D images with depth values that differ from those of the input images through the changed depth values after the depth readjustment.

III. Psychophysical Experiment

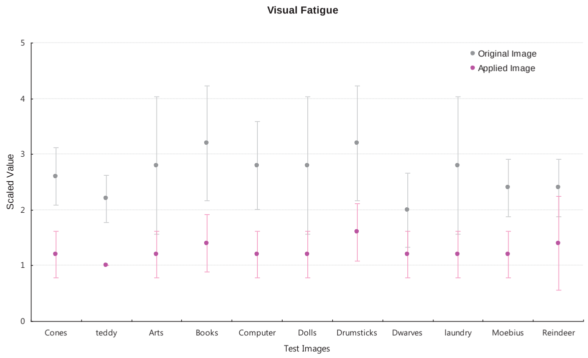


Fig. 3: Graph of the visual fatigue experiment result.

To evaluate the results of this study, 11 Middlebury stereo data set images were used. The static 3D images were toggled to compare the original images and the algorithm-applied images, after which relative evaluations were conducted.

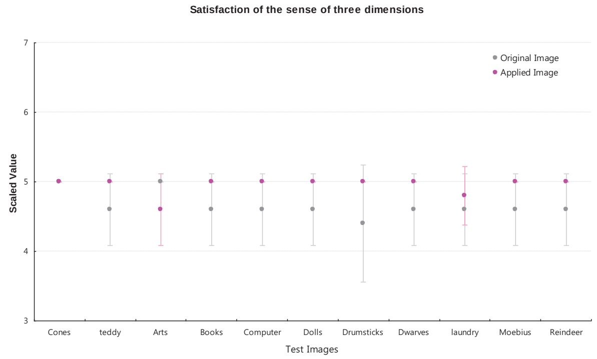


Fig. 4: Graph of the experiment results for the sense of three dimensions.

The results of the experiment are as follows. The visual fatigue of the images to which the algorithm for the reduction of visual fatigue and the enhancement of the sense of three dimensions was applied decreased by about 20 %, and the satisfaction of the sense of three dimensions improved by about 4 % from that with the original images. The visual fatigue results of the experiment with adaptive adjustment of the sense of depth of the interest objects according to the visual attention were significant because the original images and the algorithm-applied images in most of the error bar graphs did not overlap. Furthermore, the satisfaction of the sense of three dimensions increased as the visual fatigue decreased.

IV. Conclusion

In this study, a 3D image processing method of improving the visual sense of depth in line with the human visual perception characteristics was developed. The method developed in this study differs from other proposed methods in that the depth associated with selective visual attention was compromised while the relative sense of depth was maintained by focusing on the fact that the visual sense of depth is relative, rather than absolute. Thus, the method developed in this study is more satisfactory in terms of human visual perception and more effectively reduces visual fatigue and enhances the sense of three dimensions.

References

[1] Martens T. G. and Ogle K. N, "Observations on accommodative convergence; especially its nonlinear relationships," *American Journal of Ophthalmology*, pp. 455–463, 1959.

[2] Polak N. A. and Jones R, "Dynamic interactions between accommodation and convergence," *IEEE Transactionson Biomedical Engineering*, pp. 1011–1014, 1990.

[3] C Schor, "The influence of interactions between accommodation and convergence on the lag of accommodation," *Ophthalmic and Physiological Optics*, pp. 134–150, 1999.

[4] Livingstone, M. S., Hubel, and D. H, "Psychophysical evidence for separate channels for the perception of form, color, movement, and depth," *Journal of Neuroscience*, pp. 3416–3468, 1987.

[5] JiYoung Hong Youngshin Kwak and Du-Sik Park, "Preferred memory and accent colors shown on the display," *Journal of the society for information display*, pp. 649–656, 2007.

3D Hand Gesture Recognition from One Example

Myoung-Kyu Sohn, Sang-Heon Lee, Dong-Ju Kim , Byungmin Kim, Hyunduk Kim
Division of IT Convergence, DGIST, Korea

***Abstract*— In a typical recognition system, the inclusion of more training data is likely to increase the recognition rate. However, it is not easy to obtain large training sets. Focusing on practical applicability such as controlling home appliances, we propose a hand gesture recognition method from one example that is computationally efficient and can be easily implemented. 3D hand motion trajectory is achieved from a depth camera and then normalized for translation invariant feature extraction. Based on the simple K-NN classifier, we develop a pattern matching method by combining the DTW (Dynamic Time Warping) algorithm and a statistical measure for similarity between two random vectors. We conducted computational experiments on hand gesture data and compared the results with those derived via conventional DTW recognition.**

I. INTRODUCTION

Vision-based hand gesture recognition is one of the possible solutions for HCI (Human Computer Interface) as it provides natural interaction between people and all kind of devices. There have accordingly been many studies on gesture recognition techniques [1]. It is common knowledge in the field of statistics that estimating a given number of parameters requires a many-fold larger number of training examples. However, it is often difficult and expensive to acquire large sets of training examples. This leaves open the question of whether these methods are appropriate when a large amount of training data is not available [2]. Considering efficiency in computation and implementation in the case of a single training example, we apply the K-NN classifier with the DTW [3] technique, which is easy to implement and adaptable to changes in gesture types and users [4].

II. FEATURE EXTRACTION AND NORMALIZATION

A. *Feature Extraction*

The hand gesture video was obtained from a 3D depth camera. Using OpenNI [5] middleware, we obtained 3-dimensional coordinate values of the hand position from each frame as a feature vector, $\mathrm{Q} = (\mathrm{x}, \mathrm{y}, \mathrm{z})$. In general, x and y are the pixel positions of the hand in the XY plane, and z is the real distance value between the camera and the hand. Because the elements of the feature vector have different characteristics, it is difficult to consider the 3D movement equally in every axis. We converted the feature vector in projective coordinates to a value in real-world coordinates for this reason.

B. *Normalization*

Using the position as a feature is appealing, because it is simple to extract and is highly informative about the gesture content. However, it is not invariant to translate.

We can achieve translation invariance by the following normalization technique. Normalization is achieved by two steps. First, we subtract the position of the hand in the first frame from the entire trajectory. Second, we calculate the minimum enclosing sphere of the data set and find the bounding cube. We then resize the cube to a unit cube. By this normalization, the features become translation and scale invariant.

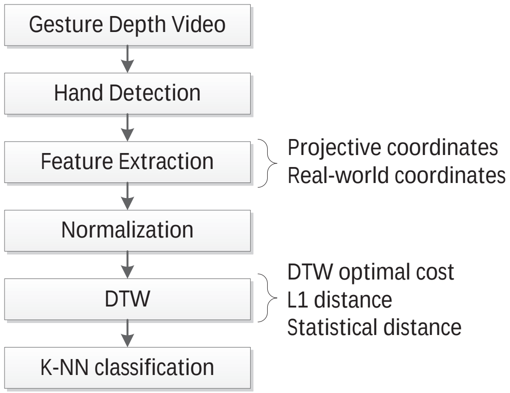


Fig. 1. Overall framework for gesture recognition

III. RECOGNITION

A. *Dynamic Time Warping*

DTW is usually used for pattern matching method between two time series. Unlike general statistical methods, a training phase is not necessary. The proposed recognition method is composed of two steps. First, we use DTW for sequence alignment of two samples with different lengths. DTW provides the matching cost of the optimal warping path but it is not appropriate for gesture data, which have many variations between users or hand location. Instead of using the DTW optimal cost, we propose a statistical measure for calculating the distance between two aligned sequences, which can be written as,

$$d(X,Y) = 1 - corr(X,Y) = 1 - \frac{cov(X,Y)}{std(X) \cdot std(Y)} \qquad (1)$$

where $std(X)$ and $cov(X,Y)$ denote the standard deviation of the feature vector and the covariance of two feature vectors, respectively. K-NN with the K=1 classifier is then applied for classification.

This work was supported by the DGIST R&D Program of the Ministry of Education, Science and Technology of Korea (12-IT-03)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

IV. Experiments

To evaluate the proposed gesture recognition method we have collected two types of gesture video clips. The first set is seven direction gestures, which are left, right, up, down, clock-wise rotation, counter clock-wise rotation, and push. The second set is ten digit gestures in the style of Palm's Grafitti Alphabet. Each gesture is performed ten times by eight different people. The video clips were captured with a Kinect camera with 320 x 240 resolution. The total body of data consists of 560 samples for the direction data and 800 samples for the digit gesture data. We choose only one sample from each class randomly for a reference sample. The remaining samples are used as the test data. We repeat each experiment 30 times.

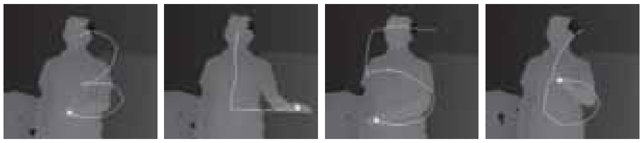


Fig. 2. Example of 3D hand gestures (digit gestures in the style of Palm's Grafitti Alphabet)

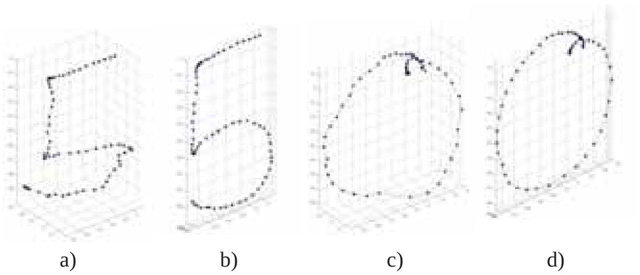


Fig. 3. An example of normalized features for two hand gesture.
a) digit 5 in projective coordinates b) digit 5 in real world coordinates
c) clock-wise rotation in projective coordinates d) clock-wise rotation in real-world coordinates

A. DTW with a different measure

The recognition rate depicted in Table 1 shows the overall performance results depending on the distance measure and the coordinates. DTW cost is the cost through the DTW optimal warping path using Euclidean distance. L1 and the proposed distance are the Manhattan distance and statistical distance between two DTW aligned sequences, respectively. The results show that the statistical distance measure is more reliable than both the DTW cost and Manhattan distance, as shown in Fig. 4, Fig. 5, and Table 1.

Table 1. Recognition rate of the digit gesture. The results are mean values over 30 trials

distance coordinate	DTW cost	L1	Proposed
Projective	67.88	66.12	69.08
Real-world	91.53	84.32	94.15

B. Feature and Normalization

Two kinds of position features are extracted and normalized. The first has positions in projective coordinates and the second has positions in real-world coordinates. The recognition rates by each feature are compared in Table 1 and Fig. 6. The performance was better when using the feature in real-world coordinates. This implies that it is important that each element in a feature vector has the same properties.

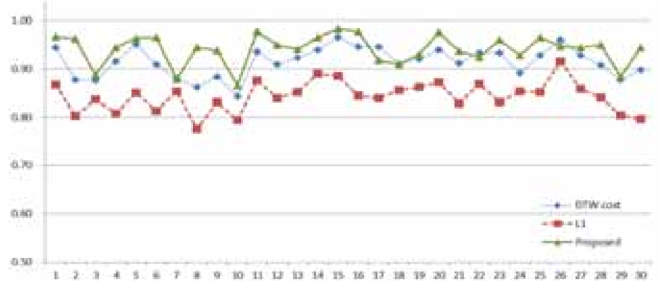


Fig. 4. Recognition rate for digit gesture over 30 trials

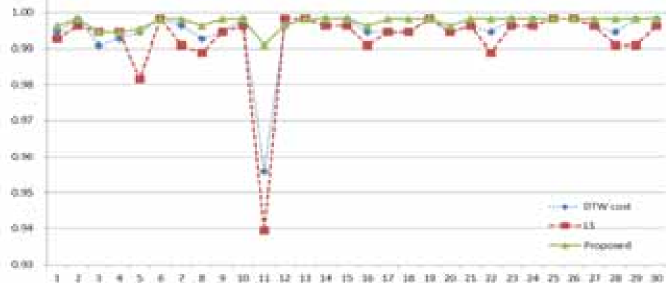
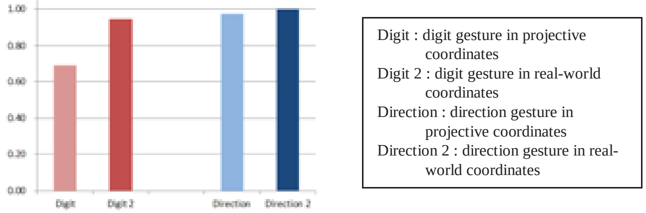


Fig. 5. Recognition rate for direction gesture over 30 trials



Digit : digit gesture in projective coordinates
Digit 2 : digit gesture in real-world coordinates
Direction : direction gesture in projective coordinates
Direction 2 : direction gesture in real-world coordinates

Fig. 6. Recognition rates of digit and direction gesture

V. Conclusion

This paper proposes a 3D hand gesture recognition method from one example. By combining the statistical distance with DTW sequence alignment, we achieved better performance. In addition, normalization using the feature that is converted to real-world coordinates provides a significantly increased recognition rate.

References

[1] S. Mitra, T. Acharya, "Gesture Recognition: A Survey," IEEE Systems, Man, and Cybernetics—-Part C: Applications and Reviews, Vol. 37, No. 3, pages 311 - 324, 2007.

[2] L. Fei-Fei, R. Fergus, and P. Perona. "One-shot learning of object categories," IEEE Trans. Pattern Analysis and Machine Intelligence, 28(4):594–611, 2006.

[3] C. S. Myers and L. R. Rabiner, "A comparative study of several dynamic time-warping algorithms for connected word recognition," The Bell System Technical Journal, vol. 60, pp. 1389-1409, 1981.

[4] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "UWave: Accelerometer-based personalized gesture recognition and its applications," Pervasive Mobile Comput., vol. 5, no. 6, pp. 657–675, 2009.

[5] http://www.openni.org/

Smart TV Interaction System Using Face and Hand Gesture Recognition

Sang-Heon Lee, Myoung-Kyu Sohn, Dong-Ju Kim, Byungmin Kim, and Hyunduk Kim

Dept. IT convergence, DGIST, Daegu, Republic of Korea

***Abstract*-- In this paper, a face and hand gesture recognition system which can be applied to a smart TV interaction system is proposed. Human face and natural hand gesture are the key component to interact with smart TV system. The face recognition system is used in viewer authentication and the hand gesture recognition in control of smart TV, for example, volume up/down, channel changing. Personalized service such as favorite channels recommendation or parental guidance can be provided using face recognition. We show that the face recognition detection rate is about 99% and the face recognition rate is about 97% by using DGIST database. Also, hand detection rate is about 98% at distance of 1 meter, 1.5 meter, and 2 meter, respectively. Overall 5 type hand gesture recognition rate is about 80% using support vector machine (SVM).**

I. INTRODUCTION

In these days, a smart TV which is also sometimes refers to as "Connected TV" or "Upgrade of IPTV" is widely spread. As the smart TV is connected to internet, that has so many functions. Smart TV can do video on demand streaming service, internet surfing, game application service and so on. Therefore, a remote controller with which human can interact to smart TV has been more and more complicated. In addition, the remote controller can be lost or just lying in the corner. The situation is very much same in the case of keyboard and mouse and they are more cumbersome than remote controller. So, vision-based natural interaction system, face and gesture recognition is widely studied and the solutions of interaction system begin to apply to smart TV.

This paper aims to develop a robust vision-based face and hand gesture recognition system for the control of smart TV. We have implemented a face and hand gesture recognition module for channel/volume changing services and personalized services such as favorite channel, parenting guidance, etc.[1][2].

For hand detection, a data fusion technique is used. To detect the hand posture the Adaboost algorithm is used and the repeated detection is used for tracking algorithm. However, the tracking based on Adaboost is limited by the static view and required hand size; thus, skin color information and a KLT tracker are applied to achieve robust tracking of hand gestures. After hand detection, five types of hand gestures are recognized using support vector machine (SVM) [3]. The five types of gestures are "left", "right", "up", "down", and "push".

For illumination-robust face detection, uniform local binary patterns (ULBPs) are used as feature vectors and SVM is used as classifier as in hand gesture recognition. Face enrollment is executed in user-cooperative environment, that is, user's face image is captured just in front of camera, user staring straight ahead. Face recognition processes are Gabor filtering, ULBP transformation, and histogram matching with chi-square distance measure.

The face and hand gesture integrated system was implemented with message passing multi-thread program. The integrated interaction system has two modules. One is RTU (Recognition To UI) and the other is IUI (Interactive User Interface).

Fig.1 shows a scenario of smart TV interaction system using face recognition and hand gesture recognition.

In this paper, we implemented interaction system of smart TV using face and gesture recognition and propose well-behaved robust detection and recognition algorithms.

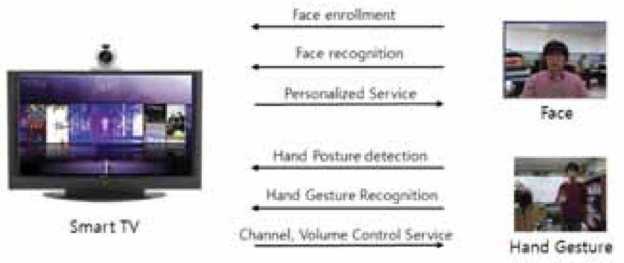


Fig. 1. Scenario of Interactive TV using a face recognition system.

II. IMPLEMENTATION OF SMART TV USING FACE AND HAND GESTURE RECOGNITION

A. Face detection and recognition

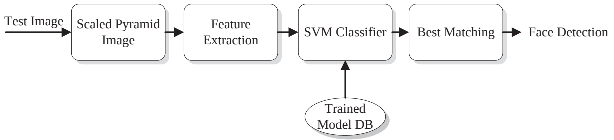


Fig. 2. Face detection process.

Fig. 2 shows a method of face detection. We used ULBP as feature vectors and SVM as classifier for face detection. Many studies have been completed using the ULBP features for face detection and face recognition applications despite the fact

This work was supported by the DGIST R&D Program of the Ministry of Education, Science and Technology of Korea (12-IT-03).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

that the LBP feature was originally designed for texture descriptions. The most important properties of the ULBP feature are its computational simplicity and its compensation for the monotonic transformation of the gray scale. ULBPs are a subset of all theoretically possible patterns, encompassing a total of 59 patterns for all LBPs [6]. With the SVM classifier, the equal error rate (EER) is determined while changing the threshold values. The training and test of face detection is done with DGIST database, about 11,000 face images, and the final accuracy is about 99.93% with 0.07% false positive rate.

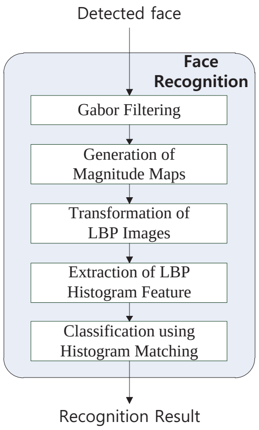

Detected face
Face Recognition
Gabor Filtering
Generation of Magnitude Maps
Transformation of LBP Images
Extraction of LBP Histogram Feature
Classification using Histogram Matching
Recognition Result

Fig. 3. Face recognition process.

Fig. 3 shows a face recognition process after face detected. For face recognition, Local Gabor Binary Pattern Histogram Sequences (LGBPHS) are used as feature vectors [4]. Chi-square measurements are then used to determine the histogram distance. Multi-resolution and multi-orientation Gabor filters are exploited to decompose the input face images for LGBPHS feature extraction. Here, four-resolution and eight-orientation Gabor magnitude images are converted to LBP images. The face recognition experiment was performed using a total of 310 images. The 310 images are composed of 31 persons and 10 images per person. Using LGBPHS, the final recognition result showed 96.77% accuracy.

B. Hand detection and gesture recognition

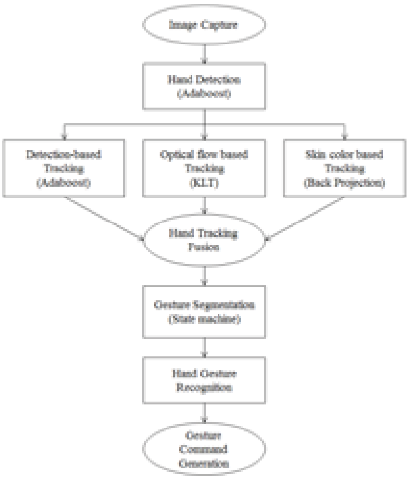


Image Capture
Hand Detection (Adaboost)
Detection-based Tracking (Adaboost)
Optical flow based Tracking (KLT)
Skin color based Tracking (Back Projection)
Hand Tracking Fusion
Gesture Segmentation (State machine)
Hand Gesture Recognition
Gesture Command Generation

Fig. 4. Hand detection and gesture recognition process.

Fast hand detection is executed by the ROI (region of interest) based cascade Adaboost method. An alternative to detecting the hand is to use the skin color feature of the hand. We proposed a hand detection method through back projection method shown in Fig.4. We used the KLT(Kanade-Lucas-Tomasi) feature tracker to track a hand, which has been previously detected by the methods in Adaboost and skin color detection. Total hand detection rate is about 98% at distance of 1 meter, 1.5 meter, and 2 meter, respectively using 1,896 test images. Overall 5 type hand gesture recognition rate is about 80% using support vector machine (SVM).

C. Smart TV interaction system

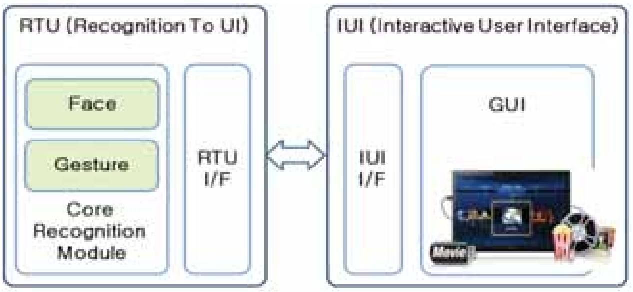


Fig. 5. Overall architecture of the interactive smart TV system.

The overall architecture of the interaction system is illustrated in Fig. 5. The system has two parts: the RTU and the IUI. The RTU (Recognition to UI) is the core recognition subsystem, which contains the face recognition module and gesture recognition module. The IUI (Interactive User Interface) is a graphic user interface subsystem controlled by the RTU.

III. CONCLUSION

In this paper, a face and hand gesture recognition system which can be applied to a smart TV interaction system is proposed. The experimental results show that the robust face recognition and hand gesture recognition method offers excellent meaningful results. Thus, the proposed interaction system is suitable for smart TV.

REFERENCES

[1] S. Mitra, T. Acharya, "Gesture Recognition: A survey," IEEE Trans. Systems, Man and Cybernetics-Part C, vol.37, no.3, pp.311-324, May 2007.

[2] S. Z. Li and A. K. Jain, Handbook of Face Recognition, Springer, 2005.

[3] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.

[4] W. Zhang, S. Shan, W. Gao, X. Chen. and H. Zhang, “Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in Proc. IEEE Int. Conf. Computer Vision, pp.786-791, Oct. 2005.

[5] Paul Viloa, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", I-511 - I-518 vol.1, CVPR 2001.

[6] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, July 2002, pp. 971-987.

Magnetic Resonance Wireless Power Transfer System for Practical Mid-Range Distance Powering Scenario References

Ki Young Kim, Changwook Yoon, Nam Yoon Kim, Jinsung Choi, Young-Ho Ryu, Dong-Zo Kim,
Keum-Su Song, Chi-Hyung Ahn, Eunseok Park, Yun-Kwon Park, and Sangwook Kwon
Future IT Research Center, Samsung Advanced Institute of Technology, Yongin 446-712, Korea

***Abstract*—Practical magnetic resonance wireless power transfer (WPT) system is presented. High-efficient power amplifier (PA) and adaptive tracking modules are combined with optimized wireless power links to propose mid-range wireless powering application scenarios in home-environments.**

I. INTRODUCTION

Recently, wireless power transfer (WPT) technology via magnetic resonance coupling [1] has been received much attention due to its versatile industrial applicability from low-power biomedical implants to high-power electric vehicle recharging applications. In order to extend the transfer distance and/or to increase the power transfer efficiency, repeater resonator can be introduced; thereby mid-range WPT applications become closer to practice. In this work, practical working WPT system with target device condition oriented tracking module and optimized wireless power link is presented such that mid-range distance wireless powering scenarios for home-environments are proposed.

II. WPT SYSTEM DESCERIPTIONS

A. Power Amplifier and Adaptive Tracking Module

High-efficient two-stage class-E power amplifier (PA) has been developed, where the electricity from the plug is transformed to 6.78MHz for magnetic resonance wireless power link. Fig. 1 shows its circuit topology and measured efficiency versus output power. High efficiency of more than 90% can be maintained in a relatively broad range higher than 150W output power. Fig. 2 shows block diagram of adaptive transmitter (TX) system and its fabricated module assembly. The out-of-band (OOB) wireless signaling based on IEEE 802.15.4 (Zigbee) has been adopted for bidirectional data transmission such that the power-level and frequency tracking operations upon requests from receiver (RX) can perform [2].

B. Magnetic Resonance Wireless Power Links

Magnetic resonance wireless power links that include at least a repeater have been optimized to enhance the power transfer efficiency (PTE) at given fixed orientations between the TX and RX resonators, which are not necessarily aligned coaxially, as shown in Fig. 3. Fig. 3 is the reference configuration for the TX and Rx resonators of 1.5m apart with coaxially aligned, which showed 87.7% in PTE, while that of the non-coaxial configuration shown in Fig. 3(b) was 83.9% but it further reduced to 74.9% when the TV was loaded as shown in Fig. 3(c). Additional repeater shown in Fig. 3(d) has been added to compensate the PTE up to 83.4%.

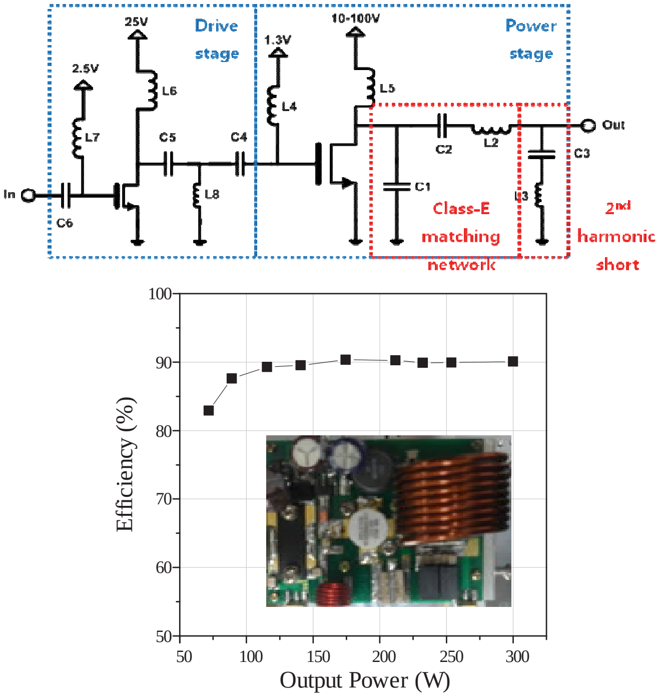


Fig. 1. High-efficient two-stage class-E PA; Circuit topology (top) and measured efficiency versus output power (bottom). [Inset: Photograph of the fabricated PA].

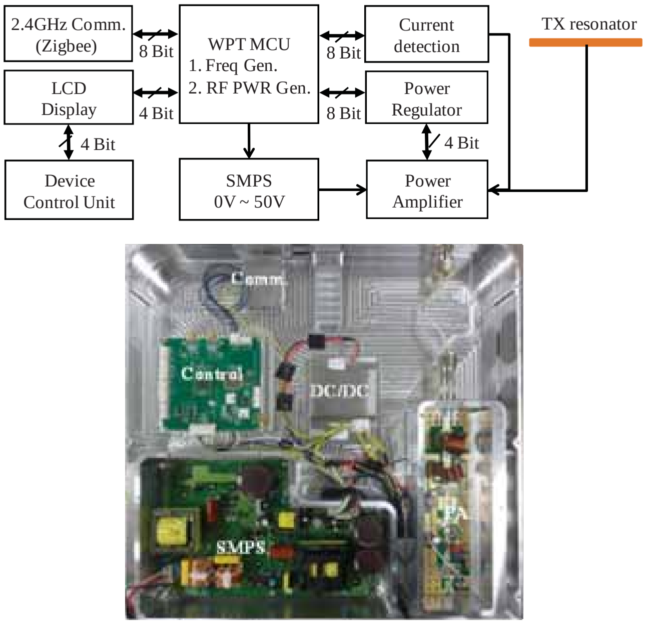


Fig. 2. Adaptive TX system; Block diagram (top) and fabricated module assembly (bottom).

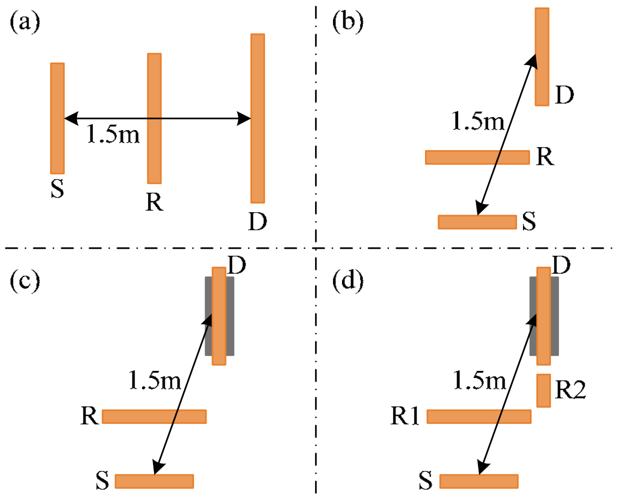


Fig. 3. Configurations of wireless power links with repeater resonators (side view).

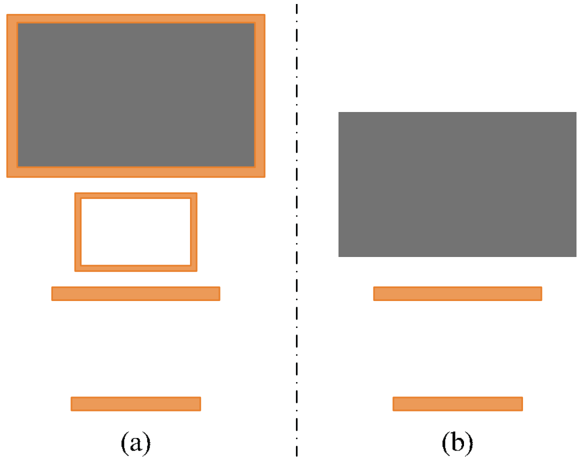


Fig. 4. Application of WPT system for installation of wall-hanging (a) and standing (b) 46-inch LED TV (front view).

III. MID-RANGE WIRELESS POWERING SCENARIOS

Some useful scenarios of mid-range distance wireless powering applications can be practically realized with our referential adaptive WPT system. Both wall-hanging and standing type installations of normal LED TV with the adaptive WPT system have been successfully demonstrated as shown in Fig. 4. Prior to transfer real “turn-on” power, a pilot low-level powering is initially performed to seek optimal conditions of output power-level considering the PTE and the working frequency via the OOB signaling. Once the optimal conditions for the power transfer have been found, the real power around 150W is transmitted via the wireless power link. Here, the wall-hanging and standing type configurations are composed of 4 and 3 resonators, respectively. Fig. 5 shows another use-case proposal of WPT scenario considering home-environment, in which the real-time mid-power delivery for wall-hanging LED TV (150W) is simultaneously performed with low-power battery charging for two smart phones (3W × 2 units). Main specifications and functions of our adaptive WPT system with wireless power links shown in Figs. 4 and 5 are summarized in Table 1.

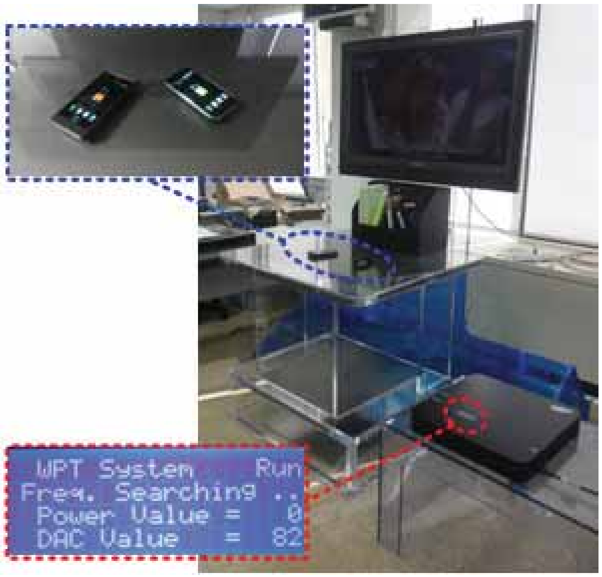


Fig. 5. Application of WPT system to simultaneous mid-power delivery for 46-inch wall-hanging LED TV and low-power battery charging for two smart phones). [Insets: Two smart phones are in charging (top); and the user interface of adaptive tracking system (bottom)]

TABLE I
MAIN SPECIFICATIONS AND FUNCTIONS OF THE ADAPTIVE WPT SYSTEM.

Efficiency	Transmitter	>90%
	Receiver	92.8% @ 150W
	Power link	83.4% @ 1.5m (90°)
Power Handling	Transmitter	>300W
	Pick-up	>150W
Functions	Wall-hanging / Standing TV	
	Mid-power delivery / Low-power charging	
	Frequency and power-level tracking	

IV. CONCLUSION

Practically working adaptive WPT system combined with non-coaxially aligned resonator configurations for mid-range distance wireless powering scenarios has been presented. High-efficient two-stage class-E PA and the powering-level with frequency tracking modules are designed and fabricated in an assembly. The variations of power transfer conditions including transfer distance, requested power-level from Rx devices, and number of Rx devices could be informed from the Rx system to the Tx system to seek optimum WPT conditions via the Zigbee based bidirectional wireless communication. The wireless power links with non-coaxially aligned 3 or 4 resonators have been optimized to obtain high PTEs. Potential use-case scenarios based on the adaptive WPT system are proposed, which are believed to be referential mid-range distance WPT applications for home-environments.

REFERENCES

[1] A. Kurs, A. Karalis, R. Moffatt, J. D. Joannopoulos, P. Fisher, and M. Soljacic, “Wireless power transfer via strongly coupled magnetic resonances,” *Science*, vol. 317, no. 5834, pp. 83-86, July 2007.

[2] N. Y. Kim, K. Y. Kim, J. Choi, and C. –W. Kim, “Adaptive frequency with power-level tracking system for efficient magnetic resonance wireless power transfer,” *Electronics Letters*, vol. 48, no. 8, pp. 452-454, Apr. 2010.

Tile Binning Algorithm for Vector Graphics Minimizing False Overlap

Jeong-Joon Yoo, Seungwon Lee, Seokyoon Jung, and Shihwa Lee, *Member, IEEE*

***Abstract*— In this paper, we present an efficient tile binning algorithm that reduces *false overlap* area for the Bezier curve. To do so, we propose a *Multiple-Bounding Box* (MBB) enclosing a Bezier curve with multiple bounding boxes. Experimental comparisons of MBB and legacy methods show that the proposed MBB method reduces 16~35% memory overhead. It also reduces 23~54% CPU latency overhead in a tile-based 2D graphics pipeline.**

I. Introduction

Tile-based Rendering (TBR) is broadly used in GPUs. Because the TBR uses a fast and small on-chip memory rather than a slow external memory, it is used in high performance and low power GPU.

In the TBR, as shown in Figure 1, the screen is divided into a number of non-overlapping regions called *Tiles*, which are processed sequentially. In the *Tile Binner*, an overlap-test between all geometries and the tiles is made. To simplify this test, we calculate a bounding box that encloses the geometry. If the bounding box overlaps with a tile, the corresponding geometry ID is dumped into a data structure called the *Tile Bin* (one *Bin* per one tile). As a result, the *Tiled Data* is generated, as shown in Figure 1. The Renderer can process the tile-based rendering with the tiled data. The performance of the rendering step depends on the number of tiled data. A *false overlap* [1], [2] irrelevantly increases the number of tiled data. Therefore, to improve the performance of the rendering, the false overlap should be removed.

In general, the tile binning algorithm consists of three steps: i) calculate a bounding box for a geometry, ii) check which tiles overlap with the bounding box for the geometry, iii) dump the overlapping geometry's ID into the corresponding *Bin*.

The efficient bounding box for a triangle was discussed [3], [4]. While the triangle is used as a primitive geometry in 3D graphics, the Bezier curve is used as one of primitive geometries in vector graphics. To implement a tile-based vector graphics (or 2D graphics) pipeline, we need a tile binning algorithm for the Bezier curve.

In this paper we propose an efficient tile binning algorithm that provides an efficient bounding box for the Bezier curve. The remainder of this paper is organized as follows. In section 2, we will explain the legacy bounding box methods and propose an efficient bounding box method called the *Multiple-Bounding Box*. In section 3, the performance will be compared with the legacy methods. Finally, we will conclude the results in section 4.

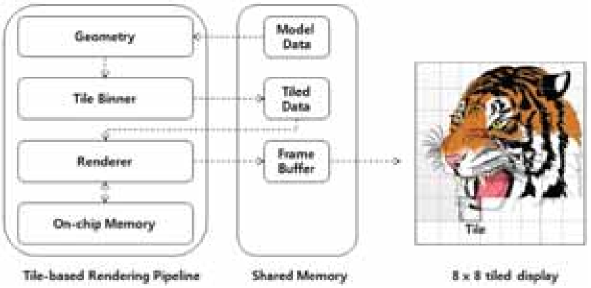
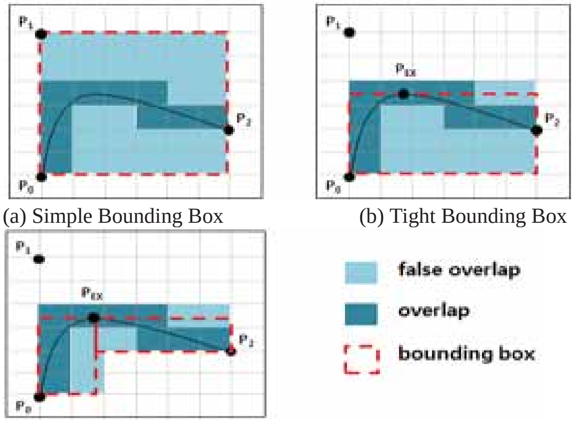


Fig. 1. Tile-based Rendering Pipeline



(a) Simple Bounding Box (b) Tight Bounding Box

(c) Multiple-Bounding Box

Fig. 2. Methods for Calculating Bounding Box

II. Multiple-Bounding Box

Figure 2 shows some different methods for the bounding box for a Bezier curve. In Figure 2 (a), the bounding box for a Bezier curve is defined as a rectangle (dotted red line), enclosing all control points of the curve [5]. The Simple Bounding Box (SBB) method is simple, but many *false overlap* areas that do not actually contain the curve are produced.

To reduce the false overlap, the Tight Bounding Box (TBB) method can be used, as shown in Figure 2 (b). To calculate a tight bounding box (dotted red line in Figure 2 (b)), we must first calculate the *y* coordinate (at P_{EX}) for the point of tangency for a curve. This can be calculated as follows:

$$y = (y_1-y_0)(y_0-y_1)/(y_0-y_1-y_1+y_2) + y_0 \qquad (1)$$

(When control points $p_0=(x_0, y_0)$, $p_1=(x_1, y_1)$, and $p_2=(x_2, y_2)$ for a curve are defined.)

To reduce the number of false overlaps, we propose the Multiple-Bounding Box (MBB) method. In this method, we also calculate the coordinate (that is P_{EX}) for the point of tangency for a curve. We can generate two sub-boxes (so, it is

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

called *Multiple Bounding Box*), one for P_0 and $P_{EX,}$ and the other for P_{EX} and P_2. We can check all of the tiles and whether they overlap with one of two sub-boxes or not. If any one of the sub-boxes overlaps with a tile, the corresponding geometry's ID is stored into the Bin of the current tile. In this method, the false overlaps can be more effectively reduced than SBB and TBB. In the following section, we will show the experimental comparisons among them.

III. Experimental Comparisons and Analysis

We compare the memory overhead, performance overhead, and the total performance of the tiled 2D graphics pipeline when we use the one method of SBB, TBB, and our MBB. To do so, we implemented a tile-based 2D graphics S/W pipeline (tile size is 32x32 pixels, screen size is 800x480 pixels with 25x15 tiles) from the open source code of Skia [6]. We implemented a tile-based 2D graphics pipeline consisting of five steps: Path Generation (PG), Tile Binning (TB), Rasterization (RA), Shading (SH), and Transfer (TR). We used the Skia benchmark (path_fill_small_long_curved test case) as a benchmark. The performance comparisons were made on a target board with 1GHz RISC CPU.

A. *Memory and CPU Overheads*

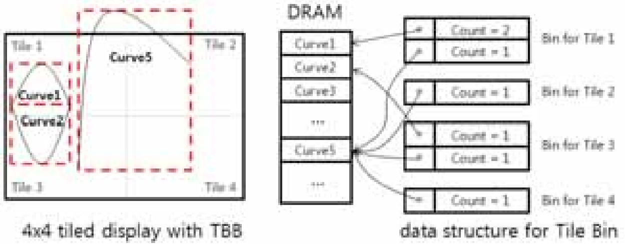
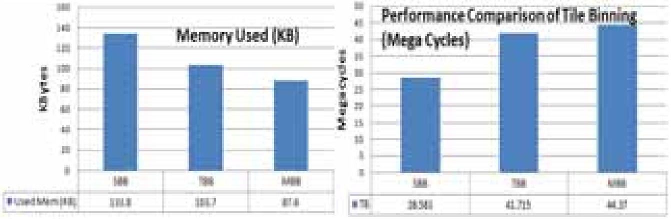


Fig. 3. Data Structure for Tile Bin

Figure 3 shows a 4x4 tiled display (for a simple understanding) and its memory structure of the Tile Bin for the TBB method. In tile 1, two consecutive curves (Count = 2), from Curve1 to Curve2, are rendered. One more curve (Count = 1), Curve5, is also rendered in tile 1. Tile 4 is a false overlap, but memory is unnecessarily occupied in TBB. However, in MBB, the false overlap can be effectively removed, as described in section 2. As a result, it is memory efficient.



(a) Memory overhead (b) Latency overhead

Fig. 4. Comparison of Bounding Box for Memory and Latency Overheads

As shown in Figure 4 (a), the SBB method requires 1.29 ~ 1.53 times the memory size of the TBB, as well as our MBB methods. Because the false overlaps are reduced in our MBB, 16~35% memory for tile binning is effectively reduced.

Figure 4 (b) compares the performance of the tile binning algorithms. In this result, the SBB method outperforms the others approximately 1.46~1.55 times, due to its simple calculation of the bounding box. That is, it does not need to calculate the point of tangency (P_{EX} in Fig 2 (b)). However, the TBB and MBB methods must calculate the coordinates for the point of tangency.

B. *Performance of Tile-based 2D Graphics Pipeline*

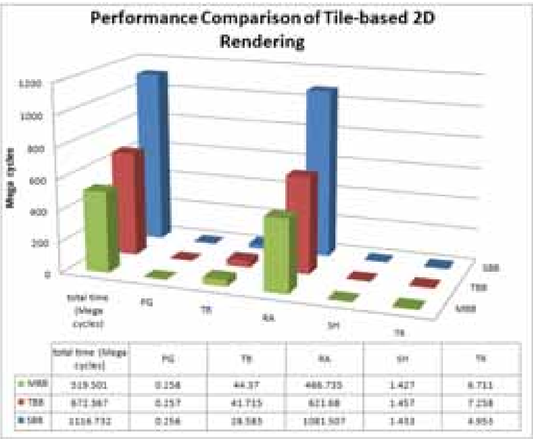


Fig. 5. Performance of Tile-based 2D Graphics Pipeline

We can compare the total performance of the 2D graphics pipeline when we adopt different methods, as explained in section 2. Although the algorithm for using the SBB method in Tile Binning has a low latency, as shown in Figure 4 (b), the Rasterization (RA) step takes more time (Figure 5) than others because SBB has many false overlaps to render. On the other hand, the latency of the RA step for MBB is effectively reduced due to the number of false overlaps removed during Tile Binning (TB) step in our proposed MBB method.

IV. Conclusions

Throughout this paper, we have proposed a tile binning algorithm for vector graphics. Our contribution is to effectively reduce false overlaps in tile-based vector graphics rendering. As a result, 16~35% of memory size is saved. 23~54% CPU latency of tile-based 2D graphics pipeline is reduced when compared with the legacy methods. Because our MBB method is efficient not only in memory, but also in performance, it can be used in both desktop GPU and mobile GPU.

References

[1] Hsiu-ching Hsieh, et al., "Methods for Precise False-Overlap Detection in Tile-based Rendering," *Int. Conf. on Computational Science and Engineering*, pp. 414-419, 2009.

[2] I. Antochi, et al., "Efficient Tile-Aware Bounding-Box Overlap Test for Tile-Based Rendering," *Int. Symp. On System-on-Chip*, pp. 165-168, 2004.

[3] K.J Min, et al., "Method and System for Tile Binning Using Half-Plane Edge Function," *US Patent US2008/0018664*, 24, Jan. 2008.

[4] Stephen Junkins, et al., "Polygon Binning Process for Tile-based Rendering," *US Patent US6975318*, 13, Dec. 2005.

[5] Hon-Wen Pon, et al., "Method and apparatus for rendering cubic curves," *US Patent6295072*, 25, Sep, 2001.

[6] http://code.google.com/p/skia, *Skia – 2D Graphics Library*.

Design and Implementation of a PIR Luminaire with Zero Standby Power Using a Photovoltaic Array

Cheng-Hung Tsai, Ying-Wen Bai, Ming-Bo Lin, Chih-Yu Chung, and Roger Jia Rong Jhang

***Abstract*-- This project further reduce the standby power consumption of a PIR luminaire. Generally, although a PIR luminaire will turn on when motion is detected and turn off when any motion is no longer present, it still consumes 1 to 3 W of power when the lamp is off. In this design the luminaire consumes 1 mW when the light is turned off. The power consumption is much lower than that in present products, and the zero standby power luminaire is not only easy to set up but also inexpensive. A more effective circuit design is used to reduce the power consumption which does not affect original functions. Furthermore, the photovoltaic array is included in this design to reduce the consumption from the local electric power company. The standby power consumption of the luminaire is 1 mW in a darkroom and less than 1 mW in a non-darkroom. When the illumination intensity is higher than 150 lx, the consumption from the local electric power company is 0 W.**

I. INTRODUCTION

The average household power consumption dedicates 11% of its energy budget to lighting [1]. The power consumption of lamps in a typical home is a factor which cannot be ignored. To save lighting energy, the pyroelectric infrared (PIR) luminaire is now in widespread use [2]. Because the PIR luminaire only comes on when the PIR sensor is activated, no energy is wasted. But what is unfortunate is that the power consumption of the PIR luminaire in the standby state cannot be switched off completely without being unplugged. There are three states of the PIR luminaire defined in this paper: the cut-off state, the standby state and the lighting state. The cut-off state means that the luminaire is unplugged from its power source and does not consume any electricity. In the standby state the luminaire is connected to a power source; but if no user is approaching, it does not turn on the light. In the lighting state the PIR sensor is activated, and as a result of reduced daylight the luminaire then produces light. Though the luminaire in the standby state is not performing its main function of lighting, it is often performing some internal functions like sensing both IR and daylight. In such a situation the luminaire cannot be switched off unless the unit is unplugged. These internal functions require power to operate, and the power consumption used by the luminaire while in the standby state is called “standby power” in this paper. Its origin lies in that these internal functions require not only a specific low DC voltage to operate but also continuous power supplied by an AC/DC converter which has no power-off switch [3]. The converter, which serves as a power supply in the luminaire, converts AC to low DC voltage for the operation of the internal functions. It is inefficient at low DC voltage and consumes between 1 and 3 W, which is many times more than the power actually used for the internal functions. The standby power of a PIR luminaire draws power 24 hours a day. This amount is typically small, but the sum of the standby power consumption of all PIR luminaires within a household is significant. Therefore, in the long run the PIR luminaire consumes much standby power while in the standby state. In fact the reduction of the standby power of the PIR luminaire is still an important issue. In this paper a new PIR luminaire design is proposed whose standby power consumption is reduced to 1 mW. When the illumination intensity suffices, i. e. is higher than 150 lx [4], the PIR luminaire’s standby power consumption is 0 W. This design is called a “zero standby power PIR luminaire”. It is easy to set up, inexpensive, saves power efficiently and is consequently suitable for use in most locations.

II. CIRCUIT DESIGN OF THE ZERO STANDBY POWER PIR LUMINAIRE

As described in the introduction, the PIR luminaire is turned on only when motion is detected and by reduced daylight. When there is no motion, both the time when the light is switched off and the duration of the lighting can be adjusted. The state diagram of the PIR luminaire is shown in Fig. 1.

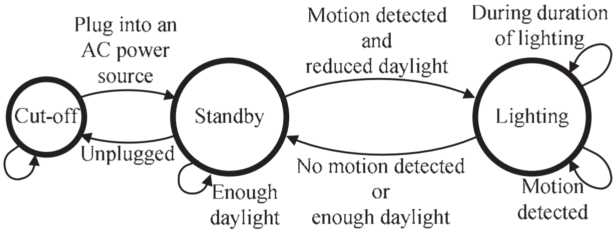

Plug into an AC power source
Motion detected and reduced daylight
During duration of lighting
Cut-off
Standby
Lighting
Unplugged
No motion detected or enough daylight
Enough daylight
Motion detected

Fig. 1. State diagram of the PIR luminaire.

This standby power consumed by the PIR luminaire is mainly that needed by the AC/DC converter and is many times more than that actually used by the luminaire itself. To reduce this standby power the converter should be turned off. The main concept of this design is that the DC voltage provided by the converter should be stored in an ultracapacitor (UC) which is used to support internal functions. If the UC needs to charge, the converter is turned on; otherwise it is turned off so it won’t use any unnecessary standby power. All power can be turned off completely by means of a solid state relay (SSR). The block diagram of this design is shown in Fig. 2. The output voltage of the AC/DC converter is denoted as V_{DC}, the ultracapacitor voltage as V_{UC} and the output voltages of the boost circuit as V_{CC} and V_{PIR}. The V_{DC} is the UC charge source, and the V_{UC} is both the UC voltage and the input of the boost circuit. The V_{CC} is the required operation voltage that supports the DC voltage module operation, and the V_{PIR} is the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

PIR module operation voltage.

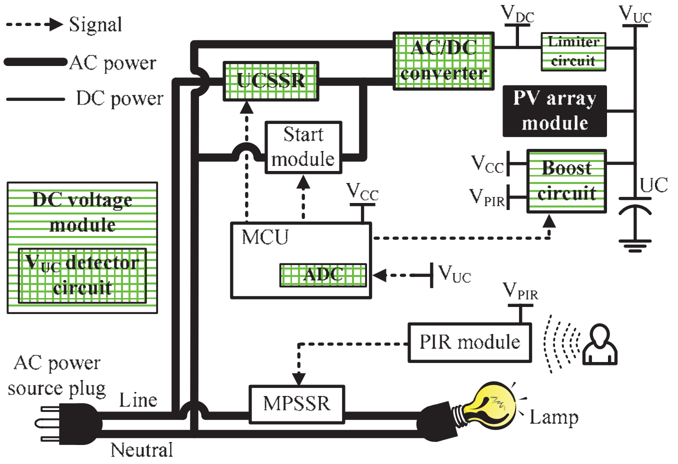


Fig. 2. Block diagram of the zero standby power PIR luminaire.

The MUC controls the V_{UC} detector circuit and the boost circuit to keep the V_{UC}, V_{CC} and V_{PIR} to the predefined voltage level. The UC thus functioning as a battery supports the boost circuit input. The boost circuit outputs two regular voltages (V_{CC} and V_{PIR}) which support the DC voltage module and the PIR module operation. The start module is designed to avoid the ultracapacitor charge SSR (UCSSR) normal open in the initiation that causes the luminaire not to work. It guarantees that the luminaire can continue to work satisfactorily when power is restored after a power failure.

III. MEASURING THE POWER CONSUMPTION OF THIS DESIGN

In this design the standby power of the AC/DC converter used in Fig. 1 is 1 W; the V_{UC} detector circuit is designed to reduce the power consumption of the converter. The power consumption of the discharge time is 0 W. The charge and discharge of the UC are a cycle whose time in standby state is more than 8 hrs.

$$T_{\text{cycle}} = T_{\text{charge}} + T_{\text{discharge}} . \tag{1}$$

We denote the average power in T_{cycle} as Pw_{ave} and

$$Pw_{\text{ave}} = \frac{\sum Pw_{\text{charge}} + \sum Pw_{\text{discharge}}}{T_{\text{charge}} + T_{\text{discharge}}}, \ \sum Pw_{\text{discharge}} = 0 \tag{2}$$

thus $$Pw_{\text{ave}} = \frac{\sum Pw_{\text{charge}}}{T_{\text{cycle}}} = 0.001 \text{ W}. \tag{3}$$

The percentage of improvement is 99.99 %. In general, the cycle life of the UC is more than 100,000 times. In addition, the life time of the UC can exceed 20 years.

The UC can be charged by a small amount of current. Therefore an amorphous silicon PV array circuit is used that converts solar energy into direct current electricity to charge the UC. The PV array is suitable for an indoor environment. The charge current produced by the PV array is less than 0.5 mA, but this amount is sufficient both to charge the UC and to reduce the power consumption of this design. The PV array module is shown in Fig. 3.

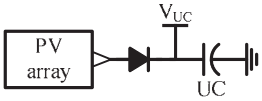


Fig. 3. PV array module.

The PV array module produces a current which reduces the average power consumption Pw_{ave} of the zero standby power PIR luminaire. It increases the illumination intensity and measures the average power consumption with different PV array areas. The experiment results show that the power produced by the PV array module within an area of 15x15 cm^2 is equal to Pw_{ave} when there is sufficient illumination (150 lx) [4].

There are several PIR lighting devices of different brands in an electric appliance store. In three products denoted as A, B and C, a 5 W lamp is fixed to these products and this design individually measures the power consumption in both the standby state and the lighting state. This measurement is carried out in a darkroom. The result is shown in Table I.

TABLE I
COMPARISON OF POWER CONSUMPTION OF THIS DESIGN WITH OTHER PIR DEVICES

Type	Standby state power	Lighting state power
Zero standby power PIR luminaire	0.001 W	5.04 W
PIR lighting device A	1.2 W	6.4 W
PIR lighting device B	1.5 W	6.6 W
PIR lighting device C	1.4 W	6.5 W

IV. CONCLUSION

Although the standby power of a PIR luminaire is not great, it affects the electricity bill in the long run. This paper proposes a new circuit design which reduces the standby power substantially. Moreover, the power consumption is much less than that of other PIR lighting devices. This new design, the zero standby power PIR luminaire, which consumes 0 W when the illumination intensity suffices for the PV array and only consumes 1 mW when not, is both easy to set up and inexpensive. In the long run our design saves more power whilst the performance of the luminaire remains unchanged.

REFERENCE

[1] U.S. Department of Energy, "Energy Saver Booklet: Tips on Saving Energy & Money at Home," May. 2009.

[2] Ying-Wen Bai, and Yi-Te Ku, "Automatic room light intensity detection and control using a microprocessor and light sensors," *IEEE Trans. Consumer Electron.*, vol.54, no.3, pp.1173-1176, August 2008.

[3] Cheng-Hung Tsai, Ying-Wen Bai, Wang Hao-Yuan and Ming-Bo Lin, "Design and Implementation of a Socket with Low Standby Power," *IEEE Trans. Consumer Electron.,* vol.55, no.3, pp.1558-1565, August 2009.

[4] Cheng-Hung Tsai, Ying-Wen Bai, Chun-An Chu and Ming-Bo Lin, "Design and Implementation of a Socket with Zero Standby Power using a Photovoltaic Array," *IEEE Trans. Consumer Electron.*, vol. 56, no. 4, pp. 2686-2693, Nov. 2010.

Dynamic Frequency Scaling based Power Saving Algorithm for a Portable Kitchen TV

Won-Jong Kim, Tae-Ho Roh, Kyou-Jung Son, Seong-Pil Moon, Chang-Hwan Jang, and Tae-Gyu Chang, *Senior Member, IEEE*

***Abstract*--This paper presents a dynamic frequency scaling technique to save the processing power of MPEG-4 video decoder implemented in a portable kitchen TV, which is an ATSC/T-DMB supporting combo platform. The operating frequency of CPU is dynamically changed in accordance to the estimated processing burden of each video frame.**

I. INTRODUCTION

Along with the rapid demand hike for high quality video in recent mobile multimedia terminals, the portion of power consumed by video processing grows as well [1][2].

This paper presents an application of dynamic frequency scaling technique to save the processing power of MPEG-4 video decoder implemented in a portable kitchen TV, where the CPU operating frequency level is changed according to the workload of each video frame. This is considered as an advanced hardware-software collaboration in the sense that the hardware adjustment is performed rapidly to the video frame interval, i.e., several tens of millisecond [3]. Considering that workload of video decoding generally shows wide and rapid variations, the proposed technique can best utilize these characteristics of video frame to maximize the power-saving effect.

The frame-based dynamic frequency scaling is applied to the MPEG4 video decoding in the portable kitchen TV system, which is a combo TV platform supporting both ATSC (Advanced Television System Committee)[4] and T-DMB (Terrestrial Digital Multimedia Broadcasting). The portable kitchen TV optionally provides a file-play mode for personal multimedia service with stored MPEG-4 video.

A prototype optional video player is implemented using a processor based software decoder to provide flexibility with the personal multimedia function. Because of the increased consumption power in the video processing, the feasibility of the processor-based software decoder heavily depends on the availability of an effective power reduction technique.

To find the feasibility of adopting the optional multimedia service in the portable kitchen TV, the proposed technique is applied to the MPEG-4 H.264 software decoder implemented on ARM9 processor, and its power reduction is experimentally measured.

II. POWER SAVING ALGORITHM FOR PORTABLE KITCHEN TV

A. Frame-based DFS algorithm

Processor consumption power is proportional to its operating frequency[5-7]. The basic principle of the proposed frame-based DFS algorithm is to estimate workload of each frame and its idle time, then apply the lowest possible operating frequency in accordance with the estimated workload of each frame, equivalently with the idle time of each frame.

Least-mean-square (LMS) based linear prediction filter is used to estimate the workload of each frame. Prediction error caused by drastic change of scenes may cause wrong selection of operating frequency resulting in excess of decoding time above a frame duration. The selection of operating frequency reflects this decoding time excess as shown in (1).

$$f_{opt} = \frac{\hat{W}_{next}}{\tau - t_{excess}} \tag{1}$$

where $\hat{W}_{next}$ is the estimated workload of the next frame, τ is the single frame duration, and t_{excess} is the last frame's excess decoding time. Frame reservoir technique using a play buffer is also provided to handle the problem of prediction error and decoding time excess.

For further exploitation of the DFS technique, one frame duration is divided into two parts then two different frequencies are applied as given in (2).

$$T_S = \frac{\hat{W}_{next} - f_2(\tau - t_{excess})}{f_1 - f_2} \tag{2}$$

where T_S is the timing index of the boundary of the two parts and f_1, f_2 are the operating frequencies of two consecutive levels applied to the frame decoding. By applying (2), the idle interval can be minimized for the maximum power saving under the limited number of discrete operating frequencies in the embedded processor environment.

B. Implementation of the power saving algorithm for a portable kitchen TV

The flow chart of the frame-based DFS algorithm is shown in Fig.1. The algorithm is implemented on the ARM9-based video processor of the portable kitchen TV. The functional block diagram of the kitchen TV platform is shown in Fig.2.

The operating frequency is changed by controlling the system timer and the PLL register. The resolution of the μC/OS kernel timer is set to 1.0msec, and an application timer function is added for the frame-wise switching of the operating frequencies. The supporting frequencies of the processor are 100MHz, 200MHz, and 400MHz.

C. Performance test of the power saving algorithm

The picture of the developed prototype portable kitchen TV is shown Fig 3. The combo TV supports both ATSC and T-DMB for mobile TV service, and an optional MPEG-4 video player.

For the feasibility test of the proposed DFS algorithm's power saving effect, amount of saved energy is measured for the three test video files of sixty seconds having QCIF resolution and 15 fps. Consumed power is summed from the recorded profile of applied frequencies. The average consumed power shows 6% decrease when it is compared to the consumed power of the decoding without applying DFS.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

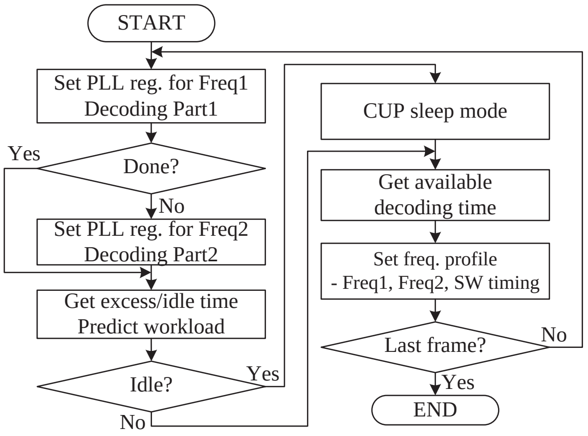


Figure 1. The flow chart of the proposed power saving algorithm.

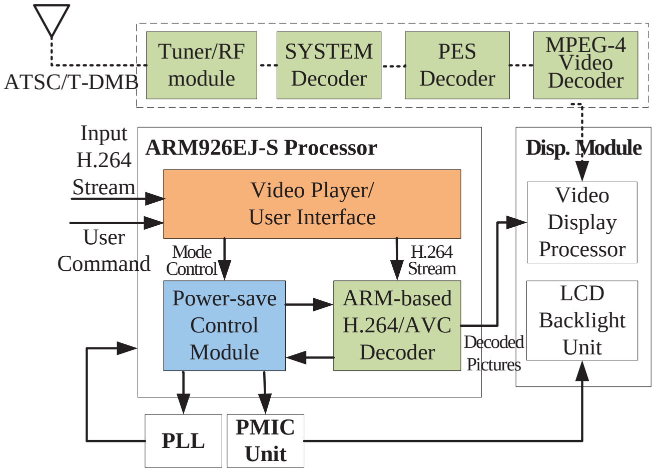


Figure. 2 The functional block diagram of the portable kitchen TV.

For detailed and comparative illustration of the processing profiles, the histogram of idle time is plotted in Fig.4. The number of frames is measured and plotted for each different idle time or excess time in units of 2.0msec. Each frame's decoding time is measured by counting processor cycles during decoding the frame and dividing the counted cycles by the operation frequency.

The average idle time is measured as 31.24msec when the DFS is not applied, and it is reduce to 4.52msec when the DFS algorithm is applied. It is confirmed that the number of frames having idle time and that of having excess time are balanced to show the drastically reduced 4.52msec average idle time resulting in minimum power consumption by applying the proposed algorithm. On the other hand, it is well illustrated that all the frames are skewed to have only idle times when the DFS is not applied.

III. Conclusion

This paper presents a frame-based DFS algorithm to save the processing power of MPEG-4 software decoder. A prototype portable kitchen TV is implemented and the proposed power saving algorithm is applied to its software decoder.

It is confirmed that the proposed algorithm gives average 6% of power save for the tested three QCIF video files.

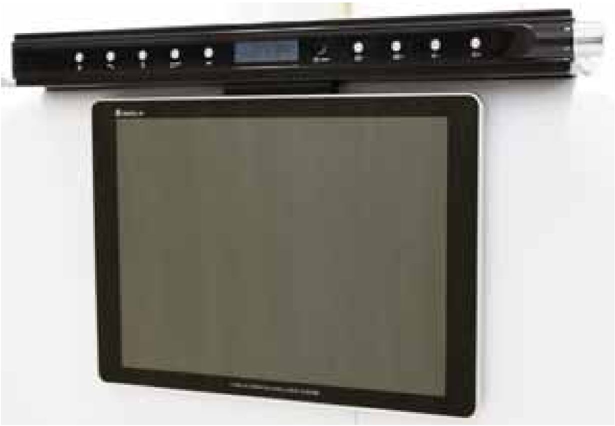


Figure 3. The implemented prototype portable kitchen TV.

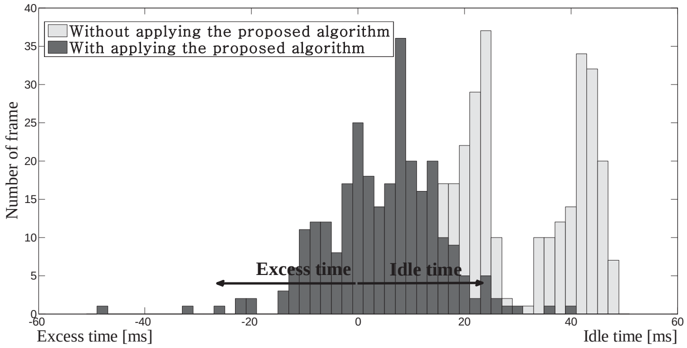


Fig 4. Histograms of measured excess / idle time.

It is also confirmed from the detailed measurement of the processing profiles that the proposed algorithm minimizes the average idle time.

Considering that the supported frequency levels are only three, i.e., 100MHz, 200MHz and 400MHz in the current implementation of the prototype TV system, the proposed power saving algorithm will contribute much more power saving if the supported frequency levels of the processor are increased.

Reference

[1] Jae-Sik Lee, Jong-Hoon Jeong, and Tae-Gyu Chang , "An efficient method of Huffman decoding for MPEG-2 AAC and its performance analysis", IEEE Transactions on Speech and Audio Processing, Vol.13, No.6, November 2005.

[2] Tse-Hua Lan, Tewfik, A.H, " A resource management strategy in wireless multimedia communications-total power saving in mobile terminals with a guaranteed QoS", IEEE Transactions on Multimedia, Vol.5, pp. 267-281, June 2003.

[3] Chandrakasan, A.P.; Gutnik, V. Xanthopoulos, T., "Data driven signal processing: an approach for energy efficient computing," International Symposium on Low Power Electronics and Design, pp. 347-352, Aug. 1996.

[4] ATSC Digital Television Standard, Part 3 –Service Multiplex and Transport Subsystem Characteristics, Advanced Television Standard Committee, Aug. 2009.

[5] T. Bud, T. Pering, A. Stratakos, and R. Brodersen, "A dynamic voltage scaled microprocessor system," Proceedings of IEEE Intemational Solid-State Circuits Conference, pp. 294-295,2000.

[6] Rabaey, Chandrakasan, Nikolic', "Digital Integrated Dircuits A Design Perspective", Prentice Hall, 2003.

[7] Seongsoo Lee; Sakurai, T., "Run-time Voltage Hopping for Low-power Real-time Systems," Proceedings of Design Automation Conference, pp. 806-809, June 2000.

Novel Level-up Shifters for High Performance and Low Power Mobile Devices

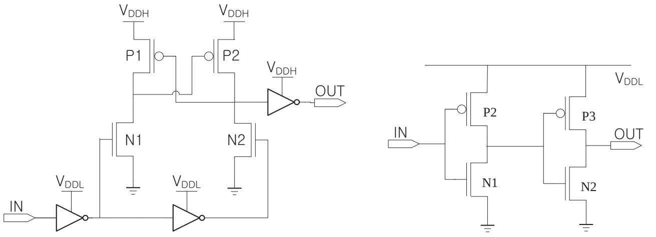
Dong-Ik Jeon and Kwang-Soo Han and Ki-Seok Chung
Department of Electronics and Computer Engineering, Hanyang University, Republic of Korea

***Abstract*--This paper presents novel level-up shifters called Dual Step Level-up Shifter (DSLS) and Stacked Dual Step Level-up Shifter (SDSLS) which are simpler, yet more efficient than conventional level-up shifters. We compare the proposed designs with two existing designs: a conventional level-up shifter and Contention-Mitigated Level Shifter (CMLS). The delay of the proposed designs is less than that of a conventional level-up shifter and CMLS by up to 4.86% and 6.51%, respectively. The power consumption of the proposed designs is less than that of a conventional level-up shifter and CMLS by up to 6.68% and 5.40%, respectively. DSLS and SDSLS act as a power gating circuit as well as a level-up shifter. Thus, we conclude that our proposed designs are very effective for low power designs.**

I. INTRODUCTION

Delay and power consumption are important in mobile application processors (APs), since the market size of mobile devices such as smart phones and tablet PCs is growing fast [1]. Low power design techniques with multiple voltage supplies are commonly used in the mobile APs where a high voltage region is used for high performance and a low voltage region is used for low power consumption to reduce the dynamic power consumption [2], [3]. When two blocks in different voltage regions are interacting with each other, we need to have an interface to adjust the voltage levels. Such interface logic is called level shifters. Among level shifters, a level-up shifter has the worse delay and power consumption, since a level-up shifter is more complicated than a level-down shifter. Therefore, active studies have been conducted to improve level-up shifters. In this paper, we propose novel level-up shifters called Dual Step Level-up Shifter (DSLS) and Stacked Dual Step Level-up Shifter (SDSLS) which are simpler, yet more efficient than conventional ones.

II. DSLS AND SDSLS

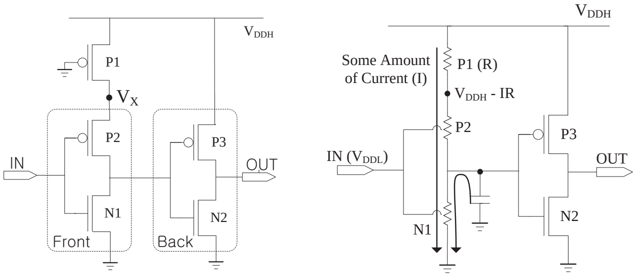


(A) Conventional Level-up Shifter (B) Conventional Level-down Shifter

Fig. 1. Conventional Level Shifters

"This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2012-H0301-12-1011)

Fig. 1 (A) and (B) show a conventional level-up shifter and a conventional level-down shifter, respectively. The conventional level-up shifter which consists of 10 transistors is more complicated than the conventional level-down shifter. The conventional level-up shifter is complicated mainly because of its differential signaling structure. Since it takes more effort to level up to a high voltage region, the conventional level-up shifter stably levels up voltage by the differential signaling structure. Thus, the conventional level-up shifter has large delay and power consumption. For high performance and low power consumption, Dual Step Level-up Shifter (DSLS) and Stacked Dual Step Level-up Shifter (SDSLS) are proposed**.**



(A) Structure of DSLS (B) Pull-down Operation of DSLS

Fig. 2. Dual Step Level-up Shifter (DSLS)

Fig. 2 (A) shows the proposed DSLS. DSLS has a stepping level-up structure where each inverter of the buffer structure is supplied by different voltages. The supply voltage of the back inverter which consists of P3 and N2 is V_{DDH}. On the other hand, the supply voltage of the front inverter which consists of P2 and N1 is lower than V_{DDH}. As shown in Fig. 2 (B), when the input is V_{DDL}, P2 stays ON due to the low voltage so that some current flows along the path of P1, P2 and N1. Thus, the voltage of the V_X is V_{DDH}-IR, because of the voltage drop in P1 transistor channel.

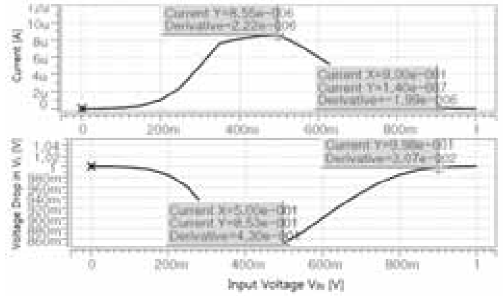


Fig. 3. Some Amount of Current and Voltage Drop in V_X of DSLS

Fig. 3 shows the amount of current and the amount of voltage drop at node V_X according to the change in V_{DDL} in Fig. 2 (A) when V_{DDH} is set to 1.0V. From Fig. 3, we confirm

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

that the voltage drop in V_X is adequate so that DSLS may be used for a level-up shifter by employing a stepping level-up. However, due to significant amount of current, DSLS is not appropriate to be used in low power designs with multi-V_{DD}. Therefore, we employ transistor stacking [4] in the DSLS to reduce the leakage power consumption while maintaining the good delay characteristics of DSLS.

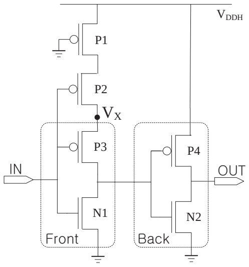


Fig. 4. Stacked Dual Step Level-up Shifter (SDSLS)

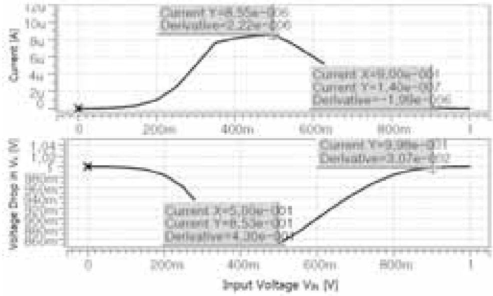
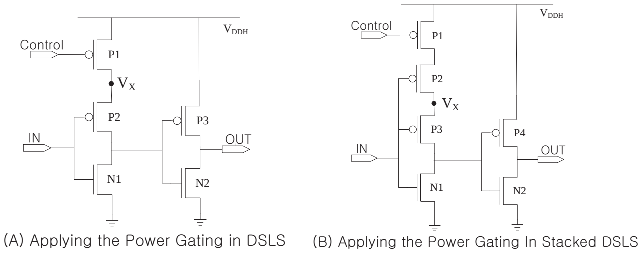


Fig. 5. Some Amount of Current and Voltage Drop in V_X of SDSLS

Fig. 4 shows the proposed SDSLS which reduces the amount of current. However, the amount of current strongly influences the voltage drop in V_X for a stepping level-up. From Fig. 5, the voltage drop of SDSLS in V_X is bigger than DSLS. The reason is that P2 has a huge resistance, since the gate of P2 is driven by V_{DDL}. However, V_X of SDSLS does not change according to the change in IN due to a huge resistance in P2. Therefore, we can maximize the performance of the level-up shifter by applying DSLS when the input voltage is more than 0.8V and applying SDSLS when the input voltage is less than 0.8V.



(A) Applying the Power Gating in DSLS (B) Applying the Power Gating In Stacked DSLS

Fig. 6. Some Amount of Current and Voltage Drop in V_X of SDSLS

Fig. 6 shows that DSLS and SDSLS can act as a power gating circuit, if the gate of PMOS (P1) which is always-on is driven by the control signal instead of the ground. Thus, applying the power gating in DSLS and SDSLS can cut off the signal from the previous block to the next block. This is very effective in low power designs such as a multi-voltage system.

III. SIMULATION

To evaluate the quality of the proposed design, we compared the proposed design with two existing designs: a conventional level-up shifter and CMLS [5]. We designed a simple mobile AP system with multiple supply voltages. We set the supply voltage of a cache memory to 0.7V and the supply voltage of a register to 1.0V. We used SDSLS as level-up shifters in the system design, since the voltage difference between a cache memory and a register is 0.3V.

TABLE I
COMPARISON OF LEVEL-UP SHIFTERS

	Conventional	CMLS	SDSLS
Rising [ns]	34.90	34.97	33.09
Falling [ns]	10.37	11.10	9.989
Delay Average	22.64	23.04	21.54
Power [mW]	214.1	211.2	199.8

Table I shows that the delay of the proposed design is less than that of a conventional level-up shifter and CMLS by up to 4.86% and 6.51%, respectively. The power consumption of the proposed design is less than that of a conventional level-up shifter and CMLS by up to 6.68% and 5.40%, respectively.

IV. CONCLUSION

Recently, low power consumption has become one of the most critical design issues. Designing a system with supply voltages is one of the most popular low power design techniques. For two regions with different voltage domains to interact properly, level shifters are required. In this paper, we proposed novel level-up shifters called Dual Step Level-up Shifter (DSLS) and Stacked Dual Step Level-up Shifter (SDSLS). In order to maximize the performance, we can use DSLS when the input voltage is more than 0.8V. On the other hand, we can use SDSLS when the input voltage is less than 0.8V. DSLS and SDSLS can act as a power gating circuit as well as a level-up shifter. Therefore, we conclude that our proposed designs are very effective for low power designs.

V. REFERENCES

[1] A. P. Chandrakasanet.al., “Minimizing power consumption in digital CMOS circuits,” Proceedings of the IEEE, vol.83, no.4, pp.498-523, 1995.

[2] A. Gayasen, K. Lee, N. Vijaykrishnan, M. Kandemir, M.J. Irwin, and T. Tuan, “A Dual-VDD Low Power FPGA Architecture”, Field-Programmable Logic and Applications – FPL, pp. 145-157, 2004

[3] Sarvesh H. Kulkarni, Dennis Sylvester, “Power distribution techniques for dual VDD circuits”, ASP-DAC '06 Proceedings of the 2006 Asia and South Pacific Design Automation Conference, pp. 838-843, 2006

[4] Siva Narendra, Vivek De, Dimitri Antoniadis, Anantha Chandrakasan, Shekhar Borkar, “Scaling of stack effect and its application for leakage reduction”, International Symposium on Low Power Electronics and Design - ISLPED , pp. 195-200, 2001

[5] Canh Q. Tran, Hirosh Kawaguchi, Takayasu Sakurai, “Low-power High-speed Level Shifter Design for Block-level Dynamic Voltage Scaling Environment”, IEEE International Conference on Integrated Circuit Design and Technology - ICICDT , 2005

Accurate GPU Power Estimation for Mobile Device Power Profiling

Minyong Kim, *Student Member*, *IEEE*, and Sung Woo Chung, *Member*, *IEEE*
Korea University, Seoul, Korea

***Abstract*—This paper not only describes the importance of GPU power estimation for smartphone power profiling, but also proposes GPU power estimation technique to enhance accuracy. The proposed technique improves the accuracy by up to 17.6%.**

I. INTRODUCTION

Smartphone power profiling techniques identify power hungry hardware components and applications [1]-[3]. The techniques help smartphone consumers to reduce power consumption. For example, users can prolong the battery lifespan by turning off unnecessary applications with massive power consumption. Even more, application developers can reduce power consumption of the applications by finding out the specific hardware component with excessive power consumption. In addition, smartphone vendors can optimize the device by replacing high power consuming hardware components with less power consuming ones.

Power profiling techniques must guarantee sufficiently high accuracy (typically, higher than 90%). Since most smartphone power profiling techniques build the power model by breaking down measured system power [1]-[3], the techniques need to model every hardware component with significant power consumption. Otherwise, the techniques may suffer from inferior accuracy, since the omitted hardware power modeled as a constant is a variable in fact.

In recent smartphones, Graphics Processing Unit (GPU) dissipates noticeable power. Most recent GPUs in smartphones have more number of cores operating at higher clock frequency levels, compared to past smartphones. Furthermore, the GPUs are more frequently utilized to provide enhanced Graphical User Interface (GUI). However, many power profiling techniques have not modeled GPU power assuming that GPU power consumption is negligible. Thus, the techniques show impractical accuracy when GPU is utilized.

In this paper, we first analyze the power behavior of the smartphone GPU to develop a GPU power model. Then, we show that GPU power model leads to higher profiling accuracy by comparing with a conventional smartphone power profiling technique without the GPU power model.

II. GPU POWER CONSUMPTION MODELING

In this section, we first show how frequently GPU is used with real smartphone applications. We then model GPU power consumption by analyzing the power consumption of four applications with different GPU utilizations on a brand-new dual-core smartphone. To exclude transient power fluctuation due to Dynamic Voltage Frequency Scaling (DVFS) and Dynamic Power Management (DPM), we manually disable both DVFS and DPM. At the same time, we keep the power state of the hardware components other than CPU, GPU and display at idle mode. To isolate GPU power consumption from the measured *system* power, we use an accurate power modeling methodology, proposed in [1].

A. GPU Applications

Though many conventional power profiling techniques omit GPU power model [1]-[3], GPU gets more frequently utilized

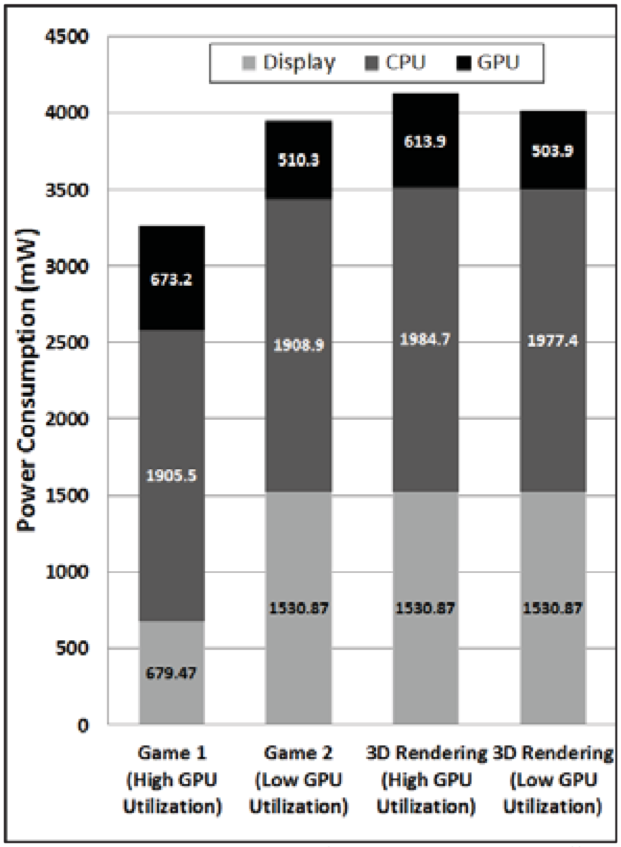


Fig. 1. GPU Power Consumption of Four Applications with Different GPU Utilization (average GPU utilization of game 1, game 2, 3D rendering 1, and 3D rendering 2 is 59.0%, 50.2%, 42.4%, and 38.6%, respectively)

This work was supported by the Basic Science Research Program of the National Research Foundation of Korea, funded by the Ministry of Education, Science and Technology under contract 2012-0002444, National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No.2012-013816), the Smart IT Convergence System Research Center funded by the Ministry of Education, Science and Technology as Global Frontier Project (SIRC-2011-0031863), and the Second Brain Korea 21 Project.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

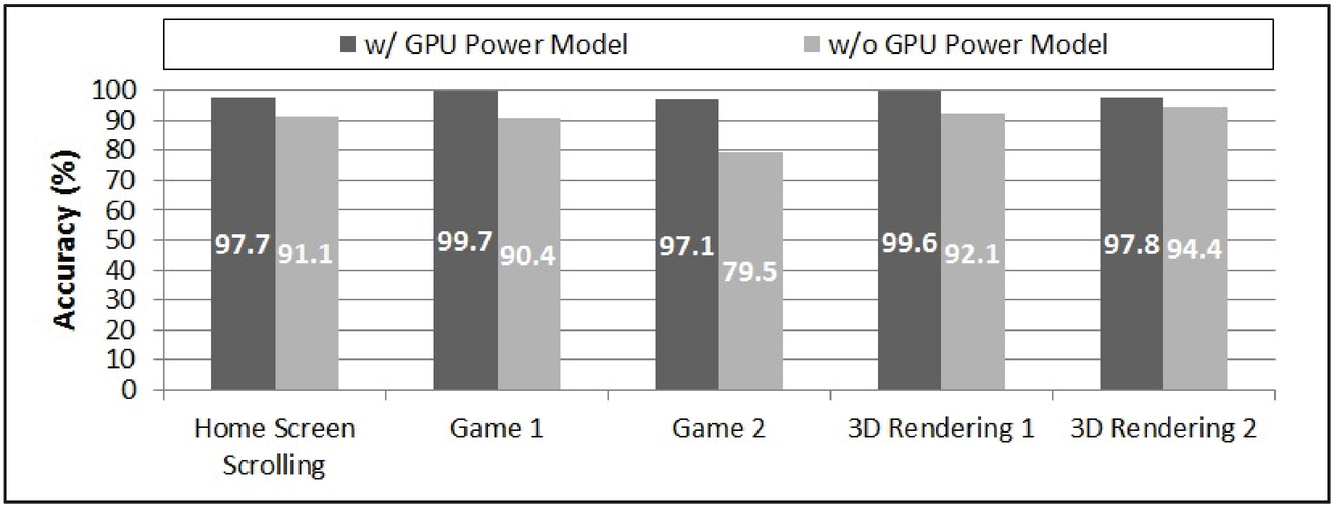


Fig. 2. Power Estimation Accuracy Enhancement by the Proposed GPU Power Model

in recent smartphones as follows: 1) home screen scrolling event is one of the most frequent events in smartphones. To accelerate the response of the event, recent smartphones utilize GPU to handle the scrolling event. 2) Game applications are the most popular applications for smartphone users. Note game applications dominate the top 5 downloads chart of the application market. 3) 3D rendering is used for many applications, even for GUI of an advanced text message applications.

B. GPU Power Consumption Analysis

Fig. 1 shows GPU power consumption of four applications with different GPU utilization. As shown in the figure, GPU power consumption accounts for up to 20.7% of total *system* power consumption. Thus, GPU power consumption needs to be modeled to guarantee sufficient accuracy.

GPU power consumption does not show much deviation across four applications, though the GPU utilizations of the applications vary (from 38.6% to 59.0%). Shown in Fig 1, the difference of GPU power consumption between game 1 and 3D rendering 2 is 169.3 mW, which is only around 3% of average *system* power consumption.

C. GPU Power Model

As explained in the previous subsection, GPU power consumption is not so sensitive to GPU utilization. In addition, most recent smartphones do not support DVFS and DPM on GPU. Hence, we model GPU power consumption as shown in (1) with one variable that simply indicates GPU on/off status (GPU_{on}). In (1), β_{GPU} is the coefficient obtained by simple regression analysis after running GPU applications.

$$\text{GPU Power} = \beta_{GPU} \cdot GPU_{on} \qquad (1)$$

III. EXPERIMENTAL RESULTS

In this section, we show how much power estimation accuracy is improved by adding the GPU power model. We examine one use case and four applications: home screen scrolling, two games, and two 3D rendering applications. We compare the power estimation accuracy between [1] (w/o GPU Power Model in Fig. 2) and [1] with the proposed GPU power model. To calculate accuracy, we compare the power profile obtained from each technique with that obtained from a hardware power monitoring device. Note that the power profiling technique [1] accurately estimates the power consumption of dual-core smartphones for applications not using GPU.

As shown in Fig. 2, our GPU power model improves the power estimation accuracy by up to 17.6% (game 2). Moreover, the error rate with GPU power model is decreased, not exceeding 3% in any case.

IV. CONCLUSION

In this paper, we propose the GPU power model to improve the estimation accuracy of smartphone power profiling techniques. Though many smartphone power profiling techniques ignore GPU power consumption, we find that GPU power consumption is significant (accounts for up to 20.7% of total *system* power) in some cases. Our experimental results show that our GPU power model revitalizes the conventional power profiling technique on recent smartphones, by substantially enhancing power estimation accuracy (by up to 17.6%).

REFERENCE

[1] M. Kim, J. Kong, and S. W. Chung, "Enhancing online power estimation accuracy for smartphones*," IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, May 2012.

[2] L. Zhang et al., "Accurate online power estimation and automatic battery behavior based power model generation for smartphones," *Proc. of 8th CODES+ISSS*, October 2010.

[3] A. Pathak, Y. C. Hu, M. Zhang, P. Bahl, and Y.-M. Yang, "Fine-grained power modeling for smartphones using system call tracing," *Proc. of the 6th Conference on Computer Systems (EuroSys'11)*, April 2011.

Low-complexity Depth Map Generation for Real-time 2D-to-3D Video Conversion

Chan-Hee Han[1], Si-Woong Lee[2], and Hyun-Soo Kang[3]

Abstract--**A low-complexity depth generation algorithm for 3D conversion of 2D videos is presented. For temporal consistency in global depth, a pattern-based depth generation method is introduced. In addition, refinement of depth from the combined depth cues of texture and motion is also addressed. Experimental results show that the proposed method outperforms previous algorithms in terms of the complexity and subjective quality**.

I. INTRODUCTION

Stereoscopic 3D display technology enables viewers to experience natural depth impression of the observed scenery by providing stereo images simultaneously. In case of 2D videos, the realistic sense of the scene can also be provided by converting them into stereoscopic videos using available depth perception cues. This 2D-to-3D conversion technology attracts considerable interest because it allows the reuse of enormous 2D contents for 3D. The most typical depth cues include focus, geometry and motion. Recently, fusion-based approaches which combine multiple depth cues to one depth map are being presented for more realistic depth generation. The challenge in this approach is how to integrate each cue reasonably to generate stable depth [1]. A priority depth fusion algorithm in [2] is based on a weighted sum of three depths. In [3], a real-time video conversion system based on human visual perception is presented. Since 3D conversion of 2D video is an ill-posed problem, generation of natural and comfortable depth is more meaningful than generation of complex real depth. In particular, temporal consistency in depth generation is of great importance because temporal discordance of depth causes terrible fatigue to viewers.

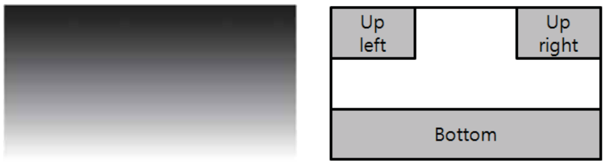
This paper proposes a low-complexity depth generation algorithm for robust 2D-to-3D video conversion which consists of two-step approach. First, the pattern-based depth generation algorithm is applied to obtain temporally consistent global depth. Then, the depth in object region is refined from the combined depth cue of texture and motion.

II. PROPOSED ALGORITHM

A. Background depth generation

The depth from geometry can be considered as the global depth because it reflects well the overall structure of the scene. Vanishing lines and vanishing points are the most commonly used geometry cue. But, the analysis process for vanishing point detection requires high computational burden. Furthermore, inaccuracies in the extracted vanishing points lead to severe temporal discordance in final depth.

To avoid these problems, pattern-based global depth generation method is newly proposed in our system. Among some representative depth patterns, the most suitable one is selected and used as the global depth for current frame. In this way, the computational complexity can be reduced, but temporal discordance is yet to be solved. Therefore, a single basic pattern in Fig.1(a) that is a kind of top-down gradual pattern is employed in the proposed method. In order to reflect the proper depth gradient for the current scene both in horizontal and vertical directions, the pattern modification process is applied to the basic pattern based on feature differences between corner regions of images. The detailed procedure for background depth generation is as follows.



(a) (b)

Fig. 1. (a) Basic pattern. (b) Corner regions.

① Calculate the average values of color and edge in the shaded regions in Fig. 1(b).

② Calculate the differences

$$\begin{aligned} diff_{UL-B} &= |avg_{UL} - avg_{B}| \\ diff_{UR-B} &= |avg_{UR} - avg_{B}| \\ ediff_{UL-B} &= |eavg_{UL} - eavg_{B}| \\ ediff_{UR-B} &= |eavg_{UR} - eavg_{B}| \end{aligned} \tag{1}$$

where avg_i and $eavg_i$ denote the average values of color and edge in region i, respectively

③ Allocate depth values for four corners as in (2).

$$\begin{aligned} depth_{UL} &= depth_{B} - [\alpha \times (diff_{UL-B} + ediff_{UL-B}) + offset] \\ depth_{UR} &= depth_{B} - [\alpha \times (diff_{UR-B} + ediff_{UR-B}) + offset] \end{aligned} \tag{2}$$

where $depth_B$ is a fixed depth value for two bottom corners which are considered to be the closest to the camera. $depth_{UL}$ and $depth_{UR}$ are flexible depth values for upper left and upper right corners.

④ Interpolate the depth of each pixel with an appropriate interpolation method using depth values of four corners.

This research was financially supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation. (No. 2012H1B8A2025982)

[1] Graduate School of IC, Hanbat National University, Korea. [2] Corresponding author, Dept. of ICE, Hanbat National University, Korea. [3] College of ECE, Chungbuk National University, Korea.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Object depth refining

The basic concept of object depth refinement is based on two observations. One is that the object region is generally in focus and appears sharp in the image. Another one is that the motion parallax in front objects tend to be larger than background region. Based on this concept, we combined focus and motion cues to compute the object depth as in (3)

$$D_{i,j}^{obj} = \sum_{x=0}^{ws}\sum_{y=0}^{ws} G_{x,y} \cdot [\omega \cdot D_{i-ws/2+x,\, j-ws/2+y}^{focus} + (1-\omega) \cdot D_{i-ws/2+x,\, j-ws/2+y}^{motion}] \quad (3)$$

where $D_{i,j}^{obj}$ represents the object depth at (i, j). $G_{x,y}$ is the Gaussian filter coefficient, and $D_{i,j}^{focus}$ and $D_{i,j}^{motion}$ denote depth-from-focus and depth-from-motion, respectively. ω is a weight for depth-from-focus. Note that Gaussian filtering is applied to generate spatially comfortable depth. To reduce the computational complexity, we made use of simple but effective measures for $D_{i,j}^{focus}$ and $D_{i,j}^{motion}$. Since humans strongly perceive the depth at boundaries of objects, the Sobel edge operator is employed for focus measure. The accurate motion estimation is pointless in depth generation since motion cue is unstable due to complex motion of camera and background. This is why motion cue should be fused with other cues. Actually, whether there is a relatively large motion or not would be more meaningful than accurate motion itself. Based on this observation, we employed simple frame difference for $D_{i,j}^{motion}$ in place of complex motion estimation such as the feature tracking or BMA.

Finally, the object depth is fused with global depth as in (4).

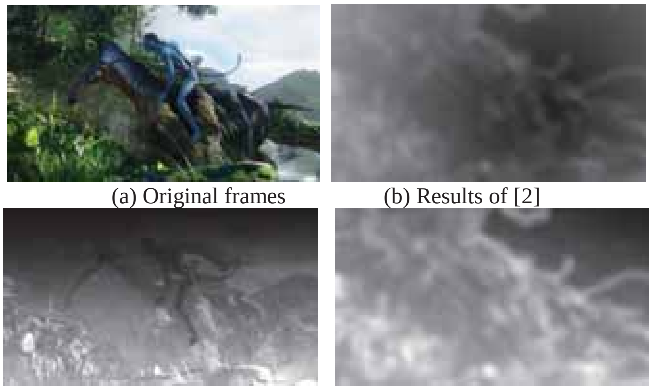
$$D_{i,j}^{final} = \varepsilon \times D_{i,j}^{obj} + (1-\varepsilon) \times D_{i,j}^{back} \quad (4)$$

where $D_{i,j}^{back}$ and $D_{i,j}^{final}$ denote the background depth and final depth, respectively. ε is a weight for the object depth. In proposed fusion, fixed weights are used because flexible weight causes temporal discordance and high cost.

III. EXPERIMENTAL RESULTS

The test sequences are parts of “Avatar” which is the most well-known commercial 3D movie. The spatial resolution is 1280 x 720 and the frame rate is 23Hz. For the performance evaluation of the proposed method, we compared previous approaches of [2, 3] and ours in terms of processing time and subjective quality. A personal computer with Intel quad-core i5 CPU and 4 GB RAM is used to measure processing time.

Fig. 2 shows the results of depth generation by 3 methods. Note that the background depth of [3] has only vertical gradient whereas the proposed method shows both horizontal and vertical gradients of global depth quite well. Table I shows processing time and subjective quality of each method. Fifty college students who have viewed the 3D movie assessed subjective quality of each method for three evaluation indices: stereoscopic effect, temporal consistency and spatial stability. In conclusion, the proposed method and [3] have a feasible complexity for the 3D viewing device and proposed method is more effective than [3] in subjective quality.



(a) Original frames (b) Results of [2]

(c) Results of [3] (d) Results of proposed method

Fig. 2. A comparison between final depths of each method

TABLE I
THE PROCESSING TIME AND SUBJECTIVE QUALITY

(a) The processing time

Method	[2]	[3]	proposed
Average processing time(sec/frame)	1.122	0.389	0.416

(b) The score for subjective quality evaluation

Grade	Excellent	Good	Normal	Bad	Terrible
Score	5	4	3	2	1

(c) The subjective quality of each method

	original	[2]	[3]	proposed
Stereoscopic effect	4.9	4.1	3.3	3.8
Temporal consistency	4.8	2.2	4.5	4.5
Spatial stability	4.8	2.1	4.2	4.3
Average score	4.87	2.8	4	4.2

IV. CONCLUSIONS

This paper proposed a novel low-complexity depth fusion algorithm for 2D-to-3D video conversion. For temporal stability, we proposed a pattern-based global depth generation method. Since the depth values of two upper corners can be adjusted according to feature differences between the upper and bottom regions, the structure of the scene can be reflected to the generated depth map effectively. To refine the depth in object region, we proposed a low-cost object depth computation method based on frame difference and edge information. In experiments, we showed that the proposed method can achieve low complexity and realistic 3D effects at the same time.

REFERENCES

[1] L. Zhang, C. Vazquez, and S. Knorr, "3D TV content creation: automatic 2D-to-3D video conversion", *IEEE Trans. Broadcast*., vol. 57, no. 2, pp. 372-383, 2011.

[2] Y.L. Chang,, J.Y. Chang,, Y.M. Tsai, C.L. Lee, and L.G. Chen, "Priority depth fusion for 2D-to-3D conversion systems", *Proc. SPIE*, vol. 6805, pp. 1-8, 2008.

[3] S.F. Tsai, C.C. Cheng, C.T. Li, and L.G. Chen, "A real-time 1080p 2D-to-3D video conversion system", *IEEE Trans. Consumer Electronics*, vol. 57, no. 2, pp. 915-922, 2011.

Real-time 2D to 3D Conversion for 3DTV Using Time Coherent Depth Map Generation Method

Seung-Woo Nam†, Hye-Sun Kim†, Yoon-Ji Ban†, and Sung-Il Chien‡
†Creative Content Research Laboratory, Electronics and Telecommunications Research Institute
‡School of Electrical Engineering and Computer Science, Kyungpook National University

Abstract--**Depth map flickering of object in the scene makes human feel fatigue for 2D-to-3D video conversion on 3DTV. In this paper, we present a fast and robust scheme to generate depth map from monocular video without depth map flickering. The proposed method is to divide input video sequence into the cuts and assign initial depth map and segment object in the image by using color and motion history. Mixed information of color and motion is applied to diminish depth error. The experimental results show that our scheme achieves real-time performance and reduce human eye fatigue on 3DTV.**

I. INTRODUCTION

Recently, many researchers have developed automatic and real-time 2D to 3D conversion method. However, lack of a non-fatigue 3D content generation approach is dilemma for 3D display industry. If any method is satisfied by two constraints, then the method is sufficient to be applied to 3DTV. The first is a real-time issue and the other is coincidently elimination of human eye fatigue for a long watching time. Therefore, we proposed an automatic and real-time 2D to 3D converting method reducing human eye fatigue from big changes of depth between image frames.

II. PROPOSED SYSTEM

Proposed system uses color information of 2D input movie to generate a 3D stereoscopic movie automatically. Using motion information of the stereo effect varies according to the amount of motion [2]. The stereo effect can't be expressed if there is no camera movement or a static scene due to lack of information [3]. So we compute segment grouping in every frame using color to generate 3D depth information.

However, color segmentation has a big problem because there is not consistency between consecutive frames. The segmentation without consistency can make depth map flickering and make unstable stereo movies. For these reasons, proposed system focuses on maintaining time coherency using {n-(n-1)} difference map. Fig.1 shows the overview of proposed system. Our system uses the difference map (red pixel area) to generate depth map of current frame.

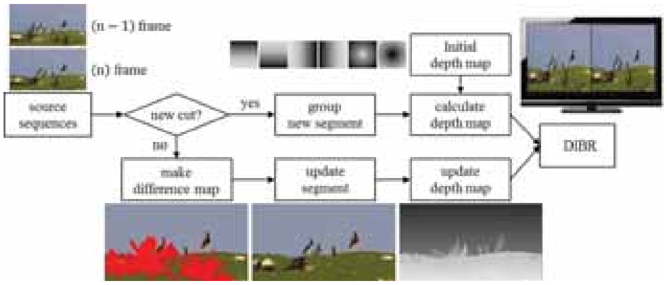


Fig. 1. Overview of the 2D to 3D auto conversion system

After generating depth maps, blur the depth maps with bilateral filter and re-render to stereoscopic images [4]-[6].

A. Cut Based Processing

2D to 3D auto conversion is processed based on a cut. If n_{th} frame of 2D input is a new cut [7], [8], we make a depth map through generally used color segmentation and depth assigning. The depth is assigned by initial depth gradient hypothesis depending on the position of region [1]·

If n_{th} frame of 2D input is not a new cut, we update segment area and depth of previous frame in order to maintain the consistency of depth map. This process is made up using difference map.

B. Segmentation Update

Within the same cut, segmentation processing of n_{th} frame changes area groups as much as necessary from the results of the $(n-1)_{th}$ frame segmenting. As limiting segmentation seed pixel position only in difference region, previous segment groups having no change can be remained. Thus, the n_{th} frame and $(n-1)_{th}$ frame are linked over the time on result of segmentation. The result of usual way of segment grouping shows in fig.2 (a), (b), (c). Subtle changes in color of the background area of n_{th} frame, due to many changes in segmentation. This can cause a flickering error. In contrast, (d), (e), (f), results of our method, can be reliably observed that segment groups have continuity.

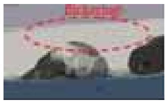
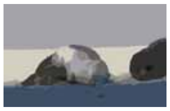
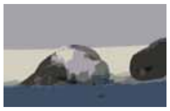
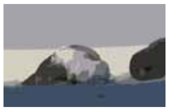
	$(n-1)_{th}$ frame	n_{th} frame	$(n+1)_{th}$ frame
Previous method			
	(a)	(b)	(c)
Ours			
	(d)	(e)	(f)

Fig. 2. Segmentation results: (a)-(c) show that flickering occurs. (d)-(f) show consistency of segmentation depending on time frame.

Korea Creative Content Agency(kocca) in the Culture Technology(CT) and Research Development Program 2012.

C. Depth Map Update

The depth value of a segment area is set to predetermined rule, depending on their size and position. In this paper, find the location of the centroid of segment area and apply initial depth gradient hypothesis [1]. However, the size and the position of segment area may change very frequently. And only if apply a simple depth generation rule, almost every frame depth values are randomly changed. It will not be able to ensure continuity of frames.

In this paper, to ensure time coherency of depth map, depth map generation refers to the value of difference map. Only the area having change will be updated. If a percentage of changed pixel number in a segment area is less than a threshold, that segment is considered as having little change, and depth value of the previous frame is used. If not, calculate a new depth value as our rule to assign. It ensures time coherency and stability of depth map. Fig.3 (b), (c) present examples of flicking of depth map at not using difference map. On contrast, image (e) and (f) refer to difference map, show depth maps very stable.

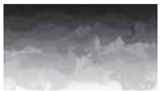
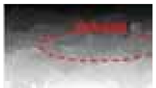
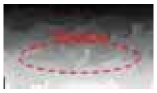
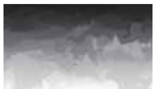
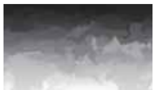
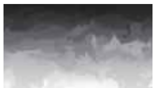
	(n-1)$_{th}$ frame	n$_{th}$ frame	(n+1)$_{th}$ frame
source			
Non Using difference map			
	(a)	(b)	(c)
Using difference map (ours)			
	(d)	(e)	(f)

Fig. 3. Results of depth map generation: (d)-(f) show that flickering reduces comparing with (a)-(c) depending on time frame.

III. EXPERIMENTAL RESULT

A. Depth Map Continuity

The experiment is designed to prove how our method can generate stable depth maps with time coherency. Ratios of depth changes are calculated with comparing every depth map sequences, and it's represented in Fig.4. A red line is a result of ours using difference map, and a blue graph line is a result of general method do not consider time continuity of depth maps. As a result, our method can generate depth maps updating gradually, so 3DTV viewers are able to feel more comfortable.

B. Real-Time Processing

Our proposed method is implemented as 8-core parallel processing. It's finished to test that 1280x720 resolution movies can be converted automatically in real-time. Processing time, depending on the resolution of the source movies can show that nearly linear increase. In future, if our method is mounted on 3DTV as a chip, it can convert HD movies in real-time also. An experimental H/W spec. is Intel® Zeon® CPU X5690, 3.47 GHz, 8GB RAM and Window 7 64bit OS.

IV. CONCLUSION

In this paper, we proposed a method of real-time 2D to 3D conversion on 3DTV. We used difference maps that help to maintain time coherency among every frames. In addition, we configured a processing pipeline with simple operations. Therefore, it can convert 2D to 3D movies in real-time, and the converted image sequence reduced the human eye fatigue.

REFERENCES

[1] C. Cheng, C. Li, , and L. Chen , "A 2D-to-3D Conversion System Using Edge Information", *in Proceedings of IEEE International Conference on Consumer Electronics*, pp.377-378, Jan. 2010.

[2] W. J. Tam and L. Zhang, "3D-TV Content Generation: 2D-to-3D Conversion", *IEEE International Conference on Multimedia and Expo(ICME)*, pp.1869-1892, 2006.

[3] I. Ideses, L.P. Yaroslavsky and B. Fishbain, "Real-time 2D to 3D video conversion", *Journal of Real-Time Image Processing*, vol. 2, Issue 1, pp. 3-9, 2007.

[4] L. Zhang and W. J. Tam, "Stereoscopic Image Generation Based on Depth Images for 3D TV", *IEEE Trans. on Broadcasting*, vol.51, no.2, 2005.

[5] W. J. Tam and L. Zhang, "3D-TV Content Generation: 2D-to-3D Conversion", *IEEE International Conference on Multimedia and Expo(ICME)*, pp. 1869-1872, 2006.

[6] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images", *International Conference on Computer Vision*, pp.839-846, 1998.

[7] J. Mas and G. Fernandez, "Video Shot Boundary Detection Based on Color Histogram", *TRECVID Workshop*, 2003.

[8] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques", *Journal of Electronic Imaging*, vol. 5, no. 2, pp. 122-128, 1996.

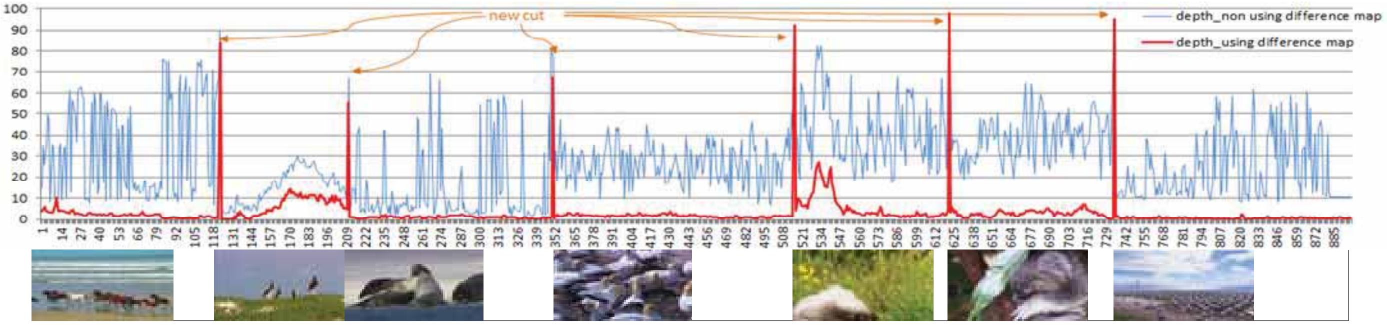


Fig. 4. Comparison of the depth change ratio: generated depth maps using difference map are more coherent than those without using difference map for the 900-frame video(wild life). Red line(ours) is more coherent than blue one for the time sequence.

Depth Adjustment of ROI using Scalable Multiplexing in 3D Display

Hong-Chang Shin, Gi-Mun Um, Chan Kim, Won-Sik Cheong, and Namho Hur
Electronics and Telecommunications Research Institute, Korea

***Abstract*— In this paper, we presents a depth adjustment technique that controls depth of region of interest using scalable multiplexing in 3D display. Autostereoscopic displays offer different 3D images depending on the viewing direction. In order to display various viewpoint images on the autostereoscopic display at the same time, we usually divide the space of display panel spatially. This is called spatial multiplexing. However, this spatial multiplexing causes degradation of each viewpoint image resolution. Each viewpoint source is resized with the ratio of the inverse proportion to the number of view due to limited resolution of 3D display. When the images are downscaled with a certain ratio and, the downscaling ratio of the object of interest is bigger than that of the others, the object of interest is enlarged more than the other region without any loss. Moreover, enlarged object region covers the other regions belong to the background of image. In this case, the covered area can offer enough space to make the object that has binocular disparity without any interpolation of holes. It can also control the object's depth in the 3D scene.**

I. INTRODUCTION

In real world, when an object comes closer to viewer, the object becomes bigger. Thus, the change of the size of an object is one of the factors give the perception of 3D depth.

Likewise, when the depth of ROI(Region of interest) or OOI(Object of interest) is adjusted in 3D display, the size of an object should also be considered. This is because it is no wonder that the size of an object coming closer to the viewer becomes enlarged as in a real world.

Binocular disparity is one of important cues when human perceive the depth in a real world. Binocular disparity is provided to the viewer when each eye of the viewer gets different view images, respectively. Most of flat panel 3D displays are using this binocular disparity.

Autostereoscopic display is one of the 3D displays. It shows two or more views that have binocular disparities at the same time, so that it can allow the viewers to freely rotate their head and even move around.

In order to provide viewers with autostereoscopic 3D contents, essentially, images captured from various viewpoints should be inputted to the 3D display. And the number of cameras required is equal to the number of viewpoints. In addition, critical problems need to be resolved, such as synchronizing the cameras, processing and transmitting a huge amount of data. Due to these problems, recent researches have focused on view synthesis techniques which generate various intermediate views using limited cameras. Recently, from the aspect of contents provider, stereo-to-multi-view conversion is the most appealing approach because this technique enables to adapt HD stereoscopic 3D contents to the various types of autostereoscopic 3D contents [1]. In general, the generation of intermediate views from stereoscopic 3D contents is based on disparity map estimated from the two input views. With the disparity map, additional views can be generated by applying depth image-based rendering technique (DIBR) [2].

After obtaining multi-view images, we need to rearrange each pixel of those images according to the pixel arrangement structure of target autostereoscopic display [3]. This is called multiplexing.

Autostereoscopic displays divide the space of display panel and display two or more views at the same time. It makes the resolution of the multi-views images reduced inversely proportional to the number of views. This means spatial multiplexing causes the downscaling of the individual source image. In practice, the resolution of the images that the eyes receive is smaller than the resolution of the individual source images.

When each source image is spatially multiplexed, ROI can be enlarged using scalable multiplexing with different downscaling factor. The occluded area by enlarged object or region offers enough space to give the ROI binocular parallax.

In this paper, we describe a way to adjust depth of the ROI using scalable multiplexing in the middle of multi-view generation process in the autostereoscopic 3D display system. Fig. 1 illustrates an overview of our proposed system.

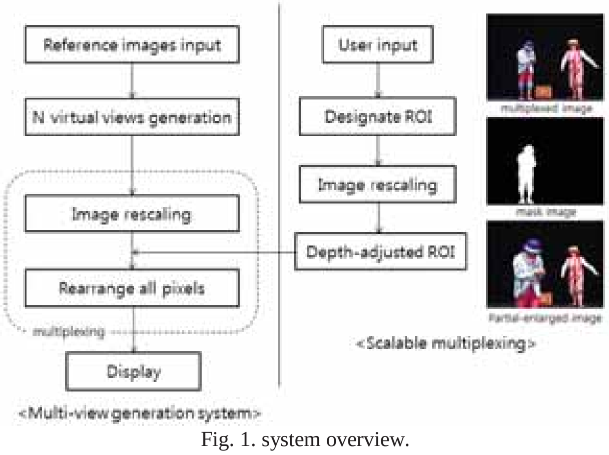


Fig. 1. system overview.

II. DEPTH ADJUSTMENT

A. Designation of ROI

At first, ROI or OOI in source image should be designated. It can be designated in rectangular-shaped or circle-shaped form by user input. And it can also be extracted by image

This research was supported by the KCC (Korea Communications Commission), Korea, under the ETRI R&D support program supervised by the KCA (Korea Communications Agency)"(KCA-2012-11921-02001).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

segmentation. In spite of various of researches for several decades, image segmentation remains a challenging problem in computer vision. However, during the multi-view generation process, the object of interest can be segmented using disparity map. In this paper, we assume that segmented ROI is given in advance.

B. *Scalable Multiplexing*

As mentioned above, downscaling is an irrevocable step in the process of spatial multiplexing due to limited resolution of display. There are many ways to multiplex source images in 3D display and it is different according to display type. The most important thing is that the each resolution of multiplexed image becomes downscaled. Assuming that the resolution of input source images is the same as the resolution of display, the resolution of each view image is reduced to about 1/N of source image. (N is the number of viewpoints). Each viewpoint images are downscaled with inverse proportion to the number of views of the display. Fig. 2 illustrates the resizing of source each view source image in the process of multiplexing.

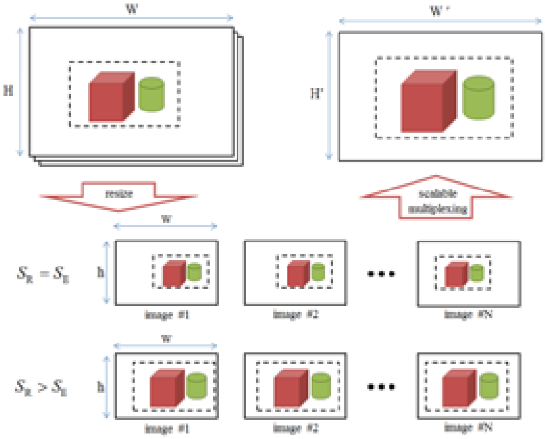


Fig. 2. Scalable multiplexing.

As shown in Fig. 2, let the resolution of source image and multiplexed image be $I_W \times I_H$ and $O_w \times O_h$, respectively. When source image I becomes resized, if the downscaling ratio of ROI is bigger than the other regions, the ROI becomes relatively enlarged more than the other regions. The scaling factor of ROI S_R has the range represented by the following.

$$S_E = \frac{O_w \times O_h}{I_W \times I_H} \le S_R \le 1. \ (O \le I)$$

C. *Depth Adjustment of ROI*

As mentioned above, different scaling factor of ROI makes it enlarged without having resolution deterioration. When the ROI becomes enlarged, the enlarged area covers the other regions belong to the background of scene. As shown in Fig. 2, the red-colored object is covered by the blue-colored object that is enlarged over it. This covered area can be used as margin space to give a binocular disparity of the object. The blue-colored object is enlarged $\frac{w'}{w}$ times horizontally compared to the red-colored object. So it can have enough range of [d_l , d_r] to shift the enlarged object leftward and rightward within depth budget without any interpolation of disoccluded holes.

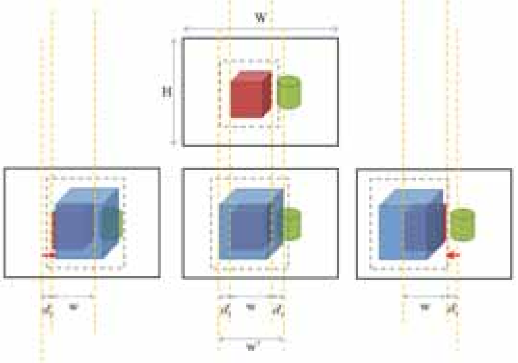


Fig. 3. Depth adjustment of ROI.

III. Experimental Result

Fig. 4 shows an example of the enlarged ROI by the proposed scalable multiplexing. And the enlarged object can be moved in both ways within certain range.

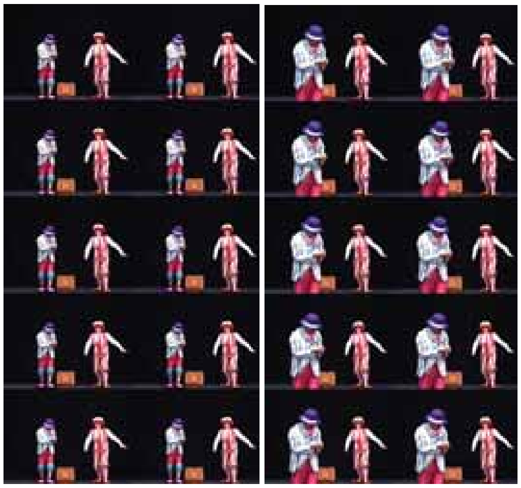


Fig. 4. Enlarged ROI using scalable multiplexing (N=10). Left: $S_E = S_R$, Right: $S_E < S_R$.

IV. Conclusions

In this paper, we propose a novel depth adjustment technique that controls the depth of ROI using scalable multiplexing for the autostereoscopic 3D display. Proposed scalable multiplexing makes ROI enlarged without having resolution deterioration. Moreover, enlarged object or region can have enough space margins for adjusting depth of it.

References

[1] http://www.hhi.fraunhofer.de/en/departments/image-processing/applications/real-time-stereo-to-multiview-conversion/

[2] C. Fehn, “A 3D-TV approach using depth-image-based rendering (DIBR)”, in Proc. VIIP 03, Benalmadena, Spain, Sep. 2003.

[3] E. Lueder, *3D Displays*, John Wiley & Sons, Ltd, Chichester, UK, 2011.

Joint Complexity and Rate Optimization for 3DTV Depth Map Encoding

Sebastiaan Van Leuven*, Hari Kalva†, Glenn Van Wallendael*, Jan De Cock*, and Rik Van de Walle*
*Multimedia Lab, Dep. of Electronics and Information Systems, Ghent University-IBBT, B-9050 Ledeberg Ghent, Belgium
Email: {sebastiaan.vanleuven; glenn.vanwallendael; jan.decock; rik.vandewalle}@ugent.be
†Dept. of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL, United States
Email: hari.kalva@fau.edu

***Abstract*—Current research towards 3D video compression within MPEG requires the compression of three texture and depth views. To reduce the additional complexity and bit rate of the depth map encoding, we present a fast mode decision model based on previously encoded macroblocks of the texture view. Meanwhile we present techniques to reduce the rate based on predicting syntax elements from the corresponding texture view. The proposed system is able to get a reduction in complexity of 71.08% with an average bit rate gain of 4.35%.**

I. Introduction

Currently, MPEG is standardizing 3D Video [1]. The goal is to improve compression compared to the Multiview Video Coding (MVC) extension of MPEG-4/AVC -H.264 (annex H) [2]. Next generation 3D devices are targeted. Such devices, like autostereoscopic displays, might require depth information to generate a large amount of intermediate views using view interpolation [3], [4]. Also more common stereo displays will benefit from the standard. However, these displays might not require depth information. Consequently, the depth is encoded solely for a limited set of users. Therefore the overhead, both in complexity and rate distortion (RD), should be limited. This preliminary research is based on the MVC extension and is aimed to reduce both the complexity and bit rate in H.264/AVC based 3D video standardized encoders.

Current research is focusing on 3D video encoding using the MVC extension, a base view is encoded as a regular H.264/AVC (AVC) bitstream, while additional views are encoded using the decoded output of the base view as an additional prediction frame. Additionally, depth can be encoded although this requires a relatively high complexity. In [5] the complexity of the depth map encoding is reduced, but the RD is not improved. An even more complex scheme is presented in [6], where the views are warped before prediction. However, such system require an even higher complexity, and also hardware design is more complex. Moreover, the applied warping should also be standardized. Optimizations have been proposed to reduce the syntax [7], however, the complexity is not reduced. Therefore we propose a model to reduce both depth map encoding complexity and the bit rate, while the only minor implementation changes are required.

II. Methodology

The MPEG CfP considers a three and two view case, for eight sequences and four rate points per sequence (resulting in 64 encoded bit streams). Our system is evaluated for the three view case, therefore the data of two sequences (*Poznan_Hall2* and *Kendo*) is analyzed for the highest and lowest rate point. This analysis results in our proposed model. Study of the syntactical overhead showed potential RD improvements. Both the reduced complexity and syntactical improvements have been evaluated with eight HD sequences (*Poznan_Hall2*, *Poznan_Street*, *Undo_Dancer*, *GT_Fly*, *Kendo*, *Balloons*, *Loverbird1*, and *Newspaper*) and four rate points for the three view case. The test conditions (GOP size, quantization, intra period) from the MPEG CfP are used.

III. Proposed Method

A. Complexity Reduction

An analysis of the mode distribution of the depth map results in the probability that a mode is selected in the depth view based on the mode in the co-located macroblock of the corresponding texture view. This probability is given by: $p = P(MODE_{Depth}|MODE_{Tex})$. Where $MODE_{Depth}$ is the macroblock mode used in the depth and $MODE_{Tex}$ the macroblock mode used in the co-located macroblock in the corresponding texture view. The analysis shows that when $MODE\_Skip$ or $MODE\_16 \times 16$ are selected in the texture, this mainly results in $MODE\_Skip$ or $MODE\_16 \times 16$ being selected in the depth. Based on this analysis, the set of probable modes for the depth is given by :

$$MODE_{Depth} = \begin{cases} S & \text{if } MODE_{Tex} \in S \\ A & \text{if } MODE_{Tex} \notin S \end{cases}$$

Where S is the subset of unpartitioned modes $S = \{MODE\_Skip, MODE\_16 \times 16\}$ and A the set of all modes $A = \{MODE\_Skip, MODE\_16 \times 16, MODE\_16 \times 8, MODE\_8 \times 16, MODE\_8 \times 8, MODE\_Intra\}$. After evaluating the probable modes, the most RD-optimal mode is selected. The analyzed sequences achieve a high accuracy, 96.76%, using this model (Table I).

B. Rate Distortion Improvement

Due to the low bit rate of depth maps, syntactical data has a high impact. For the low rate points of the MPEG CfP sequences, the syntactical data in depth maps accounts for approximately 50%. Next to the complexity reduction, our

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

TABLE I
ACCURACY OF THE PROPOSED MODEL.

	Poznan_Hall2		*Kendo*	
	R1	R4	R1	R4
accuracy (%)	98.76	96.34	97.59	94.37

TABLE II
NUMBER OF MACROBLOCKS FOR WHICH THE PROPOSED MODEL RESULTS IN LESS SYNTACTICAL DATA.

	Poznan_Hall2		*Kendo*	
	R1	R4	R1	R4
Accuracy (%)	37.09	59.48	59.54	68.09

TABLE III
COMPLEXITY REDUCTION OF THE PROPOSED MODEL (AVERAGE=71.08%) FOR LEFT (L), CENTER (C), RIGHT VIEW (R) AND TOTAL SEQUENCE.

Test Sequence	L	C	R	overall
Poznan_Hall2	72.62	70.80	70.81	71.16
Poznan_Street	76.10	70.35	72.16	72.18
GT_Fly	71.61	72.43	71.66	72.03
Undo_Dancer	66.15	68.29	70.55	68.46
Kendo	69.16	68.49	67.90	68.45
Balloons	74.91	73.00	72.45	73.20
Loverbird1	78.87	70.29	69.29	71.59
Newspaper	77.38	73.64	71.22	73.41

proposed technique is also able to reduce the *mb_type* signaling. The syntax element *mb_type* indicates the macroblock type and *mb_skip_flag* indicates whether a macroblock is a skip macroblock. Since *mb_type* of the depth map is based on the mode of the texture the *mb_type* does not have to be transmitted for every macroblock. When $MODE_{Tex} \in S$, the depth map macroblock type can be determined based on the *mb_skip_flag* in the depth. When *mb_skip_flag* $= 1$, the macroblock type of the depth is skipped, otherwise it will be $MODE\_16 \times 16$. Consequently, *mb_type* can be omitted for such cases. Table II shows the accuracy of this reduction for the analyzed sequences.

Furthermore, for each macroblock an *end_of_slice_flag* is transmitted, indicating if the current macroblock is the last macroblock of the slice. We propose to use the same slice configuration for texture and depth, so the data corresponding to the *end_of_slice_flag* should not be transmitted. Additionally, all data corresponding to chroma prediction can be neglected. This include the *intra_chroma_pred_mode* flag and chroma residual data.

IV. EXPERIMENTAL RESULTS

Results are calculated on a cluster consisting of 2.27 GHz dual quad core nodes. The software is compiled in g++ 4.1.2 for 64 bit. Each sequence is executed as a single thread. The complexity reduction is expressed as the time saving (TS) for encoding with our proposed model (T_{Fast}) compared to the reference encoder ($T_{Original}$), and is given by (1).

$$TS\ (\%) = \frac{T_{Original}\ (ms) - T_{Fast}\ (ms)}{T_{Original}\ (ms)} \tag{1}$$

Results of the complexity measurements can be found in Table III, the average complexity of four rate points per view is given because the complexity reduction is relatively constant. The overall complexity reduction is 71.08%, so only 28.92% of the complexity is required compared to the original encoder. Since the enhancement layer complexity depends on $MODE_{Tex}$, the complexity reductions are not constant.

Metrics to evaluate the quality of multiple 3D depth maps are not yet commonly used and available. Therefore, the proposed system is evaluated by the RD. Significant bit rate gains (up to 13.52%) are noticed, while a comparable quality is maintained. Worst case, a reduction of 2 dB is measured(33.1 dB to 31.1 dB for *Newspaper*). On average, 4.35% in bit rate reduction is achieved. In general, the proposed system results in higher bit rate reduction for low rate points due to the impact of the syntactical data.

V. CONCLUSIONS

To lower the cost for depth maps in 3DTV systems, the encoding complexity and bandwidth of depth maps should be reduced. Based on an analysis of encoded depth maps we propose a model to reduce the number of modes for depth map encoding depending on the selected mode in the corresponding texture view. Because of the nature of the proposed scheme, we are also able to reduce the bandwidth of the resulting bit stream. An implementation of our model shows that an average complexity reduction of 71.08% is achieved. Meanwhile, the total bit rate for the depth maps shows an average reduction of -4.35%, with a negligible quality loss.

REFERENCES

[1] MPEG, "Doc. MPEG-W12036: Call for Proposals on 3D Video Coding Technology," Tech. Rep., MPEG, Mar. 2011.

[2] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264 and ISO/IEC 14496-10 Advanced Video Coding, Edition 5.0 (incl. MVC extension)," Tech. Rep., MPEG / ITU-T, March 2010.

[3] K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, and Y. Yashima, "Multiview video coding using view interpolation and color corrections," *IEEE Tran. Circuits and Systemd for Video Technology*, vol. 17, no. 11, pp. 1436–1449, Nov. 2007.

[4] E. Martinian, A. Behrens, J. Xin, and A. Vetro, "View Synthesis for Multiview Video Compression," in *Picture Coding Symp. (PCS)*, Apr. 2006.

[5] G. Cernigliaro, M. Naccari, F Jaureguizar, J. Cabrera, E Pereira, and N. Garcia, "A new fast motion estimation and mode decision algorithm for H.264 depth maps encoding in free viewpoint TV," in *IEEE International Conference on Image Processing (ICIP) 2011*, sept. 2011.

[6] Sang-Tae Na, Kwan-Jung Oh, and Yo-Sung Ho, "Joint coding of multi-view video and corresponding depth map," in *IEEE International Conference on Image Processing (ICIP) 2008.*, oct. 2008, pp. 2468 –2471.

[7] Ismael Daribo, Christophe Tillier, and Beatrice Pesquet-Popescu, "Motion vector sharing and bitrate allocation for 3d video-plus-depth coding," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, pp. 1–13, 2009.

Fast Joint Bit-Allocation between Texture and Depth Maps for 3D Video Coding

Byung Tae Oh, Jaejoon Lee, and Du-sik Park
Advance Media Lab.
Samsung Advanced Institute of Technology

***Abstract*—This paper presents a fast joint texture/depth bit-allocation method for 3DV coding. As compared to the previous bit-allocation schemes in which the pre-encoding process is required to determine model parameters, the proposed scheme proposed the real-time joint bit-allocation scheme without any pre-encoding process. As a result, any scene change detection for adaptive model parameter changes is also unnecessary in the proposed scheme. The simulation results show the proposed scheme is comparable to the time-consuming full-search or model-based approaches, while the additional complexity for bit-allocation is almost negligible.**

I. Introduction

In these days, free-view or 3DV system is one of the hottest issues, since it will enables us to see the arbitrary view-angle of the scene. Due to large amount of multi-view source data, multi-view plus depth map (MVD) format with the depth-image based rendering (DIBR) technique becomes its promising solution [1]. One of the major activities in 3DV system is building the efficient codec for multi-view plus depth map format.

In 3DV codec, joint bit-allocation for texture and depth map is one of the most important but challenging problems, since both components contribute the rendering view quality, simultaneously. It has been theoretically proven that more bits for texture or depth maps improve the synthesized view quality [2]. Unfortunately, there is no theoretical model for optimal bit-allocation between them, yet it is observed the optimal bit-allocation is strongly dependent on its texture/depth characteristics and coding conditions. There are several works to experimentally build a model, and assign bits based on pre-encoding and model parameters derivation [2], [3]. Those methods make satisfactory results by near optimal assignment of bits. However, they contain two serious problems: First, they require the pre-encoding process to determine the model parameters, which would not be allowable in many applications despite its low additional complexity. Second, as the more serious problem, they even require the scene change detection for efficient bit-allocation, since the different texture/depth characteristics would produce the different model parameters. It means the 3DV system must include the additional scene change detector, and frequent scene change could increase the whole encoding time a lot. Furthermore, the failure of scene change detection would lead the worse bit-allocation.

In this work, instead, we propose a new bit-allocation scheme without any pre-encoding or pre-processing stages. We simply analyze the input texture signal, and determine the near-optimal bits for depth maps. The proposed bit-allocation scheme is not as accurate as those in [2], [3], yet it is much more attractive for most real-world applications, since it does not require any pre-encoding, or scene-change detection steps. Furthermore, the complexity of the additional texture analyzer is negligible, that makes the proposed scheme to be applied for real-time applications, such as live broadcasting.

The rest of this paper is organized as follows. We first briefly mention the view synthesis distortion function, and then introduce the proposed joint texture/depth bit-allocation scheme based on the view synthesis distortion metric in Sec. II. In Sec. III, experimental results are given to demonstrate the effectiveness of the proposed scheme. Finally, concluding remarks are given in Sec. IV.

II. Proposed Algorithm

A. View Synthesis Distortion Metric

It is important to notice that depth map itself is the supplement data for view synthesis in 3DV system, since depth map is not displayed to users. Instead, it is for synthesizing arbitrary views as camera parameters do. Therefore, we are required to analyze how coding error in depth map would cause the rendering distortion. In our previous work [4], we analyze the relationship between view rendering process and coding artifacts, and introduce the simple and efficient view synthesis distortion model as

$$D_v = \sum_i \left| \frac{\alpha}{2}(D_i - \widetilde{D}_i) \left[|\widetilde{T}_i - \widetilde{T}_{i-1}| + |\widetilde{T}_i - \widetilde{T}_{i+1}| \right] \right|^2 , \tag{1}$$

where D and $\widetilde{D}$ indicate the original and reconstructed depth, respectively, and $\widetilde{T}$ represents the reconstructed texture. The subscript i shows its pixel position. Finally, α is the proportional parameter to convert the depth difference to the disparity difference, which is obtained as

$$\alpha = \frac{f \cdot L}{255} \cdot \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) , \tag{2}$$

where f is the focal length, L is the baseline between the current and the rendered view, Z_{near} and Z_{far} are the values of the nearest and farthest depth of the scene, respectively. Please refer [4] for more detailed derivation of equations.

B. Joint texture and depth map bit-allocation

It is clear that the synthesized view distortion is dependent on the texture and depth distortion, simultaneously. However, it is in general not simple to determine how much each component affects the final synthesized view distortion, independently or jointly. It is often assumed the final synthesized distortion is approximately the sum of texture distortion and depth distortion, which is theoretically proven in [4]. Its relation is described as in Eq. (3).

$$D_{syn} = pD_t + qD_v , \tag{3}$$

where D_t and D_v represent the synthesized view distortion caused by texture and depth distortion, respectively.

The goal of the paper is minimizing the total distortion in rendered views, D_{syn}, given bit budget. Then, the proposed method is motivated by the following assumption : If the optimal bit allocation holds, then the rendered view distortion change by texture (pD_t), and by depth (qD_v) with sufficiently small ΔR should be the same as

$$\frac{p\Delta D_t}{\Delta R} = \frac{q\Delta D_v}{\Delta R} . \tag{4}$$

The proof of Eq. (4) is straight-forward. If the current texture/depth bit allocation is not optimal, then we can decrease D_{syn} by assigning

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

more bits for texture (or depth), and less bits for depth (or texture). Since Eq. (1) can be approximated to

$$
\begin{aligned}
D_v &= \sum_i \left| \frac{\alpha}{2}(D_i - \widetilde{D}_i) \left[|\widetilde{T}_i - \widetilde{T}_{i-1}| + |\widetilde{T}_i - \widetilde{T}_{i+1}| \right] \right|^2 \\
&\simeq E \left| \frac{\alpha}{2} \left[|\widetilde{T}_i - \widetilde{T}_{i-1}| + |\widetilde{T}_i - \widetilde{T}_{i+1}| \right] \right|^2 \sum_i (D_i - \widetilde{D}_i)^2 \\
&= \Sigma^2 \cdot D_d \,,
\end{aligned}
\tag{5}
$$

then the Eq. (4) can be expanded as follows :

$$
\begin{aligned}
\lambda &= \frac{p \Delta D_t}{\Delta R_t} = p \frac{dD_t/dQ_t}{dR_c/dQ_t} \simeq pAQ_t^2 \\
&= \frac{q \Delta D_v}{\Delta R_c} = \frac{q}{k} \frac{dD_v/dQ_d}{dR_d/dQ_d} = \frac{q \Sigma^2}{k} \frac{dD_d/dQ_d}{dR_d/dQ_d} \simeq \frac{q \Sigma^2}{k} AQ_d^2 \,,
\end{aligned}
\tag{6}
$$

where D_d is the conventionally measured depth map distortion, Q_t and Q_d is the quantization step size for texture and depth map, k is conventional texture and depth ratio, and A is the pre-determined coefficient to determine the relationship between Quantization step and λ. For example, the coefficient A is set to 0.85 in H.264/AVC. Finally Eq. (7) can be derived from Eq. (6) and Q and QP relationship as

$$
QP_t = QP_d + 3 \log_2 \frac{p}{q} k - 3 \log_2 (\Sigma^2) \,. \tag{7}
$$

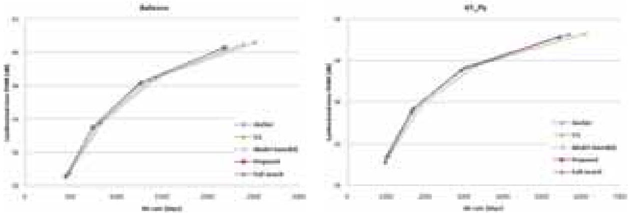
Eq. (7) clearly shows that QP_d for depth maps can be determined by QP_t and the characteristics of the corresponding texture view. In the proposed system, QP_d is determined frame-by-frame, so that any scene change detection is unnecessary. As shown the equation, with setting the texture (basic) QP value, and obtain the near-optimal texture-depth bit allocation without any pre-encoding and/or parameter derivation process. It is definitely fast and efficient for most real-life applications.

III. Experimental Results

We test the proposed method with the currently developing 3DV codec software, 3DV-ATM v0.3 with high-profile [5]. We mainly follow the common test condition (CTC) [6] : three-view cases with P-I-P view structure, GOP size 8 with full hierarchical B-structure. For B-D rate performance evaluation, the bit-rate will be total bit-rate of all multi-view color plus depth video, and the distortion will be the average PSNR of rendered views. Here, it should be noted that the ground-truth data for synthesized view is not its original view, but the synthesized view by the non-compressed texture and depth reference views as recommended in [6]. For comparison, we set the anchor with QP_t=QP_d, and compare it with the fixed ratio (texture:depth = 5:1) bit allocation, the model-based [2], the full-search, and the proposed schemes. We pre-encode the first 24 frames when it is necessary, and we vary QP_d from QP_t-6 to QP_t+8 for the full-search algorithm. For the proposed scheme, we set k=5 as fixed ratio, and set $p = q$.

First of all, we shows the R-D curves of various algorithms for 'Balloons' and 'GT Fly' sequences. As shown in Fig. 1, the proposed method is not much different from the full-search or model-based approach [2], and clearly better than fixed ratio or QP_t=QP_d case. For objective measure, we also show the B-D rate for all seven test sequences as in Table I. As before, the BD-rate performance of the proposed scheme is almost comparable to the full-search or model-based approach, while it outperforms the fixed ratio or QP_t=QP_d case.

Finally, comparison for the computational complexity of all methods is shown in Table II. It shows the complexity increase of the proposed scheme is negligible, while the fixed ratio or model-based scheme should have additional computational complexity.



(a) Balloons (b) GT Fly

Fig. 1: Comparison of R-D curves for different bit-allocation algorithms.

TABLE I: Comparison of the BD-bitrate reduction for the various schemes.

	Fixed ratio	Model-based	Proposed	Full-search
Poznan Hall2	1.07%	1.02 %	0.31%	-0.24%
Poznan Street	3.13%	-1.28%	-0.97%	-1.32%
Undo Dancer	-3.88%	-3.88%	-2.22%	-3.88%
GT Fly	6.02%	0.59%	-0.31%	-1.14%
Kendo	-9.89%	-10.27%	-10.22%	-10.27%
Balloons	-3.24%	-6.24%	-6.54%	-6.24%
Newspaper	-1.56%	-1.60%	-1.54%	-1.84%
Avg.	-1.19%	-3.09%	-3.07%	-3.56%

IV. Conclusion

In this paper, we present a fast joint texture/depth bit allocation algorithm for 3DV system. The proposed scheme enables to immediately assign the bits for depth maps without pre-encoding. The additional operation for determination of depth QP is very light, especially comparing to the encoding process. Furthermore, experimental results show the R-D performance of the proposed scheme is closed to the full-search case. As a result, the proposed scheme can be applied for various kinds of real-life applications.

References

[1] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3DTV a survey," *IEEE Transactions on Circuits System and Video Technology*, vol. 17, no. 11, pp. 16061621, Nov. 2007.

[2] H. Yuan, Y. Chang, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-D video coding," *IEEE Transactions on Circuits System and Video Technology*, vol. 21, no. 4, pp. 485–497, Apr. 2011.

[3] Y. Liu, Q. Huang, S. Ma, D. Zhao, and W. Gao, "Joint video/depth rate allocation for 3-D video coding based on view synthesis distortion model," *Signal Processing: Image Communication*, vol. 24, no. 8, pp. 661–681, Sep. 2009.

[4] B. T. Oh, J. Lee, and D.-S. Park, "Depth map coding based on synthesized view distortion function," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1344–1352, Nov. 2011.

[5] ISO/IEC JTC1 SC29 WG11 MPEG, "Test Model for AVC based 3D Video Coding," *Doc. N12558*, Mar. 2012.

[6] ISO/IEC JTC1 SC29 WG11 MPEG, "Common Test Conditions for AVC and HEVC-based 3DV," *Doc. N12560*, Mar. 2012.

TABLE II: Comparison of the encoding complexity increase for the various schemes.

	Fixed ratio	Model-based	Proposed	Full-search
Enc. complexity	109%	109 %	100 %	220 %

Optical Heart Rate Monitoring Module Validation Study

Giulio VALENTI and Klaas R WESTERTERP. Department of Human Biology, Maastricht University, Maastricht, The Netherlands

Abstract - **Optical heart rate monitoring (OHRM) offers an unobtrusive solution for continuously measuring heart rate. An OHRM prototype, able to correct for movement artifacts during physical activity, proved to be valid to continuously monitor heart rate during activities including running, allowing monitoring cardiovascular condition in response to fitness and home activities.**

I. INTRODUCTION

Photoplethysmography offers an unobtrusive solution for measuring heart rate. Movement artifacts restricted applications of this technology to confined subjects in a medical environment [1,2]. Several signal processing methods have been studied to reduce motion artifacts. The results had big delay, [3] required offline processing [4] or showed insufficient improvements in heart-rate measurement [5] and therefore were not implemented for free-living applications. A prototype of Optical Heart Rate Module[1] (OHRM, Figure 1) has been developed to be unobtrusive and allowing to correct for movement artifacts during physical activity. The electronics of the module include an accelerometer to gather data with compensation of movement artifacts in the optical signal. OHRM was validated in December 2011 with a standard electrocardiogram (ECG) as reference in order to indicate further improvements. Additionally, heart rate was monitored with a chest-strap device to compare performances between OHRM and chest-strap device[2].

II. METHODS

Subjects were 10 women and 14 men, age 28±9 year, body mass index 22.1±2.8 kg/m^2. Five were Asian (medium colored skin), one was African (dark skin) the remaining 18 were Caucasian (white skin). The standard activity protocol included, respectively, lying, standing, walking to warm up, running at increasing speed up to exhaustion, walking to cool down, and sitting. The total protocol lasted 42±5 min, where speed profile was adapted to subjects' capacities. To allow a comparison between subjects, running activity was split in three speed intervals: speed ≤9 km/h, speed between 9 and 16 km/h and speed >16km/h.

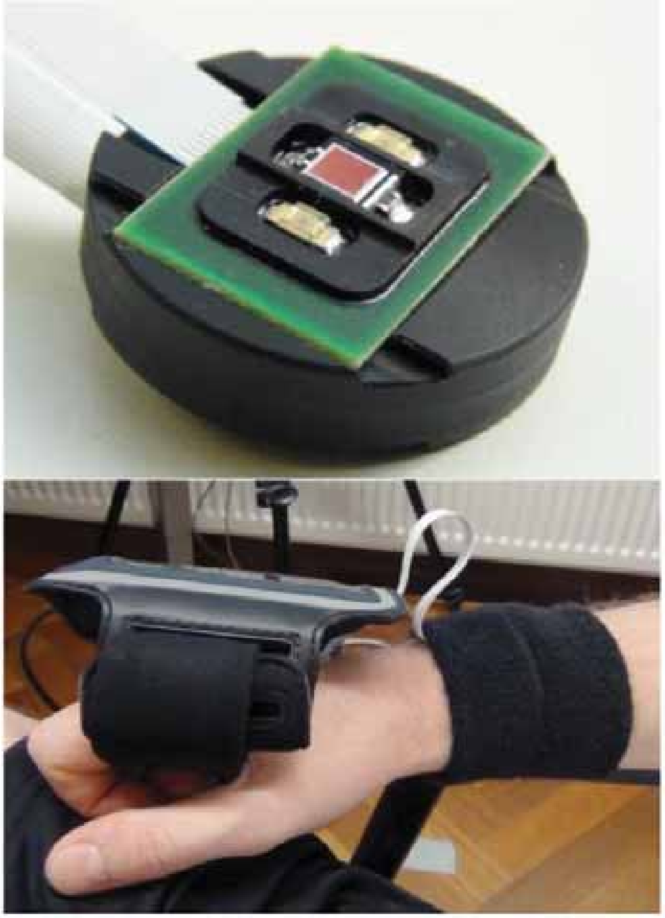


Figure 1: The optical sensor (above) and the accelerometer (below) with the connection to the optical sensor mounted on the wrist.

Each second, OHRM output consisted of a heart rate measurement and of a proprietary index of quality (D) ranging from 0 (bad quality) to 6 (good quality). The reference was a 200Hz ECG, averaging the number of beats over approximately 7s. The resulting signal was linearly interpolated and re-sampled to compare it to OHRM and chest-strap device. Chest-strap device output consisted of one heart rate measurement each second. Noisy ECG data were unreferenced and discarded. Data were valid when D≥2. Uptime was calculated as percentage of valid data. Invalid data was automatically rejected by OHRM. Chest-strap data were collected from the watch in HR mode.

[1] Alpha (Philips research, Eindhoven, The Netherlands, http://alphaheartrate.com/)

[2] Polar RS400 (Polar Electro Oy, Kempele, Finland)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Errors were calculated each second as differences between OHRM or chest-strap device and ECG. The average error over each activity and over the total time for every subject indicates accuracy. Standard deviation of errors indicates precision. Two tailed paired t-tests compared accuracy and precision between activities or between OHRM and chest-strap device. Pearson correlation coefficient (r) was used to describe the association between OHRM performances and subjects' characteristics.

III. Results

Overall OHRM showed high performances with a non-significant error of -0.1±0.3 bpm and a precision of 3±1 bpm (Table 1).

OHRM uptime (D≥2) was on average 86±14 % of the protocol time. The uptime during 'walk1' was significantly lower than the total uptime (73 % vs 86 %, $p<0.01$). Low blood perfusion of the skin after lying and standing reduced the signal amplitude and therefore the signal-noise ratio. This led to frequent data rejection. During 'run3' data uptime was also lower (55 %). During this task blood perfusion was high and heart rate was correctly measured by the OHRM. Nevertheless the index D was often lower than 2, leading to frequent rejection of accurate data. This is considered an evidence of a lack in specificity of this index during high speed running. Only 3 subjects were able to perform 'run3' therefore no t-test was performed for this activity. During activities with low uptime, accuracy and precision of the valid data were comparable to those during the total time ($p=0.26$ and $p=0.19$). This indicates that D has a high sensitivity. In this study D index was sensitive enough although there is evidence that specificity was insufficient during high speed running.

Precision tended to be higher in subjects with a higher body mass index ($r=0.35$).

OHRM had a higher accuracy (-0.1 bpm vs 0.3 bpm, $p<0.001$) but a lower precision (3.0 bpm vs 2.0 bpm, $p<0.001$) than chest-strap device. These differences were small and of no impact on any application. The absence of a belt makes OHRM less obtrusive than a chest-strap device, preserving accuracy and precision.

Improvements in the specificity of the proprietary index D could reduce the automatic rejection of valid data, increasing the uptime. Specific studies about the effect of high speed could indicate further possible improvements.

Table 1: Statistical analysis of all the activities.

	N	Uptime		O-E		CS-E	
				mean	stdev	mean	stdev
		%	min	bpm	bpm	bpm	bpm
Tot	24	86	35	-0.1	3	0.3	2
supine	24	96	10	-0.1	2	0.3	1.8
stand	24	83	4	-0.1	4.6	0.8	2.8
walk1	24	73	4	-0.5	3.7	0.3	1.9
run1	24	85	6	-0.4	1.9	0.4	1.2
run2	20	82	5	-0.4	1.8	0.3	0.9
run3	3	55	2	0.8	1.8	0.7	1.6
walk2	22	94	4	0.5	1.4	0	1.1
sit	24	94	3	0.1	1.8	0.1	1.6

N, number of subjects; Uptime, time in which the proprietary quality index (D) was >2 expressed in percentage and in minutes; O-E, difference between the heart rate as measured by the optical heart rate monitor and the electrocardiogram; CS-E, difference between the heart rate as measured by the chest-strap device and the electrocardiogram; Tot, total time of valid data; supine, lying down in supine position; stand, standing still; walk1, walking after resting; run1, running up to 9 km/h; run2, running between 9 and 16 km/h; run3, running at 20 km/h; walk2, walking after running; sit, sitting.

IV. Conclusions

This optical heart rate monitoring module is a valid and unobtrusive device to monitor heart rate, not restricted by movement artifacts during physical activities including running, allowing monitoring cardiovascular condition in response to fitness and home health care activities.

Reference

[1] Maeda Y, Sekine M, Tamura T. Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. J Med Syst 2011;35:969-76.

[2] Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas. 2007 Mar;28(3):R1-39.

[3] Kim, B. S., and Yoo, S. K., Motion artifact reduction in photoplethysmography using independent component analysis. IEEE Trans. Biomed. Eng. 2006:53:566–568

[4] Seyedtabaii, S., Seyedtabaii, L., Kalman filter based adaptive reduction of motion artifact from photoplethysmographic signal World Academy of Science, Engineering and Technology. 37 2008.

[5] Yan, Y., Poon, C., and Zhang, Y., Reduction of motion artifact in pulse oximetry by smoothed pseudo Wigner-Ville distribution. J. Neuro Eng. Rehabil. 2005:2:3.

GPU H.264 Motion Estimation with Contiguous Diagonal Parallelization and Fusion of Macroblock Processing

Fumiyo Takano and Tatsuji Moriyoshi
Green Platform Research Laboratories, NEC Corporation, Japan

***Abstract*—In this paper, we propose two methods, a contiguous diagonal parallelization and a fused execution of macroblock (MB) processing, to improve the parallel processing efficiency of GPU H.264 motion estimation. The first method can increase the degree of MB level parallelism to 2x larger with keeping coding efficiency, by relaxing the weak data dependencies. The second one can accelerate the motion estimation by a factor of 1.3, by improving the parallel processing efficiency within each MB. The evaluation results show that the processing time is reduced to 13.2 ms/frame with only 8.2% bit-rate increasing against JM16.0. As a result, the proposed methods can contribute real-time full-HD encoding with sufficiently-small bit-rate increase.**

I. INTRODUCTION

H.264, which is the latest video coding standard that can provide higher coding efficiency than conventional standards, is widely used to various applications such as video streaming and video cameras. Since the computational complexity of H.264 is significantly higher than the conventional standards, it is necessary to accelerate encoding process to encode high resolution videos, such as full-HD or higher, in real-time. One of the promising acceleration approaches is offloading heavy encoding process to GPUs (Graphics Processing Units) which consist of hundreds of processing cores and have very high computational capability with massively parallel processing. It is essential to increase the parallelism of a process to fully utilize many cores, for effective acceleration by GPUs.

We have already presented a GPU accelerated H.264 encoder[1] using CUDA (Compute Unified Device Architecture)[2] which is the software development environment for general purpose computing on GPUs. This paper presents an efficient GPU implementation of the motion estimation (ME) which is the largest process of the encoder. We propose two methods for improving the ME processing efficiency, the contiguous diagonal parallelization and the fused execution of the ME processing to reduce the total encoding time.

II. PROPOSED GPU MOTION ESTIMATION

The ME determines the optimal motion vectors (MVs) that describe the movements of objects between video frames. The ME needs to be processed in less than 16 ms/frame to encode 30 fps videos in real-time since the processing time of the ME occupies more than half of the total encoding time.

The ME processing can be parallelized using a two-level hierarchical way, macroblock (MB) level and pixel level in each MB[3]. MB is a 16x16 pixel block that is a basic coding unit of H.264. To achieve the high processing speed with such two-level parallelization, an efficient parallel processing in both inter-MB (across MBs) and intra-MB (inside of each MB) is necessary. However, the inter-MB parallelism is severely limited by the data dependencies between neighboring MBs that are caused by a property of the ME processing. As for intra-MB, the performance improvement by parallelization is restricted because there are some kinds of processing which cannot be parallelized, such as control operations. These problems should be solved to utilize the hundreds of GPU cores as many as possible.

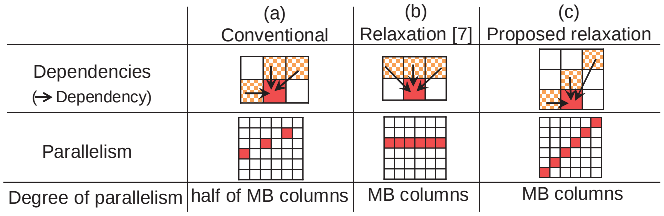


Fig. 1 Data dependencies and parallelism

Therefore, we propose the contiguous diagonal MB parallelization method to enhance the inter-MB parallelism, and the fusion of MB processing to improve the efficiency of intra-MB parallel processing.

A. *Contiguous diagonal MB parallelization*

First, we introduce the proposed contiguous diagonal MB parallelization to increase the parallelism by relaxing the data dependencies that limit the MB level parallelism. In H.264 ME[5], the predicted MV (PMV) calculated from adjacent MVs is utilized to provide the high coding efficiency by the rate-distortion optimization technique[4]. Because the PMV is derived from MVs of left, upper and upper-right MBs, there are dependencies on these MBs as illustrated in Fig.1(a). Due to this dependencies, the degree of parallelism is limited to half of MB columns, because only MBs in every other MB column can be processed in parallel[6] as colored MBs in Fig.1(a). Since the degree of parallelism for full-HD is only 60, this method is not suitable for modern GPUs which have more than 500 cores. To enhance the parallelism, a relaxation method of the inter-MB dependencies has been proposed[7]. This relaxation method uses upper-left MB instead of left MB. By using this relaxation, all MBs in a MB row can be processed in parallel and the degree of parallelism is increased to 2 times larger as Fig.1(b). However, as a result of disabling the reference to the left MB, the PMV accuracy is reduced and this also leads to degradation of coding efficiency.

Thus, we have proposed a different relaxation method which uses upper of upper-right MB instead of upper-right MB[1]. By using the proposed relaxation, MBs in a contiguous diagonal position can be processed in parallel as shown in Fig.1(c). The degree of parallelism of the proposed method is equivalent to that of the relaxation in [7]. In addition, the coding efficiency is higher than [7] because the reference to left MB, which is one of the nearest MBs and highly influences the PMV accuracy, is retained with the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

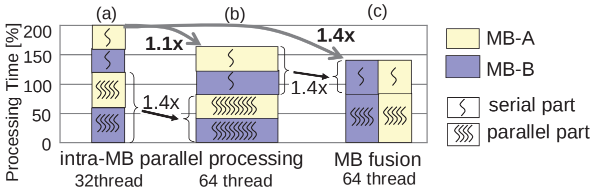


Fig. 2 Estimated processing time with different number of threads per TB

TABLE I Test conditions

Environment	NVIDIA GeForce GTX 580:512core, Intel Core i7: 2.7GHz
Parameters	Baseline Profile, QP=28,32,36,40, EPZS ME
Sequences	Full-HD (1920x1080) TUM test sequences

proposed method. The degree of parallelism for full-HD is more than 500 with the proposed method combined with additional inter-frame parallel processing[7]. This degree of parallelism is sufficiently large for current GPUs.

B. *Fusion of MB processing*

Next, we introduce the proposed fusion of MB processing to improve the efficiency of GPU processing. In CUDA, multiple threads are grouped into a “thread block” (TB). The ME process for individual MB is executed on one TB[3]. The minimum number of threads in a TB is 32[2]. However, more threads are needed in a TB for better performance since GPUs of CUDA architecture achieve the high processing throughput by hiding various latencies, such as memory access wait, by SMT (Simultaneous Multi Threading)-like execution. As a preliminary evaluation, we compared the processing time of simple vector addition processing with different number of threads (32, 64 and more) per TB on the GPU shown in Table I. The evaluation result shows that the processing speed with 64-thread/TB is 1.4 times faster than 32-thread/TB case, and no possibility of improvement in the processing speed is observed with over 64-thread/TB. From this result, the optimal number of threads per TB is expected to be 64.

To improve the processing efficiency with 64-thread/TB, the conventional GPU ME[3] enhances the intra-MB parallelism. The performance improvement by intra-MB parallelization is limited because the ME process has serial part such as PMV calculations and control operations. We estimated the ME processing time with 32 and 64-thread/TB for the clarification of the limit. Fig.2 shows the breakdown of the processing time of the ME of two MBs. As in Fig.2(a), the serial part accounts for 40% of total time with 32-thread/TB. The improvement in the processing speed with 64-thread/TB is limited to only 1.1 times since the processing time of the serial part is not reduced while the parallel part is accelerated by a factor of 1.4 (Fig.2(b)).

We propose a fusion of MB processing which can utilize inter-MB parallelism for higher performance with 64-thread/TB. The proposed method increases the number of threads per TB by processing multiple MBs in a TB, instead of by increasing the number of threads per MB. As in Fig.2(c), by using the proposed method, even the serial part is accelerated by a factor of 1.4 since the serial parts of multiple MBs can be processed in parallel. Meanwhile, the parallel processing of MBs in a TB could lead to overhead resulting from load imbalance. The processing of neighboring MBs that are expected to have similar load are fused and allocated to a TB to prevent this overhead. By using the proposed MB fusion method, up to 1.4 times speedup of the ME is estimated as shown in Fig.2(c).

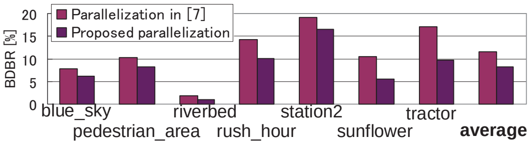


Fig. 3 Evaluation of coding efficiency

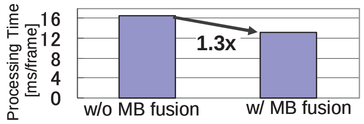


Fig. 4 Evaluation of speedup by MB fusion

III. Performance Evaluation

The coding efficiency and the processing speed of the proposed GPU ME are evaluated with test conditions shown in Table I. We evaluated the bit-rate increase against the H.264 reference encoder JM16.0[4] with BDBR[8] metric for coding efficiency comparison. Fig.3 shows the bit-rate increase results of the proposed parallelization method and the method in [7]. By using the proposed parallelization, the bit-rate increase is reduced to 8.2%, which is acceptably small for practical uses. Fig.4 shows the speedup result of the ME by the MB fusion. The ME is accelerated by a factor of 1.3 through the MB fusion processing, even including the overheads caused by the load imbalance of MBs in a TB. The ME processing time of full-HD is reduced to 13.2 ms/frame. As a result, real-time full-HD encoding is achieved.

IV. Conclusions

In this paper, we proposed two methods, the contiguous diagonal parallelization and the fused execution of the ME processing, to accelerate GPU H.264 ME. The first method can enhance the degree of inter-MB parallelism to 2x larger with keeping coding efficiency, by relaxing the weak data dependencies. The second one can accelerate the ME processing by a factor of 1.3, by improving the efficiency of intra-MB parallel processing. The evaluation results show that the processing time is reduced to 13.2 ms/frame with only 8.2% bit-rate increase against JM16.0. As a result, the propose methods can contribute real-time full-HD encoding with sufficiently-small bit-rate increase.

References

[1] T. Moriyoshi and F. Takano, "GPU Acceleration of H.264 / MPEG-4 AVC Software Video Encoder," APSIPA, 2011.

[2] NVIDIA, "CUDA Programming Guide," 2012.

[3] Y. Ko, Y. Yi and S. Ha, "An Efficient Parallel Motion Estimation algorithm and X264 Parallelization in CUDA," DASIP, pp. 1-8, 2011.

[4] A.M. Tourapis, K. Shring, and G. Sullivan, "H.264/MPEG-4 AVC Reference Software Manual," ITU-T SG16 Q.6, 2007.

[5] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," Signal Processing Mag., vol.15, no.6, pp.74-90, 1998.

[6] M.C. Kung, et al., "Block based parallel motion estimation using programmable graphics hardware," CALIP, pp.599-603, 2008.

[7] A. Obukhov, "GPU-Accelerated Video Encoding," NVIDIA GPU Technology Conference, 2010.

[8] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T VCEG M-33, 2001.

Energy-based Fair Queuing: Trading Off Energy Management and Time-constraint Meeting in Mobile Systems

J. Wei[1]*, E. Juarez*[2]*, Member IEEE, M. J. Garrido, and F. Pescador, Member IEEE*

***Abstract*--This paper extends the traditional fair queuing scheduling to the energy management domain, and presents the energy-based fair queuing scheduling, a novel class of energy-aware scheduling algorithms that support proportional energy use, effective time-constraint meeting and a flexible trade-off between them. The proposed algorithm, in combination with a mechanism that restricts the battery discharge rate, can achieve a target lifetime for Operating System (OS)-based mobile devices by including total energy consumption on all system components and systematically managing energy as the first-class resource.**

I. INTRODUCTION

OS-guided power management (PM) schemes have been widely researched to deal with the energy issue in modern mobile systems, in which a user commonly runs multiple applications simultaneously while having a lifetime in mind for a target application. Traditional OS-level PM schemes generally fall into two classes: dynamic voltage and frequency scaling and dynamic power management. They make the best effort to save energy under performance constraint, but fail to guarantee a user-specified battery lifetime, leaving the painful trading off between total application performance and battery lifetime to the user itself. In this paper, it is advocated that a strong energy-aware PM scheme should first guarantee a user-specified battery lifetime to a target application by restricting the average power of those less important applications if necessary, in addition to that, maximize the total performance of applications without harming the battery lifetime guarantee. Consequently, energy, instead of CPU or transmission bandwidth, should be globally managed as the first-class resource by the OS. The ECOSystem [1] proposed by H. Zeng is the first scheme that sets a target lifetime as the premier goal and seriously manages energy as the first-class resource. The target lifetime is achieved by dividing it into a number of epochs T_{epoch}^{i} and limiting the energy E_{epoch}^{i} in each epoch. An energy-centric scheduling is implemented to achieve proportional energy use among tasks; however, it is not well formulated and totally ignorant of time constraints. This paper follows the idea of the epoch mechanism for lifetime guarantee and proposes the energy-based fair queuing to achieve proportional energy use and time-constraint meeting.

II. ENERGY MODEL AND SCHEDULING ALGORITHM

Let us consider a system with a set of tasks $\{T_1, \dots, T_n\}$ (periodical Real-time, Interactive, and Batch) competing for a limited energy E_{epoch}^{i} during the i^{th} epoch T_{epoch}^{i}. Each task T_i is assigned a weight w_i. In an ideal model assuming energy can be simultaneously served to multiple tasks through CPU, each task runs at least with a share of power $P_i(t) = P(t) \cdot w_i / \sum_{\forall j \epsilon A(t)} w_j$, where $P(t)$ denotes the variable CPU power, and $A(t)$ denotes the set of active tasks at time t. As indicated by the equation, the power share f_i varies with the number of active tasks. Any new task joining the competition with a large weight may significantly reduce the power share of a time-sensitive task, which leads to an unstable real-time performance. To protect the power share, each time the task number changes, the weight of a time-sensitive task is fixed to its designed power share by $w_i = f_i$, and the remaining share is allocated to regular tasks based on their recalculated weights, which can be obtained by $w_i = \overline{w}_i \cdot (1 - F) / \sum_{j=k+1}^{n} \overline{w}_j$, where $\overline{w}_i$ is the initial weight of a regular task, and F is the total share reserved to k time-sensitive tasks [2]. For a periodical time-sensitive task, we define the maximum long-term power share as the average power share that guarantees meeting its total energy demand in the long term, and the worst-case power share as the one that theoretically guarantees the meeting of its maximum energy demand among all periods.

In real systems, energy is allocated to tasks along with CPU in discrete *time quanta*. The service time of task T_i is divided into a number of *service quanta* q_i^k with maximum length Q. Energy consumption during q_i^k is defined as *energy packet* e_i^k. The proposed starting-energy fair queuing (SEFQ) schedules tasks in the increasing order of the starting energy tag S_i^k, which traces the normalized energy received by task T_i. To properly compute S_i^k, a time function named *system virtual energy* $V(t)$ is defined to trace the normalized energy consumption in the system. Its value is updated to the starting tag of the currently executed task. Let $A(q_i^k)$ be the time $\mathrm{q}_\mathrm{i}^\mathrm{k}$ is requested, $S_i^k = max\,\{V(A\big(q_i^k\big), F_i^{k-1}\}$, where F_i^{k-1} is the finishing tag of q_i^{k-1}, incremented as $F_i^{k-1} = S_i^{k-1} + e_i^{k-1}/w_i$, $F_i^0 = 0$. Similar to the starting-time fair queuing [3], SEFQ provides near-optimal fairness and power guarantee under variable CPU power, but the delay, and implicitly the energy allocation error, increases linearly to the number of active tasks. Since the time-constraint meeting is sensitive to the allocation error, an overly reserved power share that is larger than the worst-case one is required for meeting all deadlines.

Borrowed starting-energy fair queuing (BSEFQ) combines a real-time friendly mechanism named warp [4] to better support time-sensitive tasks. It schedules task in the increasing order of *effective starting tag,* which is computed in a way that for a time-sensitive task it is the actual starting tag S_i^k minus a certain value named warp, and for regular tasks it equals to S_i^k. Thus, a time-sensitive task receives its share of energy earlier to meet deadlines. The priority-based scheduling is combined into SEFQ by setting different warp values, but a warp time limit is available to restrict the maximum time one task can run with priority to avoid energy starvation on unwarped tasks.

This work was supported by the Spanish Ministry of Science and Technology under grant TEC2009-14672-C02-01.

All the authors are from the *Universidad Politécnica de Madrid*, Spain. ([1]jianguo.wei@alumnos.upm.es, [2]ejuarez@sec.upm.es)

III. RESULTS

A test bench based on the producer-consumer model is designed in SystemC to assess our algorithms. For simplicity, the duration of each service quantum is set to its maximum size Q and normalized to 1 CPU time unit (Tu). A system with T_{epoch}^{i} equals to 6,300 Tus and E_{epoch}^{i} equals to 63,000 energy units (Eus) is considered. Table I characterizes the tasks considered in the simulation. Real-time and Interactive are time-sensitive tasks with periodical energy demand, while Batch 1 and 2 are regular tasks with continuous energy demand [5]. Batch 2 has an energy allocation limit of 5,000 Eus and is delayed 2,000 Tus to cause system dynamics.

TABLE I
CHARACTERIZATION OF TASKS IN THE SIMULATION

	Real-time	Inter.	Batch 1&2
Period (Tus)	7	10-15	N/A
Num. of service quanta / period	2-4	1-4	N/A
Energy packet size (Eus)	10-20	3-7	10
Max. long-term power share	0.58	0.09	x → 1
Worst-case power share	0.777	0.173	N/A

Fig. 1 compares scheduling results under SEFQ and weighted Round Robin (RR). Under SEFQ, both Real-time and Interactive tasks receive a constant long-term power share that is protected; the remaining share is allocated to Batch 1 and 2 according to their initial weights [5]. The power shares under SEFQ are workload-independent, and can be adjusted by modifying the allocated weights. However, the power shares under weighted RR are fixed once the task set is chosen and the number of active tasks is determined, since the weights of weighted RR are computed based on the task number and the average energy packet size of each task. In Fig. 1(b), when Batch 2 joins the competition, the power share of Real-time task drops due to the smaller recomputed weight; and the power share of two Batch tasks are the same because of the equal size of their energy packet.

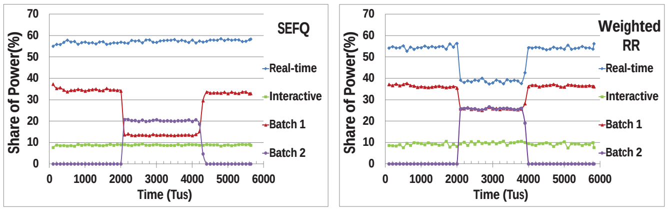


Fig. 1 Energy use under (a) SEFQ and (b) weighted Round Robin

SEFQ forces CPU into idle once E_{epoch} is exhausted, which is not appreciated by time-sensitive tasks. To extend their execution to the whole epoch and ensure exhausting E_{epoch}^{i} just at the end of one epoch, the energy allocation of Batch tasks should be properly restricted. Fig. 2 shows how the Batch 1 energy allocation affects the CPU active time. In each epoch exists an optimal point, in this case, it is 11,200 Eus.

Table II statistically compares the real-time performance under SEFQ and BSEFQ, with 2 ~ 8 active Batch tasks. Under SEFQ, the deadline meeting and response time vary with the number of active tasks because they are sensitive to the energy allocation error. However, the real-time performance under BSEFQ is not affected by allocation error. It provides strict deadline meeting if Real-time is given priority (BSEFQ_1), and achieves optimal response time if Interactive is given priority (BSEFQ_2). The performance of Interactive and Real-time is traded off when the same priority is given (BSEFQ_3).

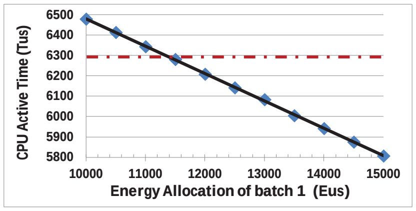


Fig.2 Relationship between Batch 1 energy allocation and CPU active time

TABLE II
COMPARISON IN PERFORMANCE OF TIME-SENSITIVE TASKS

	SEFQ	BSEFQ_1	BSEFQ_2	BSEFQ_3
Num. deadlines miss	2.5 - 13.0	0	34.8	0.03
Mean response time	3.9 - 4.1	4.2	2.5	3.8
Max. response time	10.5 - 11.3	11.9	4	11.0

* SEFQ assigns both Real-time and Interactive their worst-case power shares. BSEFQ_1 favors Real-time with a larger warp value, while BSEFQ_2 favors Interactive. BSEFQ_3 assigns Real-time and Interactive the same warp value.

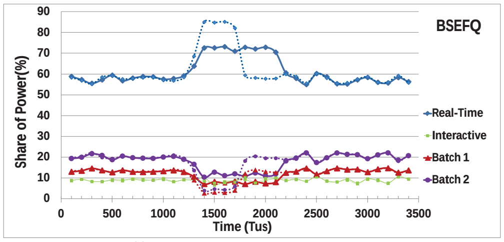


Fig. 3 Trading off power control and time-constraint meeting

Fig. 3 shows the trade-off between power control and time-constraint meeting under BSEFQ. During 1260~1680 Tus, the workload of Real-time is increased to 5 service quanta per period. All tasks start at time 0. The dotted lines represent the power shares under BSEFQ without warp time limit or under SEFQ with the worst-case power share assigned. It allows Real-time takes a maximum power share of 0.85, which nearly energy-starve Batch tasks. However, by setting the warp time limit to 4 Tus, the power share of Real-time is well controlled under 0.72 at the cost of poorer time-constraint meeting.

IV. CONCLUSION

This paper presents an energy-based fair queuing algorithm that manages energy as the first-class resource. The proposed scheme provides proportional energy use, time-constraint meeting and a trade-off between them. By combining the warp mechanism, our scheme provides more flexible and effective support for various types of time-sensitive tasks.

REFERENCES

[1] H. Zeng, C. Ellis, and A. R. Lebeck, "Experiences in managing energy with ECOSystem", IEEE Pervasive Computing, January-March 2005.

[2] J. S. Goddard and J. Tang, "EEVDF proportional share resource allocation revisited", in Work-in-Progress Sessions of the 21st IEEE Real-Time Systems Symposium (RTSSWIP00), Nov. 2000.

[3] P. Goyal, X. Guo, and H. M. Vin, "A hierarchical CPU scheduler for multimedia operating systems," Proc. Second USENIX Symposium on Operating System Design and Implementation (OSDI), 107-122, 1996.

[4] K. J. Duda and D. R. Cheriton, "Borrowed-virtual-time (BVT) scheduling: supporting latency-sensitive threads in a general-purpose scheduler," Proc. 17th ACM SOSP, pp 261-276, Dec. 1999.

[5] J. Wei, E. Juarez, F. Pescador and M. J. Garrido, "Starting-Energy Fair Queuing (SEFQ): A Novel Class of Energy-Aware Scheduling Algorithms for Mobile Systems," 16th IEEE International Symposium on Consumer Electronics (ISCE 2012). Harrisburg, USA, 4-6 June 2012.

A Remote Cardiac Monitoring System for Preventive Care

Keunjoo Kwon[1], Heasoo Hwang[1], Hyoa Kang[1], Kyoung-Gu Woo[1], Kyuseok Shim[2]
[1]Samsung Advanced Institute of Technology, Samsung Electronics, Korea
[2]School of Electrical Engineering and Computer Science, Seoul National University, Korea

Abstract--**Remote monitoring of heart disease patients has been shown to be effective for diagnosis and detection of arrhythmias. We propose a remote cardiac monitoring system for preventive care by developing a decision support system with personalized parameters and an algorithm to predict forthcoming paroxysmal atrial fibrillations. The system consists of several physiological measuring devices, mobile gateways, point-of-care devices, and a monitoring server. The proposed prediction algorithm shows 87.5% accuracy.**

I. INTRODUCTION

According to the World Health Organization (WHO), cardiovascular diseases are the leading cause of death worldwide [1]. Due to the nature of heart beating, continuous monitoring of cardiac patients is useful for diagnosis and management of heart diseases. For such purposes, remote cardiac monitoring systems have been developed and shown to provide a higher yield in identifying cardiac arrhythmia than previous systems [2]. However, most of those systems are only focused on detection of arrhythmias. With the recent advance in sensor and communication technologies, long-term ambulatory monitoring of cardiac patients has become available with little inconvenience to patients [3]. Therefore, studies exploiting such pervasive cardiac monitoring systems to provide services other than detection of arrhythmias are drawing attention.

Our study aims to develop a remote cardiac monitoring system enabling preventive care services. The system consists of several physiological measuring devices, mobile gateways, PoC (Point-of-Care) devices, and a monitoring server. Fig. 1 shows an overview of our system. Physiological measuring devices are connected to the mobile gateways through near-distance wireless communication such as Bluetooth or Zigbee. The system utilizes the smartphones of patients as mobile gateways and the tablet PCs carried by physicians as PoC devices. Mobile gateways and PoC devices communicate with the server through 3G network or Wi-Fi.

In this paper, we propose a system alerting an arrhythmia earlier than its actual onset, to enable both physicians and patients to respond to medical conditions quickly. In the monitoring server, the prediction and detection algorithms of arrhythmias are running on continuously arriving physiological signals. The predicted or detected arrhythmias can be informed to patients or their designated physicians, according to the evaluation result of personalized clinical rules performed by a CDSS (Clinical Decision Support System). To provide enough response time, we suggest an algorithm that predicts the onset of Paroxysmal Atrial Fibrillation [4] 20 minutes earlier than the actual occurrence using data mining techniques. The arrhythmia was chosen as an exemplary case to prove the feasibility of our proposed system because it is one of the most common arrhythmias and early identification of its occurrence has shown benefits to the patients in prompt treatment and prevention [5]. The experimental results demonstrate that our algorithm can predict the arrhythmia with high accuracy.

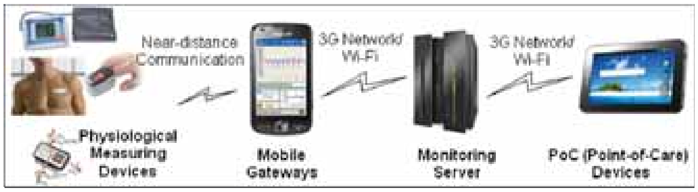

Near-distance Communication
3G Network/ Wi-Fi
3G Network/ Wi-Fi
Physiological Measuring Devices
Mobile Gateways
Monitoring Server
PoC (Point-of-Care) Devices

Fig. 1. An Overview of Our Remote Cardiac Monitoring System

II. SYSTEM DESCRIPTION

Our system is composed of physiological measuring devices, mobile gateways, PoC devices, and a monitoring server. This section describes the software architecture of each component in the proposed system.

A. Physiological Measuring Devices

The patients to be monitored are provided with wearable ECG (Electrocardiography) devices, blood pressure measuring devices, and finger mounted pulse oximeters. The ECG devices with the shape of small flexible patches attached to patients, capture 1-channel ECG signals around the clock and send the measured signals to the patient's mobile gateway. It is also equipped with an additional accelerometer which monitors postural changes of each patient. Blood pressure and SpO_2 can be collected periodically or by a request of the server. For the purpose of interoperability and extensibility of the system, we only use physiological measuring devices complying with ISO/IEEE 11073 medical/health device communication standards [6].

B. Mobile Gateways

Each mobile gateway is a smartphone with Android OS running a dedicated user application to transfer the measured physiological signals from measuring devices to the server. The application has two communication modules running concurrently which are IEEE 11073 protocol stack and HTTP (Hyper-text Transfer Protocol) stack. We designed a JSON-based light-weight message format between the server and the gateway, considering the bandwidth of 3G network and the complexity of XML-based standards such as HL7 CCD (Continuity of Care Document) [7]. The application also contains user interfaces which help patients connect measuring devices to mobile gateways and describe their symptoms which can be sent to the server.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

C. *PoC Devices*

The PoC Devices are tablet PCs with Android OS which are carried by physicians. An application is running on the device to receive alert messages from the server and to notify physicians. It also enables physicians to examine their patients by fetching the measured physiological signals and EMR (Electronic Medical Record) data from the server. After an assessment of the patient's condition, the physician may annotate ECG signals and make a diagnosis. The diagnosis is automatically stored in the server and transferred to the mobile gateway. If an immediate assistance is required, the doctor may communicate directly with the patient or paramedics using the device.

D. *The Monitoring Server*

The monitoring server consists of a CDSS, a DBMS (Database Management System), and a web application service running in a web server. Fig. 2 shows the overall architecture of the monitoring server. In the web application service, the protocol manager parses and generates JSON messages to communicate with mobile gateways and PoC devices. The database manager stores the measured signals received from the mobile gateways in the EMR database and handles the PoC devices' requests for accessing the database. The session manager maintains connection information in order to deliver push messages to mobile gateways and PoC devices.

The signal processing module in the CDSS eliminates the noises and extracts features from the received signals. Our arrhythmia prediction algorithm analyzes the features to predict a forthcoming onset of an arrhythmia and the result is transferred to the workflow execution engine via the temporal abstraction module. Meanwhile, an arrhythmia detection algorithm identifies arrhythmias at the moment to compensate for the occurrences missed by the prediction algorithm. The temporal abstraction module calculates the temporal relations such as overlap or intersection between timestamps or interval values of data, as well as temporal aggregations such as heart rate or the number of the occurrences over a given interval.

The workflow engine executes workflows for each patient in parallel, which contain steps to receive the measurement data and analysis results, to evaluate rules, and to interact with external systems. The rule engine in the workflow engine evaluates the clinical rule set based on detected or predicted arrhythmias, EMR data and personalized parameters of each patient stored in the database. As a result of rule evaluation, the CDSS may perform two actions. The first one is to send a request message to a mobile gateway for the additional measurements such as current symptoms, blood pressure, or SpO_2. The other one is to notify the patient requiring a medical attention to its designated physician.

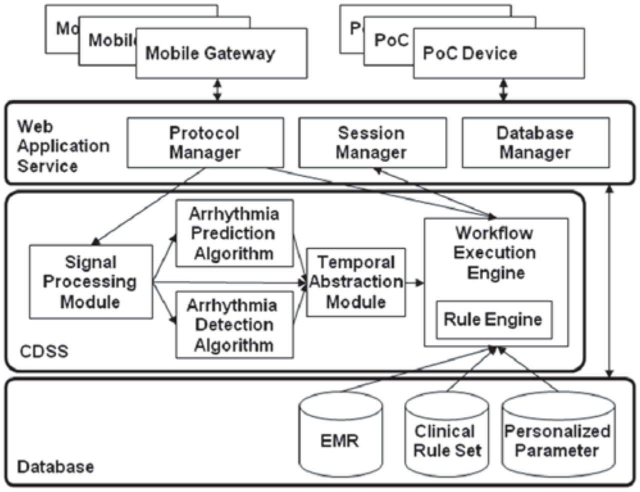


Fig. 2. Overall Architecture of the Monitoring Server

III. The Clinical Decision Support System For Remote Monitoring

A. *Design of Our CDSS*

To cope with the vast amount of information available from the remote monitoring system, medical practitioners need a system to aid their decision-making on the information. Rule-based systems are popular architectures to support decision processes in clinical domain. To implement our CDSS, we chose a workflow system with a rule engine because it allows the users to add new functionalities easily by defining rules or workflows and to represent the flows of the decision processes in understandable graphical forms. The workflow system runs multiple instances of a workflow which emulate the process of reviewing patients' physiological signals and reporting them to the physicians. The rule engine executes production rules, of the form A→C where the antecedent A is the set of conditions and the consequent C is the set of actions to be taken. The clinical rules of remote cardiac monitoring were collected by interviewing expert cardiologists and are inserted to the rule set database using the rule editor of the rule engine.

B. *Supporting Personalized Clinical Decision*

Despite the fact that the clinical rules are based on the experiences of expert doctors, the evaluation of those rules on continuous signals may generate many repeated alerts. To help physicians keep paying attention to the messages generated by the system, we provide personalized clinical decisions. Considering that the medical conditions of patients may vary from person to person as well as change even for the same person as time goes on, we identified the adjustable parameters in the clinical rules so that the physicians can change those parameters remotely by using PoC devices. For example, the number of PVCs (Premature Ventricular Contraction) within a period is a varying parameter of the rule to notify physicians. A patient who just has undergone a cardiac surgery might have many PVCs which are not particularly unusual. But for a patient without any previously diagnosed arrhythmia, a few numbers of PVCs can be a sign of abnormality

IV. Data Mining for Prediction of Paroxysmal Atrial Fibrillation

A. *Early Prediction of Paroxysmal Atrial Fibrillation*

AF (Atrial Fibrillation) is a cardiac state in which the atrium does not beat to a normal rhythm and some parts of the atrium tremble, resulting in fast and irregular heartbeats. AF is one of the most common arrhythmias, accounting for approximately one third of hospitalizations for cardiac rhythm disturbances [8]. PAF (Paroxysmal AF) is a type of AF occurring sporadically and terminating spontaneously. Even though it is usually not considered as an immediate threat to life, it may cause intolerable pains or serious complications such as strokes or heart failures. Therefore, predicting the onset of PAF is clinically important, enabling us to prevent or stabilize various types of arrhythmias using atrial pacing techniques.

To provide reasonable response time to stakeholders, we aim to predict a PAF 20 minutes earlier than its actual onset. Such an early prediction allows physicians to have enough time to examine patients and make proper decisions including anti-arrhythmic drug therapies and immediate electrical cardioversion [5]. Early warning of PAF onsets could also be useful when it is triggered by neurohumoral factors such as exercises, anxiety, or emotional upsets [4].

To the best of our knowledge, there does not exist any algorithm that specifically solves the problem of predicting PAF onsets minutes earlier. Since the Physionet challenge 2001 [9], several approaches have been proposed to address the problem of identifying the characteristics of the ECG segments just before its onsets, which is different from our problem. For example, the best algorithm in the Physionet challenge 2001 [9] is based on the weighted count of APCs (Atrial Premature Contraction) for 10 minutes before the onset, giving higher weight values to APCs close to the onset [10]. Recently, [11] suggested a method of predicting PAF onsets using recurrent plots, which also used the ECG segments immediately preceding the onsets. These existing algorithms simply exploited the abnormalities in the ECG segments immediately before the PAF onsets. On the contrary, predicting onsets far from actual occurrences requires to search for various indicators hidden in ECG signals. Note that this paper focuses on predicting whether a PAF will occur after 20 minutes, not on deciding whether it will occur at this moment.

B. Data Mining Process

In order to capture ECG patterns preceding an actual PAF onset, we need to examine many features that we can extract from long-term ECG signals. We assume that these ECG patterns reflect the progressive development of cardiac abnormalities such as PAF. Fig. 3 shows the ECG data mining process we set in order to build a prediction algorithm using features highly relevant to the PAF onsets.

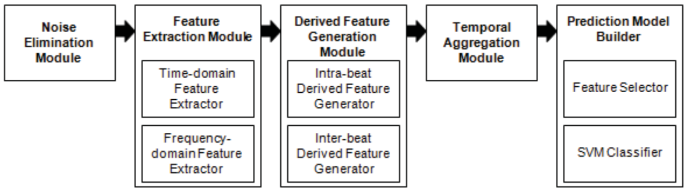

Noise Elimination Module
Feature Extraction Module
Time-domain Feature Extractor
Frequency-domain Feature Extractor
Derived Feature Generation Module
Intra-beat Derived Feature Generator
Inter-beat Derived Feature Generator
Temporal Aggregation Module
Prediction Model Builder
Feature Selector
SVM Classifier

Fig. 3. Process of ECG Data Mining

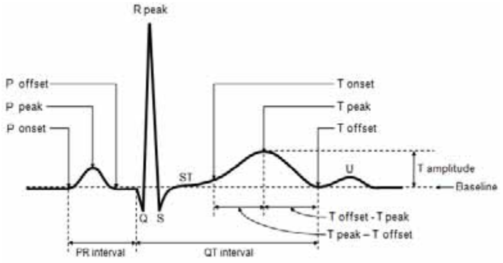


Fig. 4. ECG Morphology Decomposition

We are given two sets of ECG data; one with normal ECG, the other containing ECG signals of PAF patients. After noise elimination, we extract both time-domain features and frequency-domain features, producing a list of features per heartbeat. We then generate new features by combining either multiple features of a heartbeat or features of multiple heartbeats. These features are aggregated into statistical features such as mean or standard deviation of various time intervals as representative values of each time interval per patient. Upon the many features available, we select a set of features with highest information gain and then build multiple SVM (Support Vector Machine) classifiers to find the classifier with best accuracy among them.

C. Feature Extraction

Firstly, we extract time-domain features per heartbeat by performing ECG morphology decomposition in Fig. 4. We decompose a heartbeat into waves, each of which corresponding to a major stage of the cardiac depolarization and polarization mechanism of a heartbeat. The waves are named by Einthoven with the letters such as P, Q, R, S, and T. After de-noising the given ECG signals using band-pass filters, we extract time-domain features of each wave by using a first-derivative based method [12]. In this paper, we use time-domain features such as P onset, P peak, P offset, T onset, T peak, T offset and R peak. In addition, we extract frequency-domain features by applying FFT (Fast Fourier Transform) to each beat of the ECG signals. The frequency range is partitioned into 10 bins and we compute the sum, average, and standard deviation of the coefficients in each bin.

Using the time and frequency domain features, we derive new features, intra-beat and inter-beat features. Examples of intra-beat features include time durations or height differences between two feature points of a heartbeat. Inter-beat features are generated from features of consecutive beats. The R-R interval and the difference of T wave widths are the inter-beat features we mainly use in this paper.

Since we extract features and generate derived features per heartbeat, each feature has values as many as the number of heartbeats in a given ECG signal. In order to represent an ECG signal as a compact feature vector, we need to perform dimensionality reduction. For that, we compute the mean and standard deviation of feature values with varying the size of

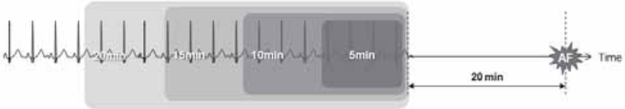


Fig. 5. Various Intervals for Temporal Aggregation

intervals to be 5, 10, 15, and 20 minutes. Since our algorithm aims to predict the onset of PAF 20 minutes ahead, we leverage only the ECG segments 20 minutes distant from the actual PAF onset as shown in Fig. 5.

V. Experimental Results

In this section, we present the experiment results of our algorithm that predicts whether a PAF would occur after 20 minutes. We used the MIT-BIH AF database (AFDB) and MIT-BIH Normal Sinus Rhythm database (NSRDB) [13]. MIT-BIH AF database includes 25 long-term ECG recordings of human subjects with AF, whereas MIT-BIH Normal Sinus Rhythm database contains 18 long-term ECG recordings of subjects with no significant arrhythmias. Since our algorithm needs to use signals from 40 minutes before the onset of PAF, we selected only the recordings with such ECG segments available. In this way, we obtained 14 recordings with PAF and 18 normal recordings.

After extracting features from the given ECG dataset as described in Section IV.C., we select the features highly related to PAF onsets. To do so, we calculated the information gain of each features and selected 6 features with the information gains significantly higher than the rest. The details on the 6 selected features are found in TABLE I. After selecting the features, we built SVM classifiers with all possible subsets of the 6 features and compared their prediction accuracies to find the best SVM model. We measured the classification accuracy by the 4-fold cross validation. The performance of the best SVM classifiers is shown in TABLE II. The highest accuracy we observed is 87.5%.

VI. Conclusion and Future Work

In this paper, we proposed a remote cardiac monitoring system for preventive care, which is composed of measuring devices, mobile gateways, PoC devices and a monitoring server. Mobile gateways transfer physiological signals collected from various measuring devices to the server and PoC devices enable physicians to examine a patient's physiological signals remotely. Meanwhile, a CDSS in the monitoring server analyzes the signals continuously by exploiting arrhythmia detection and prediction algorithms. The CDSS also evaluates clinical rules based on the analysis result and may alert physicians or patients accordingly. To facilitate preventive care, we developed an algorithm that predicts whether a PAF onset would occur after 20 minutes. We achieved the prediction accuracy of 87.5% by fully exploring an ECG segment distant from a PAF onset.

TABLE I. Six Selected Features

	Time Interval	Aggregation Function	ECG Features
f_1	5 mins	Stdev	T width
f_2	10 mins	Stdev	T width
f_3	20 mins	Stdev	T width
f_4	20 mins	Stdev	Difference between T widths of consecutive heartbeats
f_5	20 mins	Avg	4th FFT interval
f_6	10 mins	Avg	5th FFT interval

TABLE II. Classification Performance

Feature Vector	Recall	Accuracy
$<f_4,f_5,f_6>$	71.4%	87.5%
$<f_1,f_2,f_3,f_4,f_5,f_6>$	71.4%	84.4%

In the future, we would perform large-scale validation experiments by using the patient ECG signals that we're currently collecting in hospitals. Then, we plan to apply our data mining framework to predict life-threatening arrhythmias such as Ventricular Tachycardia. In addition to arrhythmia prediction, we plan to perform clinical trials to verify the effect of our system on improving the actual clinical outcomes. We expect this would provide useful insights in developing new preventive methods in cardiology.

References

[1] WHO, *Cardiovascular diseases* [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/en/index.html

[2] P. Zimetbaum and A. Goldman, "Ambulatory Arrhythmia Monitoring: Choosing the Right Device," *Circulation,* vol. 122, pp. 1629-1636, 2010.

[3] B. Lee, S. W. Booh, and K. Shin, "Embedded discrete passives technology for bandage-type medical sensors of E-healthcare system," in *IEEE Elect. Components and Technology Conf.,* 2011, pp. 1325 – 1331.

[4] G.Y.H. Lip and F.L. Li Saw Hee, "Paroxysmal Atrial Fibrillation," *QJM: An International Journal of Medicine*, vol. 94, pp. 665-678, 2001.

[5] G. Engel and R. H. Mead, "Remote Monitoring for Atrial Fibrillation," *Congestive Heart Failure*, Vol. 14, no. s2, pp. 14-18, 2008.

[6] IEEE, IEEE 11073-10406: IEEE Draft Standard for Health informatics - Personal health device communication - Basic Electrocardiograph (ECG) [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5719562

[7] Health Level Seven International, *Continuity of Care Document (CCD)* [Online]. Available: http://wiki.hl7.org/index.php?title=Continuity_of_Care_Document_(CCD)

[8] ACC/AHA/ESC, "2006 Guidelines for the Management of Patients With Atrial Fibrillation," *Circulation*, vol. 114, pp. 700-752, 2006

[9] G. B. Moody, A. L. Goldberger, S. McClennen, and S. P. Swiryn, "Predicting the Onset of Paroxysmal Atrial Fibrillation: The Computers in Cardiology Challenge 2001," in *Computers in Cardiology,* 2001, pp. 113-116.

[10] W. Zong, R. Mukkamala, and R. G. Mark, "A Methodology For Predicting Paroxysmal Atrial Fibrillation Based On ECG Arrhythmia Feature Analysis", in *Computers In Cardiology*, 2001, pp. 125-128.

[11] M. Mohebbi and H. Ghassemian, "Prediction of Paroxysmal Atrial Fibrillation Using Recurrence Plot-based Features of the RR-interval Signal," *Physiological Measurement*, vol. 32, no. 8, pp. 1147, 2011.

[12] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database,*" IEEE Trans. Biomed. Eng.*, vol. 33, pp. 1157-1165, Dec. 1986.

[13] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation,* vol. 101, no. 23, pp. e215-e220, 2000.

Performance Increase by using a EEG Sparse Representation based Classification Method

Younghak Shin, Seungchan Lee, Soogil Woo and Heung-No Lee*
School of Information and Communications
Gwangju Institute of Science and Technology
Gwangju, Republic of Korea
{shinyh, seungchan, woo, heungno}@gist.ac.kr

Abstract— **Attempts are being made to make brain-computer interface system (BCIs) commercially viable for normal person. Stable performance is essential so that BCIs could widely be used for general public. We propose a new classification method based on sparse representation of EEG signals and L1 minimization. The proposed method use the common spatial filtering (CSP) and band power feature for classification. We compare the classification accuracy of proposed method to that of the conventional linear discriminant analysis (LDA) method. Our method shows improved accuracy over the LDA classification method regardless of the number of CSP filters.**

Keywords- Electroencephalogram (EEG), Brain-Computer Interface (BCI), Sparse Representation, Compressed Sensing (CS), Common Spatial Pattern (CSP).

I. INTRODUCTION

Brain-computer interface system (BCIs) provides a new communication and control channel between human brain and an external device without any muscle movements [1]. In the past, BCIs have been developed mostly to provide alternative communication means to people who have severe motor disabilities [2]. These days there are some companies applying electroencephalogram (EEG) based BCIs to normal person by using headset shaped scalp electrodes, such as Emotiv EPOC [3] and MindWave [4]. For these commercial BCIs going beyond laboratory researches, important issue is stable performance, *viz*. classification accuracy.

In this paper we propose a *sparse representation* based classification (SRC) scheme for the purpose of increasing the classification accuracy of EEG based BCIs. This SRC method has been used in the face recognition field [5]. The SRC method works by finding a sparse representation of the test signal in terms of a set of training signals inside a dictionary. This sparse representation is efficiently done by using an L1 minimization which is motivated from the compressive sensing (CS) theory [6]. The dictionary design is the critical step for this method. We use band power as a feature, and common spatial pattern (CSP) filtering for making the EEG signals distinguishable for different classes [7].

II. METHODS

A. Experimental data

In this study, we use a BCI Competition III data set (Data set IVa) [8] which were recorded from five subjects. Subjects have taken the same procedure of a BCI experiment in which there are two classes, Right hand, and Right foot of motor imagery movements. The data recording was made using BrainAmp amplifiers and a 128 channel Ag/AgCl electrode cap from ECI. 118 EEG channels were measured at positions of the extended international 10/20-system. Signals were band-pass filtered between 0.05 and 200Hz and then digitized at 1000Hz. For off line analysis signals were downsampled to 100Hz.

B. Data analysis

We take a data segmentation for following analysis. We use 1000~2000ms of signal samples (100 samples) after the Cue has been presented. Next, to eliminate the noise that is not related with sensorimotor rhythms (SMRs), we use a band-pass filter with 8~15Hz cut off frequency.

To reduce the dimension of feature vector and make distinguishable features, we use the CSP filtering. CSP is a powerful signal processing technique that has been successfully applied for EEG-based BCIs [7].

Let $\mathbf{X} \in \mathbb{R}^{C\times T}$ be a segment of EEG signals where C is the number of EEG channels. In this study, C is 118, and T is the number of sampled time points collected in all the trials. We use 100 samples (one second). We have two classes of EEG training trials $\mathbf{X}_R \in \mathbb{R}^{C\times T}$ and $\mathbf{X}_F \in \mathbb{R}^{C\times T}$ each corresponding to the Right hand '*R*' and Foot '*F*' movement. Using the CSP method, we obtain the CSP filters $\mathbf{W} \in \mathbb{R}^{C\times C}$. We call each column vector $\mathbf{w}_i \in \mathbb{R}^{C} (i = 1, 2, ..., C)$ of $\mathbf{W}$ a spatial filter. Among them, we use n CSP filters from the front and another set from the back. Then, we can make this as the CSP filtering matrix $\overline{\mathbf{W}} \in \mathbb{R}^{C\times 2n}$, i.e., $\overline{\mathbf{W}} := [\mathbf{w}_1, ..., \mathbf{w}_n, \mathbf{w}_{C-n+1}, \mathbf{w}_C]$. Given the two classes of EEG training signals, we define the CSP filtered signals, i.e.,

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$\overline{\mathbf{X}}_R \in \mathbb{R}^{2n\times T} := \overline{\mathbf{W}}^{\mathrm{T}}\mathbf{X}_R$$
$$\overline{\mathbf{X}}_F \in \mathbb{R}^{2n\times T} := \overline{\mathbf{W}}^{\mathrm{T}}\mathbf{X}_F \tag{1}$$

Next, we compute band power of each class signal. In this study, the power of the CSP filtered signal, i.e., the second moment of each row of $\overline{\mathbf{X}}_R$ and $\overline{\mathbf{X}}_F$ is the band power from 8 to 15 Hz.

C. *Linear Sparse Representation Model*

In this section, we aim to introduce the sparse representation of the test signal. Let N_t be the number of total training signals for each class $i = R, F$. We define the dictionary matrix $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, ..., \mathbf{a}_{i,N_t}]$ for $i = R, F$ where each column vector $\mathbf{a} \in \mathbb{R}^{m\times 1}$ having dimension $m = 2n$ is obtained by concatenating the 2*n* band power features. The same procedure is repeated for the right hand and right foot classes. By combining the two matrices, we form the complete dictionary, $\mathbf{A} := [\mathbf{A}_R ; \mathbf{A}_F]$. Thus, the dimension of $\mathbf{A}$ is $m \times 2N_t$.

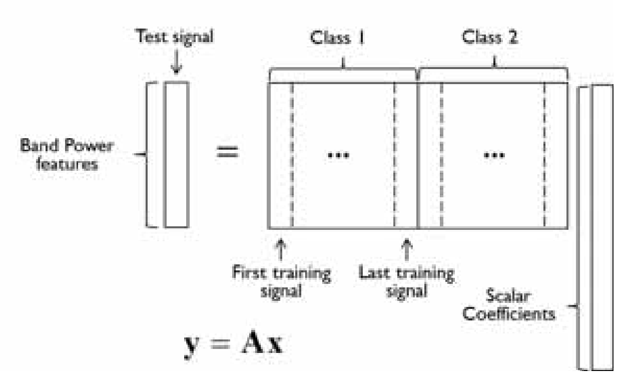


Figure 1. Design a dictionary and linear sparse representation model

Figure 1 shows the proposed model. We apply the same procedure done to obtain the columns of the dictionary to the test signal. Thus, the dimension of **y** is the same as the dimension of the columns of the dictionary **A**. Then, this test signal **y** can be sparsely represented as a linear combination of some columns of **A**:

$$\mathbf{y} = \sum_{i=R,F} x_{i,1}\mathbf{a}_{i,1} + x_{i,2}\mathbf{a}_{i,2} + \cdots + x_{i,n_t}\mathbf{a}_{i,N_t} \tag{2}$$

where $x_{i,j} \in \mathbb{R},\ j = 1, 2, ..., N_t$ are scalar coefficients. Then, we can represent this as a matrix algebraic form:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{3}$$

where $\mathbf{x} = [x_{R,1}, x_{R,2}, ..., x_{R,N_t} x_{F,1}, x_{F,2}, ..., x_{F,N_t}]^T \in \mathbb{R}^{2\cdot N_t}$. For example, we expect that the test signal $\mathbf{y}$ of class R can be represented as the training signals of class R .

$$\mathbf{y}_R = \mathbf{A}\mathbf{x}_R \in \mathbb{R}^{m\times 1} \tag{4}$$

where $\mathbf{x}_R = [\mathbf{a}_{R,1}, \mathbf{a}_{R,2}, ..., \mathbf{a}_{R,N_t}, 0, ..., 0]^T \in \mathbb{R}^{2N_t}$ is a coefficient vector whose elements are zero except some elements associated with test signals of class R . Sparse representation of the test signal **y** can be made when the number of non-zero coefficients of **x** is much smaller than N_t .

D. *Sparse Representation by L1 Minimization*

We have the number of total training signals $2N_t$ which is larger than the number of CSP filters $(m = 2n)$. Thus, the linear equation (4) is under-determined $(m < 2N_t)$. Recent studies in the Compressed Sensing theory have shown that the L1 norm minimization, given below, can solve this under-determined system well in polynomial time [9]:

$$\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x} \tag{5}$$

There are many L1 minimization algorithms. In this paper, we use one of the standard linear programming methods [10], the ‘SolveBP’ function implements the basis pursuit algorithm available in the SparseLab, which is a free MATLAB software package [11].

E. *Sparse Respresentation based Classification*

After solving the L1 minimization problem, the nonzero elements of $\mathbf{x}$ must be corresponding to the column of class i . Because the EEG signals are very noisy and non-stationary, the nonzero elements may appear in the indices corresponding to the column of another class. To make use of the sparse representation result, the coefficient vector **x**, in a classification problem, we introduce the characteristic function δ [5]. For each class i , we define its characteristic function $\delta_i : \mathbb{R}^{2N_t} \to \mathbb{R}^{2N_t}$ which selects the coefficients associated with class i . For $\mathbf{x} \in \mathbb{R}^{2N_t}$, $\delta_i(\mathbf{x}) \in \mathbb{R}^{2N_t}$ is a new vector which is obtained by nulling all the elements of $\mathbf{x}$ that are associated with the other class. Then we can obtain the residuals $r_i(\mathbf{y}) := \|\mathbf{y} - \mathbf{A}\delta_i(\mathbf{x})\|_2$ for *L* and *R*. Then, the classification rule is given by:

$$\text{class}(\mathbf{y})\text{=arg } \min_i r_i(\mathbf{y}) \tag{6}$$

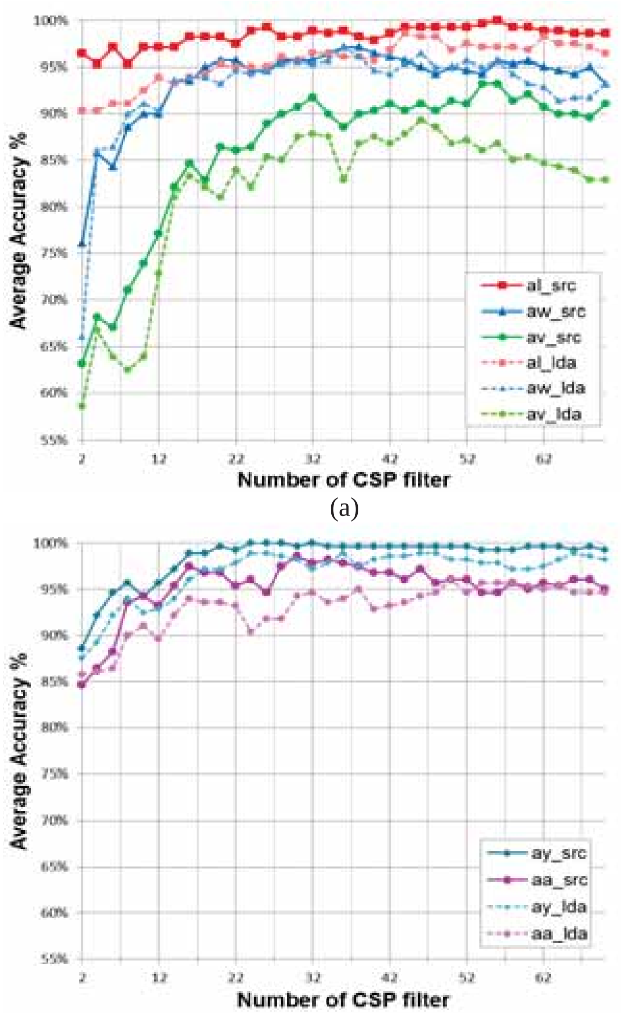
Thus, we determine the class *i* that has the minimum residuals.

III. Results

We have analyzed five data sets using proposed SRC method and conventional linear discriminant analysis (LDA) method. To evaluate the average classification accuracy using limited size datasets, we use the statistical leave-one-out (LOO) cross-validation method with the same total number of data trials for each subject [12]. The classification accuracy is calculated from the following equation:

$$\text{Accuracy}(\%) = \frac{\text{correct test trials}}{\text{total test trials}} \times 100 \qquad (7)$$

Figure 2 shows the classification accuracy (%) of SRC and LDA as a function of the number of CSP filters for each subject. Figure 2 (a) shows the results of subject al, aw and av. Solid line represents the SRC accuracy and dashed line represents the LDA accuracy. Figure 2 (b) shows the results of subject ay and aa. For each selection on the number of CSP filters, SRC performs better than LDA does with few exceptions. Thus, it can be said that SRC has better classification accuracy than LDA regardless of the number of CSP filters in Figure 2. To investigate the statistical significance of the observed accuracies in Figure 10, we performed a paired t-test for each subject. The obtained p-value of the t-test was less than 0.05 for all subjects, which indicates that the difference was significant.



(a)

(b)

Figure 2. Classification accuracy (%) per subject with different number of CSP filters. (a) Classification accuracies for subject al, aw and av. Solid line represents SRC results and dashed line represents LDA results. (b) Classification accuracies for subject ay and aa.

IV. Conclusions

We apply the idea of sparse representation as a new classification method for the motor imagery EEG based BCIs. The sparse representation method needs a well-designed dictionary matrix made of a given set of training data. We use the CSP filtering and the band power to produce the columns of the dictionary matrix. We have shown that a good classification result can be obtained by the proposed method. In addition, we have compared with the conventional approach, *viz.*, the LDA method, which is well known for robust classification performance. Our method shows improved accuracy over the LDA classification method regardless of the number of CSP filters.

V. Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (Do-Yak Research Program, No. 2012-0005656)

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller and T. M. Vaughan, “Brain-computer interfaces for communication and control” *Clin. Neurophysiol.* vol. 113, no. 6, pp. 767-791, 2002.

[2] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, “Brain-computer interface-a new communication device for handicapped persons,” *J. Microcomput. Appl.,* vol. 16, pp. 293-299, 1993.

[3] Emotiv Lifesciences, http://www.emotivlifesciences.com/

[4] NeuroSky, http://www.neurosky.com/

[5] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, Yi Ma, “Robust Face Recognition via Sparse Representation” IEEE Trans. *Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210~227, February 2009.

[6] D. Donoho, “Compressed sensing,” *IEEE Trans. Information Theory*, vol. 52, pp. 1289-1306, 2006.

[7] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, Klaus-Robert Müller. “Optimizing Spatial Filters for Robust EEG Single-Trial Analysis,” *IEEE Signal Proc. Magazine*, 25(1):41-56, January 2008.

[8] Benjamin Blankertz, Berlin Brain-Computer Interface, http://www.bbci.de/

[9] E. Cande`s, J. Romberg, and T. Tao, “Stable Signal Recovery from Incomplete and Inaccurate Measurements,” *Comm. Pure and Applied Math.*, vol. 59, no. 8, pp. 1207-1223, 2006.

[10] S. Chen, D. Donoho, and M. Saunders, “Atomic Decomposition by Basis Pursuit,” *SIAM Rev*., vol. 43, no. 1, pp. 129-159, 2001.

[11] David Donoho, Victoria Stodden and Ysskov Tsaig, SparseLab, http://sparselab.stanford.edu/

[12] Larry Wasserman, “All of Statistics: A Concise Course in Statistical Inference”, Springer, 2010, pp. 63-64.

Compact Wireless EEG System with Active Electrodes for Daily Healthcare Monitoring

Koji Morikawa*, Akinori Matsumoto*, Shrishail Patki†, Bernard Grundlehner†,
Auryn Verwegen†, Jiawei Xu†, Srinjoy Mitra‡ and Julien Penders†
*Panasonic Corporation, Japan
Email: morikawa.koji@jp.panasonic.com
†Holst Centre/imec-nl, The Netherlands
‡imec, Leuven, Beigium

***Abstract*—Development of Wireless EEG system is described. Realtime impedance monitoring and active electrodes are introduced in order to reduce noise from impedance changes caused due to body motion, and to prevent noise from power line interference, respectively. EEG ASICs are developed for the system. The complete system has a low noise (60nV/√Hz) and is packaged in a compact enclosure (38mm x 38mm x 16mm). The system is evaluated against different types of artefacts and possible applications with the system are discussed.**

I. INTRODUCTION

In hospitals, bio-potential signals such as Electroencephalogram (EEG) and Electrocardiogram (ECG) are measured as indicators for medical diagnosis. Continuous monitoring of bio-potential signals helps to detect the indications of medical disorders earlier. This requires bringing the bio-potential signals monitoring to the home environment while maintaining the high signal quality offered by conventional hospital based EEG systems. One of the main reasons for medical grade signal quality in the hospitals is use of skin preparation and gel electrodes. These practices reduce the skin-electrode contact impedance improving the signal quality of the bio-potential signals. However, this restricts the use of the conventional bio-potential monitoring systems to the ambulatory environment. Daily healthcare monitoring requires miniaturized and low noise wearable sensors. Another important aspect of such wearable sensors is the need for quick and easy set-up without the medical technician intervention. Dry electrodes prove beneficial due to their quick setup but their high impedance makes them more susceptible to artefacts. Solutions to reduce different types of artefacts are required on the system level as well as on the application level.

This paper presents a miniaturized and low noise wireless EEG system with active electrodes in order to address the issue of reducing different types of artefacts. The active electrodes play an important role in reducing the impedance especially in case of dry electrodes. The literature also suggests that electrode tissue impedance (ETI) can be used to provide information regarding motion artefacts. The readout ASIC in the system enables simultaneous recording of EEG and ETI. The system is evaluated for key performance parameters in order to verify the susceptibility against specific types of artefacts. The system with an extended dymanic range can also be used to monitor other bio-potential signals such as ECG or EOG enabling various applications ranging from medical domain to the lifestyle and gaming domain. Finally some possible applications with the system are also discussed.

II. SYSTEM REQUIREMENTS

The table 1 summarizes various artefacts that impose serious challenges during EEG monitoring in ambulatory conditions. Possible solutions are also summarized in the table 1. These artefacts can be divided into physiological artefacts coming from internal body such as other bio-potential signals, motion artefacts caused due to movement of electrodes or electrode lead wires or external artefacts such as power line interference.

Active electrodes reduce the impedance of electrode lead wire making the wire less susceptible to power line interference. Real-time ETI monitoring with EEG provides an additional parameter to detect the motion artefacts caused due to the electrode movement. Integration of active electrodes with the EEG and ETI monitoring provides additional information which can be used for detecting and reducing the artefacts mentioned in (d), (e) and (f).

Background EEG and signal from muscle or eye movements are inevitably mixed to the target signal. Physiological artefacts (mentioned as (a)-(c) in table 1) removal needs efficient signal processing algorithms.

Typically, eight channels of bio-potential signals are necessary to enable sleep monitoring at home. Standard sleep monitoring [1] requires two to four channels for EEG, two channels for eye movement and one to two for chin movement.

In summary, compact, eight-channel bio-potential sensor that can work without the use of gel electrodes is necessary for enabling healthcare monitoring at home.

Table 1 Source of Artefacts

	Place	Signal Source	Solution
Target Signal	Internal Brain	Brain (EEG)	
Artefacts		(a) Background EEG	Signal Processing
	Internal Body	(b) Muscle movement (c) Eye movement	Signal Processing (Signal separation/Extraction)
	Boundary	(d) Electrode-tissue Contact Impedance	Signal correction by Realtime impedance monitoring
	External Body	(e) Power line Inference (f) Cable movement	Lower input impedance by Active Electrodes

III. DEVELOPED SYSTEM

A. Systems overview

Based on the system requirements discussed above, active

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

electrodes and EEG analog front end ASIC were developed [2]. The complete system consists of 9 active electrodes and a back-end analog signal processor. It is capable of continuously recording EEG signals and electrode-tissue contact impedance (ETI). The EEG channels have 1.2GΩ input impedance, 1.75µVrms noise (0.5-100Hz), 84dB CMRR, and can reject ±250mV of electrode offset.

The integrated wireless sensor module consists of active electrode ASICs, EEG analog front end (EEG AFE) ASIC, microcontroller, radio and power management circuit. The active electrode is a low power buffer (30µW) with very high input impedance and built-in current generator for ETI measurement. The output of the active electrode is demodulated into the EEG and ETI. The EEG AFE has 8 readout channels and has a built-in 12 bit SAR ADC to digitize the EEG, ETI-I and ETI-Q information. Each active electrode is connected with AFE using 6 wires. The analog and digital modules are separated in order to reduce the interference from the radio on the AFE. An accelerometer is also provided on the system in order to track the movement.

B. Evaluation

Benefit of active electrode is demonstrated in terms of its susceptibility against power line interference (e) and cable motion artefacts (f). For quantifying the susceptibility against power line interference, the sensor module was isolated in an EMI shielded box and subjected to a controllable amount of 50Hz noise. The power density levels of 50Hz were measured by introducing a resistor at the input. The experiment was performed with and without active electrodes. The noise in the case where active electrodes were used remains consistent despite the increase in resistance which suggests better susceptibility against power line interference [3].

Cable motion artefact is also evaluated by vibrating an electrode wire at a known frequency and amplitude. A cable of length 80cm was used in order to simulate the electrode lead wire. The vibrating cable was mechanically decoupled by fixating it with a tape to a non-vibrating surface. The power spectral density at the vibration frequency was recorded. The measurements were performed with and without active electrode. The noise level with active electrode is lower than the case where passive electrodes were used [3].

IV. APPLICATIONS

A. Prototype

In order to have a wearable bio-potential sensor, the system is integrated and packaged into a compact enclosure . Figure 1 shows the picture of the prototype. The dimensions of the prorotype are 38mm x 38mm x 16mm and it weighs 23.5g including battery and electrodes for reference and one measurement channel. The default gain is 1200 and dynamic range of measurement is 1.5mV peak to peak. The wider dynamic range allows ECG signals to be measured with EEG simultaneously.

The Ag/AgCl electrodes are installed since they have low half-cell potential.. The material of electrodes affects the signal settling time. The settling time is an important parameter especially in cases where frequent measurements are necessary, such as an instant health checker.

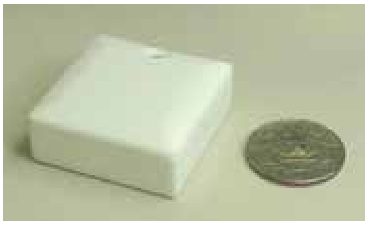


Fig.1. Sensor module prototype

B. Examples of monitoring applications

The miniaturized and low noise wireless system proposed in this paper has the potential to be used in various applications. In brain machine interfacing (BMI) [4], one of the barriers for the prevalence of BMI systems is the scarcity of compact, reliable EEG system. In sleep monitoring [1], multi bio-potentials from brain, eyes, and chin need to be monitored in order to determine the sleep stages. A wireless system proposed in this paper has the potential to enhance the quality of the sleep monitoring by providing more freedom of movement to the subjects.

In hearing-aid fitting [5], EEG can be used for visualizing the hearing abilities. The compact and reliable system increases the possibility of usage at hearing-aid stores. Driver monitoring is another new field of bio-potential application. EEG is good index for sleepiness, drowsiness and attention distraction while driving.

V. CONCLUSIONS

Compact wireless monitoring system with active electrodes is developed which can enable daily healthcare monitoring. Active electrodes and ETI monitoring improves the signal quality and allows the use of system with dry electrodes avoiding the use of gel electrodes.

Future work includes long term bio-potential recordings, algorithm development based on additional signals from the system and mining useful information.

REFERENCES

[1] J. Fell, J. Roschke, K, Mann, C. Schaffner, “Discrimination of sleep stages: a comparison between spectral and nonlinear EEG measures”, Electroencephalography and clinical Neurophysiology, Vol. 98, pp. 401-410, 1996.

[2] S. Mitra, J. Xu, A. Matsumoto, K. A. A. Makinwa, C. Van Hoof, R. F. Yazicioglu, “A 700µW Active-Electrode Based Eight-Channel EEG Acquisition System”, Digest of 2012 Symposium on VLSI Circuits, pp.68-69, 2012.

[3] S. Patki, B. Grundlehner, A. Verwegen, J. Xu, J. Penders, S. Mitra, A. Matsumoto, “Wire.ess EEG system with Real Time Impedance Monitoring and Active Electrodes”, BioCAS 2012, Submitted

[4] J.R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T.M. Vaughan, “Brain-Computer interfaces for communication and control”, Clinical Neurophysiology, Vol. 113, pp. 767-791, 2002

[5] S. Adachi, K. Morikawa, Y. O. Kato, J. Ozawa, H. Nittono, “Estimating Uncomfortable Loudness Levels using Evoked Potentials to Auditory Stimuli for Hearing Aid Fitting”, 34th Annual Intl. Conf. on the IEEE EMBS, pp. 2108-2111, 2012

Progressive Monitoring and Treatment Planning of Diabetes Mellitus in Smart Home Environment

Topi Pulkkinen[1], *Member, IEEE*, Young-Sung Son[2], Joohyung Lee[3], Yann-Hang Lee[3], Mikko Sallinen[1] and Jun-Hee Park[2]

[1]VTT Technical Research Centre of Finland, [2]ETRI, [3]Arizona State University

***Abstract*--This paper describes a method and a platform for generating a treatment plan for a patient with a chronic disease such as diabetes. The planning process considers the user's living environment so the doctor's prescription can be easily implemented as a monitored activity. The process utilizes OWL representation for the knowledge modeling and Answer Set Programming for managing the patient's goals.**

I. INTRODUCTION

Currently healthcare related costs are increasing especially in case of chronic diseases such as diabetes, which encourages improving homecare. Because the real progress monitoring is lacking it is difficult for the doctor to see the symptoms between the check-ups of a homecare patient. Additionally the self-monitoring equipment can be faulty or the patient can make systematic mistakes by measuring the body-signals incorrectly.

It can be envisioned that the home network gateway or a cloud server (like Continua alliance architecture suggests) could be acting as shared data storage between the patient and the doctor. The doctor can set-up the treatment plans, follow the progress of the disease and the patient would benefit by getting helpful notifications from the system automatically. Additionally by monitoring the user's activities for a long period of time it is possible to distinguish user's life patterns, which in general can help to improve and to optimize the process to reach the intended goal as well as verify the correctness of the data [1,2].

The paper is organized as follows: In chapter 2 we describe the active monitoring architecture for diabetes case, which is the basis for the treatment plan generation. Chapter 3 defines the process and the data models that are used for generating the treatment goals. Finally Chapter 4 draws the conclusions.

II. ACTIVE MONITORING

When the patient has been diagnosed as a diabetic the general goal is to prevent diabetes complications and guarantee a good standard of living without suffering from diabetes symptoms. If the diabetes is very serious a patient should utilize self-measurement devices, e.g. blood glucose meter and blood pressure meter. He should also keep food diary and exercise diary and monitor his body weight.

A. Test case for active monitoring

To analyze a life-pattern of a diabetic person various home measurements has to be gathered. Table 1 illustrates example measurements that were utilized in a test case for diabetes monitoring at home.

TABLE I
EXAMPLE DATA SOURCES FOR DIABETES MONITORING

Variable	Sensor or data source
Activity (four states: walk, run, stable, fall)	Smartphone context recognition software
Location cluster	Clustered GPS data from Smartphone
Weight, Body fat, Body water	Body scale at home
Blood glucose, Blood pressure, heart rate	Multi-function stationary bio-sensor at home
Door sensor, noise, temperature, luminance, humidity, CO_2	Home network sensors
Patient status in kitchen (eat, sit, fall down)	Depth-camera at home
User's calendar, food diary, exercise diary, local weather	Web service
Daily status (overall condition, detailed condition, symptom, sleep time and type, stress level, pain level)	Web service
User's biological profile (age, sex, height)	Web service
User's medical profile (diabetes type, other diseases, prescribed medicine	Web service
User's sociological profile (hasCar, hasBicycle, home location, workplace location, family members)	Web service

All the measurements are produced as time series data that can be abstracted into more suitable format for the treatment plan generation. For example a doctor doesn't need know that the patient is walking every morning, but instead a doctor is interested to monitor the exercise level during six months or the correlation in between exercise and blood sugar.

To abstract the time series data into concepts we utilize an ontology description of the home domain and health domain. Home domain ontology describes the devices and the properties at home while the health domain has semantics for patient's health status. These are described in more detail in chapter 3.

B. Active Monitoring Platform

The home devices are connected in various ways: some devices provide stand-alone cloud service interface such as wi-fi body scale and others are connected via home network.

The deployment diagram presented in Fig. 2 describes the architecture of the data management process within the patient's home. The server where the data is collected is in the middle of the Fig. 2. The home gateway on top left runs visualization module and reasoner. The data processing and data mining has its own server where the user can also give input for supervised learning purposes.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

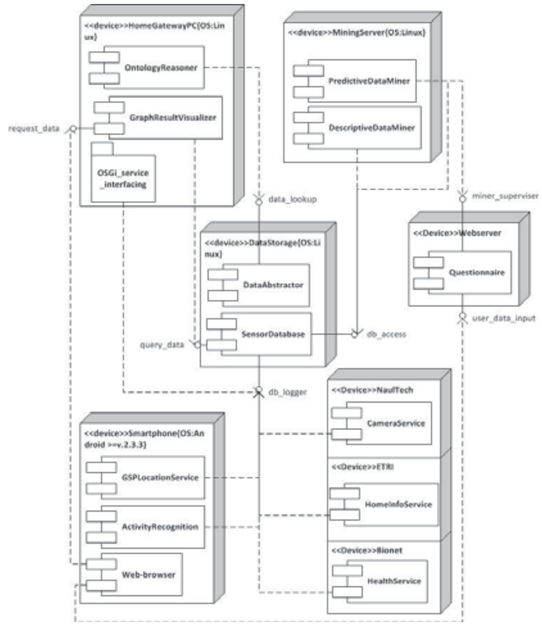


Fig. 2. Deployment diagram of the active monitoring platform.

III. KNOWLEDGE MODELING

The system knowledge which the medical professional uses for creating the treatment plans is stored in OWL format. There are two main domains that contribute to the treatment process (i) smart home domain and (ii) health domain. Additionally the treatment planning should be able to utilize common medical knowledge information such as UMLS (Unified Medical Language System) for supporting the doctor in case of a patient with multiple diagnosed diseases that can affect the treatment, medication and lifestyle of the patient.

A. *Health domain ontology*

The devices that are modeled in the smart home domain ontology provide monitoring function, which is mapped to the health domain ontology by HealthMonitor class. This class contains aggregated or abstracted information from various sensors that could be utilized as the system knowledge. The ontology model for health domain is presented in Fig. 3.

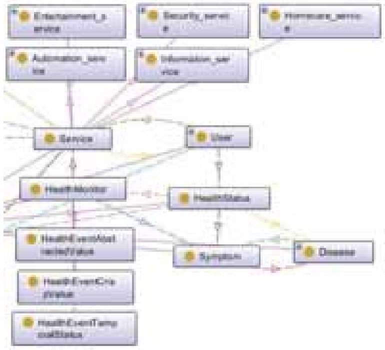


Fig. 3. Health domain ontology classes.

The HealthMonitor class contains three HealthEvent subclasses that are used depending on the data: for example the instance HealthEventTemporalStatus describes a long-term temporal state of health e.g. fever with start and end times, while the HealthEventCrispValue can describe a medical variable such as SMBG (Self Monitoring Blood Glucose).

B. *Treatment plan generation*

The treatment plan is based on a set of rules and constraints that a doctor creates for a patient. Fig. 4. presents the process that the doctor uses when the treatment plan is generated. First the planner visualizes the possible home monitors that the patient has at home. For example, if the user has a body scale and a food diary the doctor could select a goal of 5% weight loss and some constraints like daily energy intake over 1400 kCal. The created rule is translated to Answer Set Programming (ASP) syntax, which has be proven to be effective for solving multivariate logic problems [3].

If there are some important goals the doctor wants to set for the patient but a monitor does not exist at patient's home, the doctor can make a new monitor that needs to be installed by the patient.

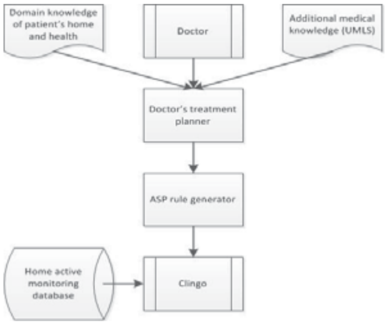

Domain knowledge of patient's home and health
Doctor
Additional medical knowledge (UMLS)
Doctor's treatment planner
ASP rule generator
Home active monitoring database
Clingo

Fig. 4. General process for treatment plan generation.

IV. CONCLUSIONS

This paper describes a method and a platform for generating a treatment plan for a patient with a chronic disease. The medical professional can utilize a simple gui on a tablet computer to create treatment goals for the user. The goals are constructed based on the patient's actual living environment modeled with OWL. As a result it is possible for the doctor and the patient to see the progress of the monitored disease.

REFERENCES

[1] Y. Son, T. Pulkkinen, K. Moon, and C. Kim, "Home energy management system based on Power Line Communication," IEEE Trans. Consumer Electron., Vol. 56, No. 3, pp. 1380-1386, Aug. 2010.

[2] T. Pulkkinen, M. Sallinen, J. Son, J-H. Park, Y-H. Lee, "Home Network Semantic Modeling and Reasoning - A Case Study," Proceedings of the 15th International Conference on Information Fusion in Singapore 10-12th July 2012, pp. 338-345.

[3] L. Chen, C.D. Nugent and H. Wang, "A Knowledge-Driven Approach to Activity Recognition in Smart Homes," IEEE Trans. on Knowledge and Data Engineering, Vol. 24, No. 6, pp. 961-974, June 2012.

User Adaptive Application Program Management among Multi-devices for Personal Cloud Computing Services

Hyewon SONG, Eunjeong CHOI, Chang Seok BAE, and Jeunwoo LEE, *Member, IEEE*

Abstract**-- This paper presents a User Adaptive Application Program Management (UAAM) framework, which supports a self-management rule based on usage state pattern of applications and devices, in the Personal Cloud Computing Service Environment. It provides to adaptively control and manage applications among devices like a smartphone and tablet PC according to usage properties of a user which possesses the devices. Consequentially, it enables for the users to conveniently keep the consistency of application programs among devices with personal rules customized by own usage properties.**

I. INTRODUCTION

As a user has one and more mobile devices such as smart phones and tablet PCs, they need to use and manage their devices more easily. To address the requirements, some researchers focus on a cloud computing service technology personally provided by various methods. [1] As one of methods, we focus on maintaining a consistent user experience (UX) in an aspect of applications and their execution among multi-devices, and adaptively controlling the process based on application usage properties in devices belonging to a user in the Personal Cloud Computing Service environment [2]. In connection with this issue, there are various researches to study the data sharing using synchronization method and the context-aware methods for supporting the data sharing, which is adaptive to diverse personal preference and status information of service users, among multi-devices. [3]-[6] However, they focus on the data synchronization to provide the consistency of data to users whenever the users access any data in their devices, so that, they do not consider the case of application programs. Therefore, we propose a User Adaptive Application Program Management (UAAM) framework in this paper. It makes it possible to adaptively manage application programs according to user’s state information such as usage pattern of applications or devices. At first, we propose the architecture for the UAAM framework and the basic process for managing various application programs installed in multi-devices based on management rules, which is adaptive to the usage state pattern of users. Finally, we show a use case of the UAAM framework using Android mobile devices and Personal Cloud Computing (PCC) Server developed in the Personal Cloud Computing Project [2].

II. USER ADAPTIVE APPLICATION PROGRAM MANAGEMENT USING USAGE STATE PATTERNS

As mentioned in a previous section, the UAAM framework is an application program management framework in multi-devices environments for Personal Cloud Computing. In addition, the UAAM framework is a user-convenient and device-adaptive framework for executing and managing application in diverse devices as well as accessing data given by various contents of devices. It mainly provides application program management among devices belonging to a specific user using usage state value given by user and device usage pattern model. In order to support this framework, we consider agent and web based server architecture in which it enables to model the usage state pattern per devices of a specific user and to control a management process of applications and their execution among devices.

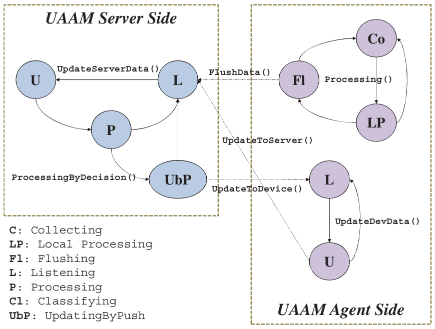


Fig. 1. UAAM Functional Architecture with Function State Flows

As shown in Fig. 1, the UAAM framework is consists of 2 basic components: a UAAM Server using PCC resources, PCC server and PCC storages, and a UAAM agent installed in users devices. In the agent side, there are 2 kinds of closed flows: 1) updating an application status, which is similar to an application sync process in [7], when it is changed, and 2) collecting and sending information, which is either raw data or processed data from the device. The UAAM server processes main functions to support efficiently and automatically managing applications among devices per a user. The basic flow in the server is also similar to an application management process in [7], however, the UAAM server has a function to select a management rule adaptively.

For this, we propose a simple decision algorithm using the usage state and its pattern of applications. In order to appropriately quantify the usage state of an application, we consider an RFT (Recency, Frequency, and Term) model [8]. Namely, the usage state value of applications is defined as a rate value to represent how often an application is executed during a time interval, how recently an application is executed, and how long an application is executed. Based on this model, we let r_i^k , f_i^k and d_i^k be the calibrated value to reflect recency, frequency, and term (duration) for the *k*th application installed in the *i*th device. Also, we let *L* be the number of total devices belonging to a user, *M* is the number of total users to consume the PCC service with the UAAM framework,

This work was supported by the IT R&D program of MKE/KEIT. [K10035321, Terminal Independent Personal Cloud System].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

and N is the number of total application programs installed in a device. When the time interval for a device i, which the jth user possesses, is T_i, we first define the usage state value with respect to recency (r_i^k), frequency (f_i^k), and term (d_i^k) as follows:

$$u_{r_i^k} = \frac{r_i^k - r_{max}}{r_i^k}, \quad u_{f_i^k} = \frac{f_i^k - f_{max}}{f_i^k}, \quad u_{d_i^k} = \frac{d_i^k - d_{max}}{d_i^k} \qquad (1)$$

where $i = 1, 2, \ldots, L$ and $k = 1, 2, \ldots, N$. r_i^k is the moment value to least recently execute the application k in the device i during T_i, and r_{max} is the maximum of r_i^k. Also, f_i^k is the rate value to execute the application k in the device i during T_i, and f_{max} is the maximum of f_i^k. In addition, d_i^k is the duration value to execute the application k in the device i during T_i, and d_{max} is the maximum of d_i^k. Finally, the usage state value of an application k installed in the device i, u_i^k, is defined as follows:

$$u_i^k = \omega_r u_{r_i^k} + \omega_f u_{f_i^k} + \omega_d u_{d_i^k} \qquad (2)$$

Where ω_r, ω_f and ω_d are the weighted value based on their relative importance. Using this value, we derive the usage pattern of applications for the jth user as follows:

$$\hat{p}_j = \begin{bmatrix} \left(u_1^1,\ u_1^2,\ \ldots,\ u_1^N\right) \\ \vdots \\ \left(u_L^1,\ u_L^2,\ \ldots,\ u_L^N\right) \end{bmatrix} \qquad (3)$$

where $j = 1, 2, \ldots, M$. This usage pattern represents the usage state of total applications installed in all devices of the jth user. Meanwhile, using this pattern, we can derive the usage state value of the ith device (u_i) as follows:

$$u_i = \frac{1}{N} \sum_{k=1,2,\ldots,N} u_i^k \qquad (4)$$

where $i = 1, 2, \ldots, L$. This value is applied to making a decision in a management process. The simple process is as shown in Fig. 2.

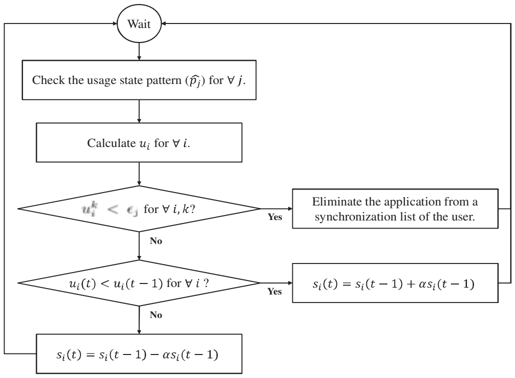

Wait
Check the usage state pattern ($\hat{p}_j$) for ∀ j.
Calculate u_i for ∀ i.
$u_i^k < \epsilon_j$ for ∀ i, k?
Yes
Eliminate the application from a synchronization list of the user.
No
$u_i(t) < u_i(t-1)$ for ∀ i ?
Yes
$s_i(t) = s_i(t-1) + \alpha s_i(t-1)$
No
$s_i(t) = s_i(t-1) - \alpha s_i(t-1)$

Fig. 2. Self-Management Process in the UAAM framework

In a case of ε_j, it should represent the enough value to remove an application which is seldom executed by a user. Also, if the application deleted from a list is run by a user again, it can be added in the list.

III. IMPLEMENTATION AND RESULTS

For developing the UAAM framework, we use not only some devices with Android platform such as smartphones and tablet PCs but also a web server connected with the PCC server and PCC storage service, which has already implemented as a result from PCC project [2]. Fig. 3 shows the scenario for a use case, which we design as a reference model of the proposed framework, and we can administrate the management process based on the usage pattern for executing the basic application program synchronization service.

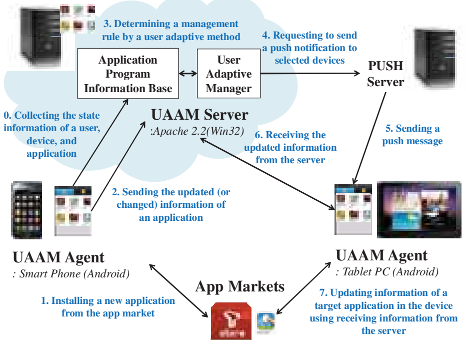

3. Determining a management rule by a user adaptive method
Application Program Information Base
User Adaptive Manager
4. Requesting to send a push notification to selected devices
PUSH Server
0. Collecting the state information of a user, device, and application
UAAM Server
:Apache 2.2(Win32)
6. Receiving the updated information from the server
5. Sending a push message
2. Sending the updated (or changed) information of an application
UAAM Agent
: Smart Phone (Android)
UAAM Agent
: Tablet PC (Android)
App Markets
1. Installing a new application from the app market
7. Updating information of a target application in the device using receiving information from the server

Fig. 3. UAAM Use Case with a Scenario

IV. CONCLUSION

In this paper, we propose application program management framework with a user adaptive method, the UAAM framework. This concept is based on a management method proposed in [7], however, the UAAM framework provides automatic adaptive control for application program management among devices, so that, it adaptively reflects properties of a user which possesses multi-devices like a smartphone, and tablet PC. Namely, the UAAM framework facilitates to adaptively manage application programs over multi-devices throughout the management rule controlled by the usability of applications and devices.

REFERENCE

[1] L. Ardissono, A. Goy, G. Petrone, and M. Segnan, "From Service Clouds to User-centric Personal Clouds," *Proceedings of IEEE International Conference on Cloud Computing*, pp. 1-8 (2009)

[2] Personal Cloud Computing Project, http://pcc.sktelecom.com

[3] T. Lindholm, J. Kangasharju, and S. "Tarkoma: Syxaw, Data Synchronization Middleware for the Mobile Web," *Journal of Mobile Networks and Applications*, Vol. 14, No. 5, pp. 661-676 (2009)

[4] V. Ramasubramanian, K. Veeraraghavan, K. P. N. Puttaswamy, T. L. Rodeheffer, D. B. Terry, and T. Wobber, "Fidelity-Aware Replication for Mobile Devices," *IEEE Trans. on Mobile Computing*, Vol. 9, Issue 12, pp. 1697-1712 (2010)

[5] iCloud, http://www.apple.com/icloud/

[6] P. Stuedi, I. Mohomed, and D. Terry, "WhereStore: Location-based Data Storage for Mobile Devices Interacting with the Cloud," *Proceedings of the 1st ACM Workshop on Mobile Cloud Computing and Service (MCS)*, pp. 1-8 (2010)

[7] H. Song, E. Choi, C. S. Bae, and J. W. Lee, "Web based Application Program Management Framework in Multi-Device Environments for Personal Cloud Computing," *IT Convergence and Services, Lecture Notes in Electrical Engineering*, Vol. 107, pp. 529-536 (2011)

[8] B. Yan and G. Chen, "AppJoy: Personalized Mobile Application Discovery,*" Proceedings of the 9th international conference on Mobile systems, applications, and services (MobiSys)*, pp. 113-126 (2011)

Efficient Seamless Content Sharing Among Separate Multiple WLANs for Pervasive Mobile Network Environment

Hiroyuki Kasai, The University of Electro-Communications, JAPAN

Abstract—**This paper presents a proposal of a scheme to collect and manage content and service data belonging to separate multiple WLANs so that Mobile CE devices can use them. Content and service list sharing for seamless multiple WLANs access by network switching is proposed. This paper shows the effectiveness of the proposed algorithm using simulations, and describes future work[1].**

I. Introduction

Recently, various wireless access modes such as WiFi have become widely spread throughout our society in places such as transport stations, shopping malls, and streets. In this environment, an important necessity is that the end-user wants to use contents and services seamlessly even though they are located separately across multiple WLANs. These contents and services might include AV/image media files, and content provisioning services inside users' mobile terminals or provider servers. Managing the contents and services on the internet is expected to entail higher costs because mobile terminals are movable and unstable. This paper presents a proposal for an innovative mechanism that collects and shares content and service lists belonging to multiple WLANs so that the end-user can use them. Without modifying WLAN access points, end terminals collect the content and service lists on multiple WLANs autonomously by switching themselves collaboratively. Then the terminals share contents with other terminals, providing seamless accessibility to enable construction of, for instance, a P2P network across separated multiple WLANs.

II. Proposed Method

In our proposal, content sharing is achieved between a pair of WLANs (Fig. 1). As Fig. 1 depicts, WLAN1 and WLAN2 exist, with one "Switching Terminal" belonging to WLAN1 switching to WLAN2 (ii) after broadcasting a switching announcement (i), and exchanging the content and service lists (not content itself) with WLAN2 (iii). Then, it returns to WLAN1 (iv) and distributes the new list to others (v). Each terminal quickly shares it with others in the visiting WLAN using MAC messages without IP address allocation. Some conventional studies, especially in ad-hoc network research, have been conducted for service discovery or management [1][2]. However, those targets are not located in seamless access to multiple WLANs for contents and sharing.

Because network switching might disturb existing communication sessions, we propose an autonomous determination algorithm for switching timings based on terminal states to reduce interference. Also, the sharing delay time for the change of content lists to reach all terminals in all WLANs should be reduced as much as possible. Therefore, a cost-efficient and fast sharing mechanism is proposed, designated as an "Event-Driven Information Sharing Scheme."

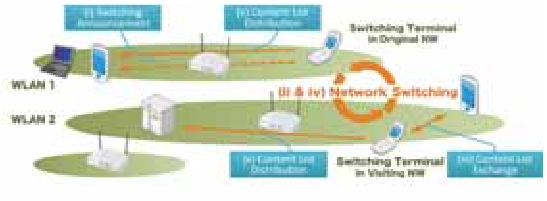

(i) Switching Announcement
(v) Content List Distribution
Switching Terminal in Original NW
WLAN 1
(ii & iv) Network Switching
WLAN 2
(v) Content List Distribution
Switching Terminal in Visiting NW
(iii) Content List Exchange

Fig. 1. Basic Operation of Content List Sharing.

A. Autonomous Switching Determination inside WLAN

Using an autonomous switching time determination algorithm without a central control point, the session interruption can be reduced as low as possible. Our proposed mechanism determines a switching absolute time, t_{switch}, by predicting the terminal's future communication session status. Specifically, each terminal observes its communication sessions during "Observation Time Period, T_{obsv}", and determines t_{switch} for switching during "Switching Time Period, T_{sw}." It estimates a switchable ratio R during the prior T_{obsv} by adding the time lengths that are practically used for communication sessions such as web browsing or IPTV. These are calculable from the start absolute time, t_{bi} $(i=0,\ldots)$ and the end absolute time t_{ei} $(i=0,\ldots)$ of the i-th communication session. If the required time period to switch between two WLANs is defined as TP_s, then the switchable ratio R is calculated using $R=\Sigma(t_{ei} - t_{bi} + TP_s)/ TP_{obsv.}$ Consequently, we can calculate $t_{switch} = R$ x $TP_{sw}+ t_c$, where t_c is the current absolute time.

B. Time-Slot Based Event-Driven Information Sharing

We propose a time slot based event driven content sharing scheme for all WLANs to share content information. Here, "event" means joining of new contents and services or changes to existing content lists. In each time slot, a pair of two WLANs shares its lists by switching one terminal belonging to each. This operation is performed recursively in multiple WLANs. Regarding time slot synchronization, its start time can be chosen by discovering the AP with the smallest MAC address automatically. This paper presents a proposal of four methods. The "*Base Method*" in Fig. 2(i) is that a terminal inside the WLAN switches to the next WLAN every time an event occurs.

[1] This work is supported in part by the National Institute of Information and Communications Technology (NICT), Japan, under the new generation network R&D program for innovative network virtualization platform and its application.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

The following networks recursively switch until the event reaches all WLANs. In the "*Aggregation Method*" in (ii), an event is stored without starting sharing. The sharing operation starts only at the assigned timeslot time (a). It takes a longer time to reach all terminals in all WLANs. The "*Follow Method*" in (iii) directly switches, i.e. follows, the WLAN with the current timeslot (b) to reduce the time latency in (ii). The Follow Method brings more switching according to the higher event frequency. Therefore, the final method named the "*Threshold Method*" in (iv) controls the following operation by considering event frequency.

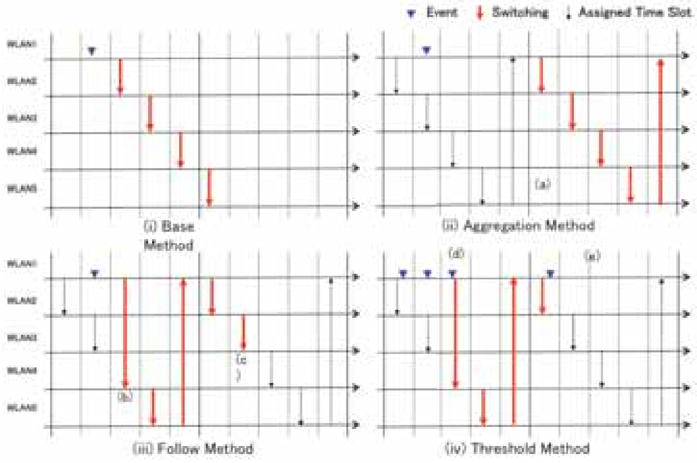


Fig. 2. Time Slot Based Content List Sharing Schemes.

III. EVALUATIONS

In the evaluation, the distribution of the end-user's sessions follows an exponential distribution [3]. The event frequency is configured by changing the exponential distribution, lambda. The simulation time is 5 hr. TP_{sw} and TP_{obsv} are 1 and 9 min. The number of WLANs is 3. T_s is 5 s.

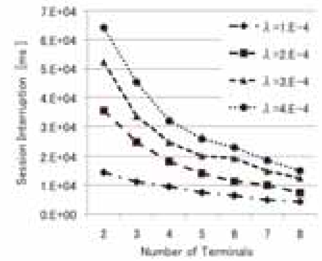


Fig. 3. Random method.

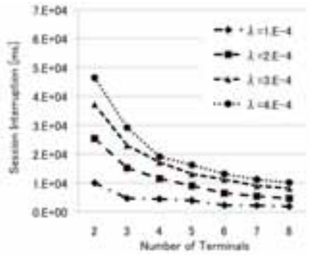


Fig. 4. Proposed method.

We show results of the random method and the proposed method respectively in Figs. 3 and 4. The x-axis and y-axis respectively represent the amounts of the interrupted session per terminal, called "session interruption", and the number of terminals. The random method decides the switching timing randomly within TP_{sw}. Session interruption is defined as the time lapping over the user's session and the network switching time. By predicting future sessions from the past state, the proposed method can better reduce interruption than the random method can.

Fig. 5 depicts the relation between the event frequency (x-axis) and the session interruption (y-axis). When the event frequency increases, the session interruption increases. The interruption session of the Base Method is the highest of all. The other methods aggregate network switches or follow the current switch based on the assigned timeslots. Frequent interruptions can be avoided. Fig. 6 presents results of sharing delay on the y-axis. Results show that the delay of the Follow Method is 320 [s] lower than that of the Aggregation Method, where the terminal does not switch until the subsequent time slot.

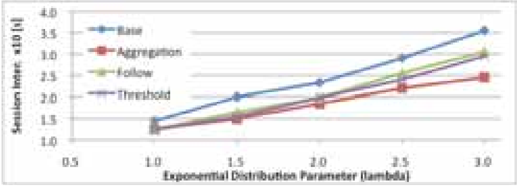


Fig. 5. Session Interruption.

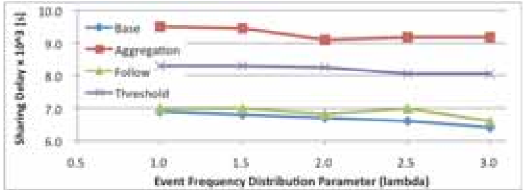


Fig. 6. Event Frequency and Sharing Delay.

The evaluations show that the Follow method produces the fewest session interruptions among the proposed methods. In addition, regarding the sharing delay, the Follow method provides the fastest performance among the proposals. Its performance is almost identical to that of the Base Method.

IV. CONCLUSION

The simulation demonstrated the effectiveness of the proposed time-slot-based event-driven method, especially the Follow Method. We will implement this proposed component on a Linux mobile terminal using open-source WLAN drivers.

REFERENCES

[1] D. Noh, and H. Shin, "SPIZ: An Effective Service Discovery Protocol for Mobile Ad Hoc Networks", *EURASIP Journal of Wireless Communications and Networking*, vol. 2007, 2007.

[2] L. Chou, W. Lai, C. Lin, Y. Lin, and C. Huang, "Seamless Handover in WLAN and Cellular Networks through Intelligent Agents", *Journal of Information Science and Enginnering*, vol. 23, pp. 1087-1101, 2007.

[3] H. Shino, K. Kitazawa, and K. Yana, "A Method for the Nonstationary Analysis of HTTP Communication Request Occurrences in Internet Access Networks", *IEICE-B* vol. J84-B, no. 8, pp. 1494-1504, 2001. (in Japanese)

Flexible Computing for Personal Electronic Devices

Daniel Díaz-Sánchez, *Member, IEEE,* Andres Marín López, *Member, IEEE,* Florina Almenares, *Member,* Rosa Sánchez, *Member, IEEE* and Patricia Arias, *Member, IEEE*

Abstract**-- This article describes an experimental framework for Android called Light Weight Map Reduce that pursues enabling Elastic Personal Computing, a refinement of the Elastic Computing concept that allows personal electronics to automatically distribute the load among devices constituting a computing fabric seamlessly.**

I. Introduction

The dream of flexible computing or computing utility has been pursued for a long time since it was first introduced [1]. The concept has now become a commercial reality with the name of Cloud Computing. However, current cloud systems allow accessing and manipulating resources that are located in a different place to the client device [2]. They usually require the data to be placed near the processing power (large data centers). Thus, in the end, the way client devices interact with the system is similar to the old-fashioned mainframes, there is single entry point that accepts requests and delivers the outcome, so the client device is just a client and not part of the process. Due to that, current clouds provide only part of the dream of flexible computing. It would be desirable to let devices to automatically discover resources, manage them and distribute the load among devices constituting a computing fabric seamlessly. For instance, a group of friends that have stored some pictures from a recent travel in their mobile phones could make a presentation with those pictures distributing the work among their mobile phones reducing the time it takes to process the pictures and keeping their privacy avoiding data centers to store personal data. In this article an experimental framework for Android called Light Weight Map Reduce (LWMR) that enables what we call Elastic Personal Computing (EPC) is presented.

II. The Concept and Related Work

EPC is a refinement of the Elastic Computing concept intended to provide such a computing fabric distributing the work load among consumer electronics belonging to a single, many individuals or even relying on shared "environment resources". The idea behind the concept is to move the processing power, better said "the job to be done", where the information is persisted to and not the other way around, minimizing bandwidth consumption, preserving privacy and improving consumer electronics resources by federating devices in an opportunistic fashion.

Among consumer electronics, personal devices are exceptional candidates to participate in some distributed computing tasks since are rich in personal and context information. This fact, together with the increasing computing power of personal devices, makes it more attractive distributing an operation over a set of data minimizing the communication among devices. Moreover, the social character of nowadays personal electronics would let the system to scale up beyond ownership limits since they are the preferred hub to interact with social networks. This leads to an scenario in which the correlation of the information persisted to personal devices belonging to individuals of the same social network would be meaningful.

A. Related work

A distributed computing system can be implemented in different flavors, and the approach taken to implement it influences several usability and scalability aspects. A direct approach would suggest to share processing power, disk and memory directly among devices. However, to maximize compatibility it would be necessary to rely on virtualization techniques that do not fit well in personal devices as mobile phones, STBs, TVs and non-commodity hardware. The highest level of abstraction would be to provide a programming and communication model that would be eventually implemented over heterogeneous hardware. A example of this approach is MapReduce [3] based systems, as Apache Hadoop, that are widely used in distributed computing solutions for data centers. LWMR follows the latest approach.

Prominent works in the area employ the same approach however many of them have just ported Hadoop to mobile scenarios [4] preserving the work balancing strategy. Hadoop manages the data knowing at any time the physical location of every worker node and replicates the data among adjacent nodes over a local network. A Hadoop cluster includes a master and multiple worker or slaves nodes. The master executes a JobTracker to which clients submit MapReduce jobs and that pushes the work to available TaskTrackers (worker nodes) pursuing to keep the work as close to the data as possible. Hadoop works very well in data centers however it is complex to be used in the target scenarios. Personal devices cannot be always located within a local networks, they can move frequently (mobile devices) and roam from LAN to WAN almost spontaneously. Moreover, the master role cannot be assumed by a single device as in Hadoop. To fully achieve the EPC concept every personal device should be able to act as a master as well as a worker node. Moreover, in EPC a job can be submitted by any device or group of devices, the outcome of the job can be collected by more than one device and it would be possible worker nodes to opportunistically delegate batch tasks to other devices upon battery, network or location

This work was partially founded by the Spanish Ministry of Science and Innovation TEC2010-20572-C02-01 CONSEQUENCE

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

changes. Finally, Hadoop assumes commodity hardware and tolerates big doses of hardware failures that cannot be assumed in EPC scenarios.

III. LIGHT WEIGHT MAP REDUCE FRAMEWORK

A distributed operation under the EPC concept requires one or several devices (job originators-JO) to distribute the load among originators and other devices (job slice executors - JSE). A job is a set of operations over a data set. The operations required by a job in EPC are usually calls to a common set of APIs that are already provided by the LWMR framework or calls to any custom piece of software. The latter requires the custom piece of software to be distributed among the JSE nodes. In regard to the data, EPC consider as candidates to act as JSE nodes those nodes that are near in terms of information availability so it is possible to distribute the load among nodes that are located far away from the JO in terms of network location. Thus JSE nodes should have the entire data set or part of it already stored in their storage space. Finally, if requested by JOs, the outcome of a job can be collected by another device. For instance, redirecting some processed images or video to a projector.

The LWMR framework enables any device to act as JO, JSE or both at the same time. For that reason, every device should instantiate the following LWMR services. Moreover, LWMR

Job Slice Executor Manager: this service keeps a list of candidate JSE nodes. Nodes in the list are given two scores, Affinity and Availability. The affinity measures the probability of the node to accept a job in terms of information availability. The affinity considers past operations so whenever a node accepts to act as a JSE node the affinity is increased pointing out that the node use to have enough data to execute the operation (could signal the owner of that device may have similar tastes) and the other way around. The affinity use to be a stable value. The availability measures how available is the device to cooperate. This information is provided by devices directly to the JO or updated in central service synchronized with the contact list in the case of mobile devices. A low availability value could mean the remaining battery is low, the network coverage is poor or just the device is busy.

Data Manager: The data manager keeps a list of the data sets available to be used. Every item within a data set is given an unique identifier using a one way mathematical operation. The way data are split in items is domain specific and depends on the nature of the information. For instance, a book can be split in chapters, sections or pages whereas a picture can be split in regions or kept entire... The data manager informs JOs if the operation can be conducted or not given the data.

Job Manager: The Job Manager is in charge of distributing a Job among devices using the information provided locally by the Job Slice Executor Manager and remotely by candidate's Data Manager services. The job distribution can follow different strategies depending on the job nature. If a job is critical or cannot be procrastinated, the Job Manager preselects candidate nodes with high affinity and availability scores. Other tasks can be given more tolerance so the Job Manager can preselect nodes with a different criteria. Once the pre-selection is done, the Job Manager checks the data availability for the particular job against the Data Manager services and filters the candidate list. Then it notifies selected nodes its intention to distribute a work and compose the final JSE list. After that it splits the job according to the data nature and assigns job slices to JSE nodes. A job slice univocally identifies the operation to be done and the data set over which it should be conducted as well as the outcome collection points.

Job Slice Executor: This service takes a job slice, executes the operation and delivers the outcome to the appropriate collection point.

IV. PROTOTYPE DETAILS AND CONCLUSIONS

The LWMR framework has been fully implemented in Java for the Android operating system. The current prototype implements the aforementioned services and three common APIs: text, mathematics and contacts. Service instances communicate among them by means of stateless HTTP messages using POST methods. The HTTP payload consist on JavaScript Object Notation, a text-based human-readable data interchange format. Services are implemented on top of a light weight open source android server called NanoHttp. Some simple experiments have been conducted to test the system using built-in APIs and custom Jobs. To test common APIs in a distributed fashion jobs calculating the word frequency of several big books (in plain text) and calculating sequences of numbers have been executed demonstrating the flexibility and reliability of the system. Custom jobs binaries and their dependencies are distributed through HTTP towards JSE nodes. Custom jobs should be programmed according an API and exported in Dalvik Executable (DEX) format.

We have designed and developed a distributed computing framework that pursues making EPC concept a reality allowing personal electronics to integrate themselves automatically into a computing fabric. We are on the most exciting phase of the project adding more functionality to the built-in APIs looking forward to increase device resources automatically and seamlessly for several application domains.

REFERENCES

[1] D. F. Parkhill, *The challenge of the computer utility*, Addison-Wesley Pub. Co. Reading, MA, 1966

[2] T. Velte, *et. al.*, *Cloud Computing, A Practical Approach,* McGraw-Hill, Inc., NY, 2010.

[3] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters", *6th Symposium on Operating Systems Design and Implementation*, pp. 137–150, 2004

[4] E. Marinelli, "Hyrax: Cloud Computing on Mobile Devices using MapReduce," M.S. thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, USA, 2009.

Implementation of an FPGA-Based Low-Power Video Processing Module for a Head-Mounted Display System

Assem A. M. Bsoul, *Student Member, IEEE,* Reynald Hoskinson, *Associate Member, IEEE,* Milen Ivanov, Shahriar Mirabbasi, *Member, IEEE,* and Hamid Abdollahi

Abstract-- **Portable head-mounted display (HMD) systems must balance functionality against battery life. To maximize operation time between battery recharges, we present a power-optimized field-programmable gate array (FPGA)-based implementation of an HMD video processing system. In this paper, power reduction is achieved using adaptive hardware-based sleep mode; this technique is performed by applying clock gating to the embedded microprocessor in our HMD system during idle times. Clock gating is available in many FPGA devices. Resource utilization and power dissipation results for the FPGA-based system are presented for different performance configurations.**

I. Introduction

State-of-the-art integrated head-mounted display (HMD) systems for fast-paced environments, such as action sports, are emerging and have been developed, e.g., by [1]. A sample goggle for snow sports is shown in Fig. 1. Such goggles provide continuous real-time information such as speed, latitude/longitude, etc. to the user. The data is shown to the user on an embedded display.

Due to its strict form-factor requirements, one of the main challenges for the HMD system in [1] is power consumption. Lower power consumption prolongs the battery life and leads to reduction of the total system cost, which in turn results in a better user experience.

The system in [1], however, can be further optimized to reduce its power consumption. Instead of having a separate chip to work as display driver and video buffer (explained further in Section II), all the system components can be integrated on the same chip to further reduce the power consumption and the chip-count.

In this paper, we investigate a field-programmable gate array (FPGA)-based prototype implementation for the video processing system in the HMD in [1]. Using an FPGA platform enables integrating most of the system components on the same chip and facilitates lowering the overall power consumption.

In this paper, we extend the results of [2] and report a low-power, FPGA-based architecture that can deliver the required graphical functionalities for HMDs. Furthermore, we propose applying an adaptive hardware-based sleep mode (HBSM) technique to the embedded microprocessor in the FPGA-based implementation to reduce its power consumption. Since the microprocessor executing the firmware in the proposed system becomes periodically idle for significant amounts of time, applying sleep mode to the microprocessor while it is idle results in significant power reductions.

Our study and measurements show that the power consumption of the proposed system can be lowered to 131 mW. In our system, HBSM can reduce power consumption by up to 60% compared to putting the processor in an infinite loop during its idle time. Compared to the work in [2], 50% power reduction is obtained using HBSM.

Fig. 1. Sample HMD for active sports.

This paper is organized as follows. Section II discusses the design challenges and provides a brief review of the FPGA-based system that has been investigated in [2]. Section III provides the details of the proposed low-power, FPGA-based implementation. In Section IV, we discuss the power measurement methodology that is used in this study. Section V provides a summary of experiments and the results. Finally, we conclude the paper in Section VI.

II. Background

A. Design Challenges

Fig. 2 shows the block diagram of the video processing system in the HMD in [1]. The embedded processor is responsible for rendering images to be displayed, and it responds to interrupts from external sensors. This system has an adequate battery life for the core functionality, but for more complex feature sets, or to enable smaller form factors, further reduction in overall power consumption is critical.

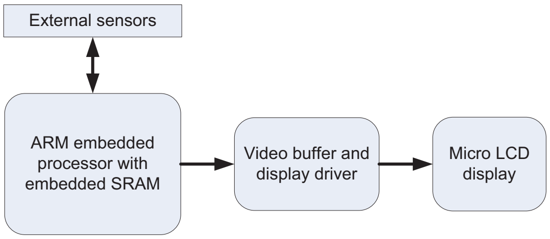

External sensors
ARM embedded processor with embedded SRAM
Video buffer and display driver
Micro LCD display

Fig. 2. Hardware architecture for current HMD system.

One approach for reducing power is to place the processor in sleep mode during idle times. This is realized in the system

This work was supported in part by MITACS Accelerate Internship Program and Recon Instruments Inc.

in [2] by placing the processor in an infinite loop for the duration of the sleep period. We refer to this approach as software-based sleep mode (SBSM). In SBSM, the processor's power is reduced because it is only executing the same few instructions while in sleep mode, thus limiting the usage of the processor's datapath to only a small part of it. Ideally, however, we would like to use a hardware-based sleep mode by using power or clock gating. Power gating [4] may cause losing the state of the processor, and it is not supported in many embedded processors. Clock gating [3], however, is performed by stopping the clock that synchronizes the operation of the system from switching without losing state information, thus eliminating the flow of data in the processor's datapath; this leads to significant savings in the dynamic power. In this study, we exploit the clock gating capability of an FPGA device to realize HBSM.

B. *Previous Work*

In a previous study [2], two preliminary FPGA-based implementations of the video system were investigated. These two systems focus on the memory subsystem organization, and its implication on the power consumption. Thus, further power saving techniques can be incorporated. The work in this paper is focused on developing a fully functional power-optimized version of the FPGA-based systems presented in [2].

III. Low-Power FPGA-Based Design

In this section, we discuss the details of the designed video processing module, and we show the proposed adaptive hardware-based sleep mode (HBSM) that is used to reduce the power consumption of the system.[2]

A. *Basic Design*

In the basic design, the input from sensors is emulated in the software by using random values for the data that changes every update cycle of the micro liquid crystal display (LCD), with a frequency of 1 Hz.

Fig. 3 shows the block diagram of the FPGA system. A synchronous dynamic random access memory (SDRAM) is used as the storage space for the firmware of the system and for the frame buffer. The frame buffer size is 210 KB, and stores an image of 300×224 resolution with 24-bit RGB pixels.

The processor has instruction and data cache memories. The sizes of these caches will be varied in our experiments in order to obtain different performance points as explained in Section V. The effect of the processor's clock frequency on power consumption is also discussed in Section V.

The video chain consists of the following components: a frame reader (FR), a color plane sequencer (CPS), and a clocked video output (CVO) block. The FR is a direct-memory-access (DMA)-like block that reads image data (24-bit RGB pixels) from the frame buffer (SDRAM) through the shared bus in bursts (32 words per burst, with 32 bytes per word). It then sends the data to the next stage in the video chain through a dedicated streaming interface. The CPS converts the parallel data (24-bit RGB) into a sequence of color planes (8 bits per plane) since the micro LCD has an 8-bit data bus. Finally, the CVO block generates the required *VSync*, *HSync*, and *Valid* signals to drive the micro LCD, and feeds it with an 8-bit video data per clock cycle.

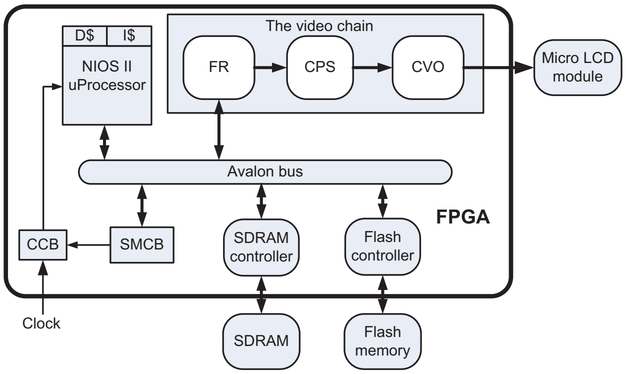


Fig. 3. FPGA-based design for HMD video processing system.

B. *Power Optimization*

The firmware running on the central processing unit (CPU) renders images to be displayed on the micro LCD based on the input from the user and data from the sensors, and stores the rendered image in the frame buffer. In our experiments, we found that this process could take between 160 to 250 ms when the CPU runs at 100 MHz with data and instruction cache size of 4 KB. For an update rate of 1 Hz, this means that the processor is idle for about 75 to 84% of the time, and can be placed into sleep mode to reduce power consumption.

We realized the HBSM using the clock gating functionality in the FPGA device. Fig. 4 shows an illustrative diagram describing the operation of the sleep mode by turning off the clock during the idle periods of the CPU.

The hardware consists of two components: sleep-mode control block (SMCB) and a clock control block (CCB). The SMCB can be configured with the length of sleep time using a software driver, and the CCB controls the clock network based on the input from the SMCB (see Fig. 3).

When the CPU finishes rendering image data and storing it in the video buffer, it executes a software routine to enter the hardware-based sleep mode, which in turn programs the SMCB with the amount of sleep time based on the type of the screen that has been rendered. This sleep period is calculated adaptively based on the time it takes the CPU to perform the rendering operation for the image. The SMCB in turn provides the required signals for the CCB in order to disable the clock network that provides clock for the CPU and the instruction and data caches, and starts a hardware counter to count down until the specified sleep interval passes.

Once the hardware counter hits zero, the SMCB is notified, and it in turn notifies the CCB to enable the clock network in order to resume the operation of the CPU. This starts the CPU execution from the point it stopped at before entering the sleep

[2] We use Altera Cyclone II device on the DE2 development and educational board.

mode, without losing any data or state information.

Clock gating during the idle period of the CPU eliminates the dynamic power consumed in the clock network, and the energy dissipated in the datapath. Since no data moves in the datapath of the CPU when the clock is stopped, no interconnect wires or gates are switching.

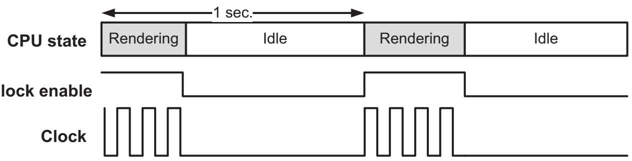


Fig. 4. Illustrating the use of clock gating to apply sleep mode.

IV. Power Measurement Methodology

In this section, we explain the methodology used to measure the power of the FPGA device used in the design described in Section III.

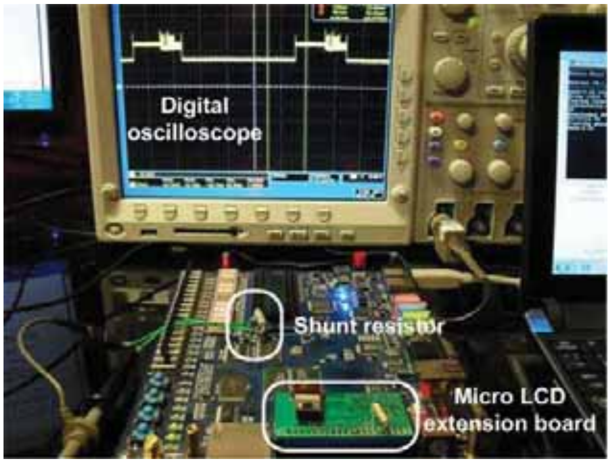


Fig. 5. Power measurement system including the FPGA development board with the micro LCD and a digital oscilloscope

We insert a 0.1-Ω resistor in series between the power regulator and the V_{CCINT} pin of the FPGA. A digital oscilloscope is used to measure the voltage drop across the resistor, which can be converted to the current drawn by the FPGA device core. Fig. 5 shows the power measurement system, which includes the development board with the micro LCD on a custom-designed extension board. The measured current is multiplied by the core voltage (1.2 V) to obtain the instantaneous power. Averaging the instantaneous power over a long measurement period provides an estimate of the FPGA power consumption.

Since the majority of the resources in the used FPGA device are unused, we have removed part of the static power which is due to the unused resources. We do this because in the ideal case we can implement our design using a smaller FPGA chip, which means less static power consumption.

V. Experiments and Results

In this section, we report results related to the amount of resources used to implement the power-optimized FPGA design, and the power measurement results for our system.

A. FPGA Resources Usage

Table I reports the amount of resources in the FPGA chip used to implement our design. Our implementation utilizes less than 1/3 of the resources available on the target FPGA chip. Only one phase-locked-loop (PLL) clock generator is used to implement the clock domains in our system. This PLL is used to generate 3 clock signals: 27 MHz signal for the micro LCD module, 70 MHz signal for the SDRAM controller and the video chain, and the third clock signal is used for the CPU. The frequency of the CPU's clock is varied in our experiments in order to investigate its effect on power dissipation.

TABLE I
FPGA resource usage for the FPGA-based implementation

Metric	Resources count
Logic elements	9586 (29%)
➔ Combinational	7180 (20%)
➔ Registers	6845 (22%)
Memory (KB)	17.2 (29%)
Embedded multipliers (9-bit elements)	4 (6%)
PLLs	1 (25%)

B. Power Dissipation

The firmware of the HMD system in [1] was ported to the NIOS II processor in the FPGA-based design. Fig. 6 shows snapshots of some screens that can be displayed by the system on the micro LCD. Since the CPU's clock frequency and cache size affects performance, and thus power consumption, we explore their effect in this subsection.

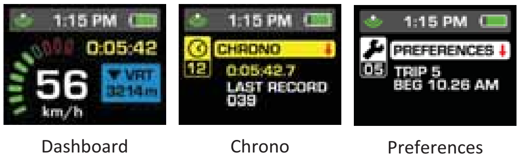


Fig. 6. Snapshots of example screens that can be displayed by the HMD.

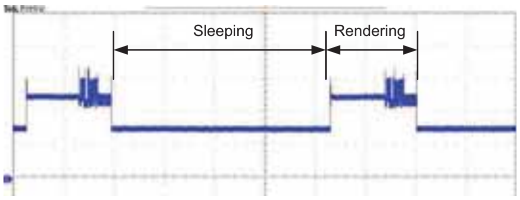


Fig. 7. Power consumption during one cycle of the micro LCD using HBSM.

We averaged the measured power over 20 cycles of the display update. Fig. 7 shows a snapshot from the oscilloscope for measuring power in one experiment for one display update cycle. The labels in the figure indicate the regions in the voltage measurement that correspond to different phases of the processor execution.

CPU Clock Frequency

Fig. 8 shows the power dissipation for two versions of the implemented system, one using SBSM and the other using HBSM. The figure also shows the percentage of power savings (HBSM compared to SBSM) as we vary the CPU's clock frequency.

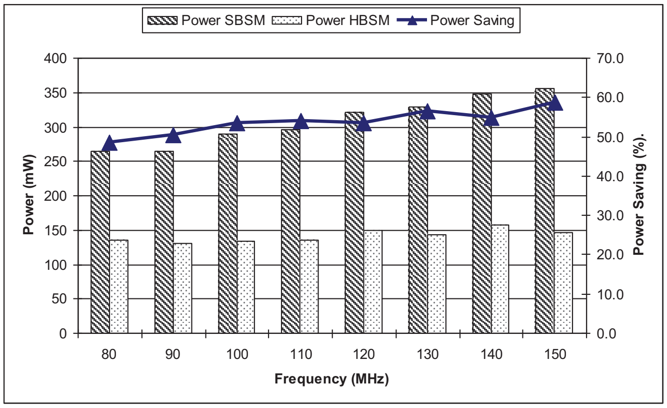


Fig. 8. Power dissipation and power savings (HBSM compared to SBSM) for different CPU clock frequencies.

The power consumption using SBSM increases as the frequency increases because of the linear relationship between the CPU's clock frequency and power consumption [5]. However, the power consumption for HBSM does not have the same linear relationship with the CPU's clock frequency. This behavior is a result of two conflicting factors. The first is the clock frequency itself; as the frequency is increased it increases the power consumption during the rendering phase, and at the same time increases the length of the idle period. The second is the length of the idle period, as it is increased (due to the increase in the clock frequency), it increases the length of the time that the processor spends in sleep mode, thus increasing the power savings.

In summary, the power-saving results using HBSM increase as the clock frequency increases. This is because the processor's idle time increases as the frequency increases, leading to more power savings. The obtained power savings are very promising, ranging between 49% and 59%. The lowest power consumption using HBSM is 131 mW at 90 MHz CPU clock frequency, which is 48% less than what was reported in [2]. Note that the power in [2] is less than the power in our system with SBSM because [2] does not fully implement all video display functionality. This is why the power savings in our system using HBSM compared to [2] are less than the savings compared to using SBSM in our system.

CPU Cache Size

We have also experimented with varying the CPU cache size and fixing the clock frequency at 100 MHz. Increasing the cache size potentially increases performance and the length of idle periods, which leads to less power consumption.

Fig. 9 shows the power results for SBSM, HBSM, and the power savings as the cache size is increased. As the cache size increases up to 4 KB, the CPU performance increases, hence increasing the amount of sleep time and power savings in HBSM. The savings range between 25% and 56% of the SBSM power. Interestingly, increasing the cache size by more than 4 KB does not help in increasing the performance; this can be attributed to the limited SDRAM bandwidth; hence, HBSM power is not reducing further at that point.

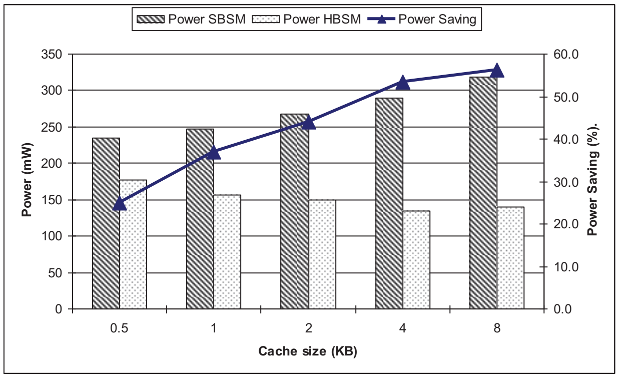


Fig. 9. Power dissipation and power savings (HBSM compared to SBSM) for different CPU cache size.

VI. Conclusion

In this paper, we presented a low-power, FPGA-based implementation for an HMD system. In our design, we observed that the power consumed by the CPU contributes to a significant portion of the total power dissipation. Therefore, we proposed to use an adaptive hardware-based sleep mode that is realized using clock gating.

Our results indicate that there is a strong relationship between the CPU's clock frequency and cache size, and the power consumption. Clock frequency and cache size affect the length of the idle periods during which the CPU could be placed in sleep mode to save power.

The proposed HBSM can reduce the power consumption by about 60% as compared to SBSM in our system. The proposed HBSM technique results in a system's power consumption as low as 131 mW, which is 48% less power consumption compared to the work in [2].

References

[1] Recon Instruments Inc. http://www.reconinstruments.com.

[2] D. Sengupta, R. Hoskinson, S. Mirabbasi, M. Ivanov, and H. Abdollahi, "Low-power FPGA-based display processing module for head-mounted displays," in Proceedings of the International Conference on Consumer Electronics, pp. 673-675, Jan. 2011.

[3] Q. Wang, S. Gupta, and J. H. Anderson, "Clock power reduction for virtex-5 FPGAs," in Proceeding of the 17th ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, pp. 13-22, Feb. 2009.

[4] S. Henzler, "Power management of digital circuits in deep sub-micron CMOS technologies," Springer Series in Advanced Microelectronics, Springer-Verlag New York, Inc., 2007.

[5] J. M. Rabaey, A. Chandrakasan, and B. Nikolić, *Digital Integrated Circuits: A Design Perspective*, Prentice-Hall, Inc., 2003.

[6] Altera Corporation, *DE2 Development and Educational Board User Manual v1.4*, 2006.

A Novel Mobile GPU Architecture based on Ray Tracing

Won-Jong Lee[1], Youngsam Shin[1], Jaedon Lee[1], Jin-Woo Kim[2], Jae-Ho Nah[2],
Hyun-Sang Park[3], Seokyoon Jung[1], and Shihwa Lee[1]
SAIT Samsung Electronics[1], Yonsei University[2], Kongju National University[3], Korea

Abstract-- **Recently, with the increasing demand for photorealistic graphics and the rapid advances in desktop CPUs/GPUs, real-time ray tracing has attracted considerable attention. Unfortunately, ray tracing in the current mobile environment is difficult because of inadequate computing power, memory bandwidth, and flexibility in mobile GPUs. In this paper, we present a novel mobile GPU architecture called the SGRT (Samsung reconfigurable GPU based on Ray Tracing) with the following features: 1) a fast compact hardware engine that accelerates a traversal and intersection operation, 2) a flexible reconfigurable processor that supports software ray generation and shading, and 3) a parallelization framework that achieves scalable performance. Experimental results show that the SGRT can be a versatile graphics solution, as it supports compatible performance compared to desktop GPU ray tracers.**

I. INTRODUCTION

Ray tracing is a physically correct rendering algorithm efficiently modeling the interaction between objects and lights, which produces highly realistic graphics images. Due to the requirements of massive computing power and memory bandwidth, ray tracing has been mainly used in off-line rendering field. However, recent rapid advances in desktop CPUs/GPUs and a variety of researches have made real-time ray tracing possible [1]. As a result, the ray tracing is expected to be a new graphics paradigm to create a new market in near future [2].

Mobile graphics has been another trend introducing a new user experiences. Mobile devices are widely used all over the world, and these platforms provide an opportunity creating the new graphics applications. Increased interest in mobile graphics can be seen in the activities of industry standard like OpenGL|ES. In order to maximize user experience, ray tracing is expected to be demonstrated on the mobile devices in near future.

Though mobile graphics capabilities and performance have advanced considerably in recent years, real-time ray tracing in current mobile GPU is very difficult due to the following reasons. First, computational power is inadequate. Ray tracing of a real-world application at HD resolution requires the performance of 300Mray/sec (about 1~2TFLOPS) is needed [3], but the peak performance of current flagship mobile GPU is no more than 256GFLOPS (ARM Mali T658 [4]). Second, mobile GPU lacks efficient branching supports. Ray tracing is a control-flow-intensive algorithm, but mobile GPU cannot fully support branches with limited stack memory. Third, execution model of the mobile GPU is multithreaded SIMD which is not suited for ray tracing, because it causes a divergent branching and memory access in secondary rays. These incoherent rays can lead to a poor SIMD efficiency.

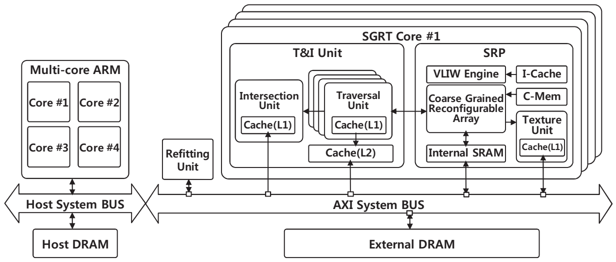


Fig. 1. Our system architecture including the SGRT cores and host processor

In this paper, we propose a new mobile GPU architecture, called SGRT (Samsung reconfigurable GPU based on Ray Tracing), which can solve the problems previously mentioned. The SGRT has three key features. First, it has a fast compact hardware engine that accelerates a traversal and intersection (T&I) which are computationally dominant operations in ray tracing. Second, it employs a flexible reconfigurable processor that supports software ray generation and shading (RGS). Third, it exploits a parallelization framework with real-time operation system (RTOS) that achieves scalable performance.

In addition, our system architecture is designed for recent application processor (AP) that integrates CPUs, GPUs, and DSPs into a single chip with SoC technology. We assign the major modules of the ray tracing into the appropriate computing resources of AP, which is a combination of the tree-rebuild module to reconstruct whole acceleration data structures (on multi-core CPUs), the tree-refit module to update only the changed nodes in the tree (on dedicated H/Ws) and the rendering (on the SGRT). Experimental results show that our GPU can be a versatile graphics solution for future application processor by exposing equivalent performance of recent desktop GPU ray tracers.

II. SGRT CORE ARCHITECTURE

Figure 1 shows the overall system architecture including the SGRT cores and host CPUs. This section describes our architecture in detail.

A. Dedicated Hardware for Traversal and Intersection

The lack of computational power and memory bandwidth of current mobile GPUs motivated us to design a dedicated hardware. Our H/W, called T&I engine, consists of multiple traversal and intersection units with multi-level caches for efficient memory usage, which is similar with previous ray tracing architecture [5][6]. But, unlikely the previous works, our H/W is optimized for mobile environment with the following features. First, it has a smaller area (3.89 mm^2 per core,

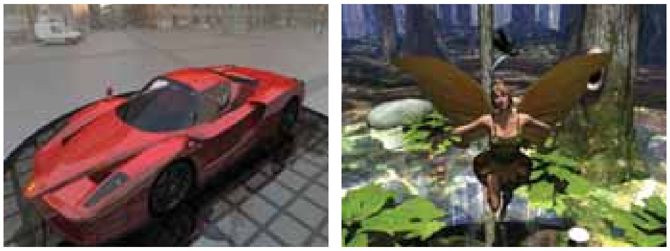


Fig. 2. Rendered images by the SGRT simulator: *Ferrari* (left, 210K triangles) and *Fairy* (right, 170K triangles).

65nm). For processing dynamic scenes, our H/W uses bounding volume hierarchies (BVH) that is an object hierarchy, which negates the need for LIST units to manage primitives. In addition, the traversal unit performs both BVH traversal and an intersection test between the ray and the primitive's axis-aligned bounding box (*primAABB*), which can significantly save the area. Second, we minimize the SRAM usage by employing short-stack based traversal algorithm [7]. Third, we combine a primAABB and pre-computed triangle data (*triAccel*) into a 32-byte aligned compact data, which increase the cache efficiency.

High performance features of our previous work [6] like the MIMD architecture for incoherent rays and a ray accumulation unit for latency hiding are directly reused in our H/W. Moreover, we can selectively utilize a specific BVH between the variants (e.g. Full SAH, Binned, SBVH, and LBVH) that are supported by the T&I engine.

B. *Reconfigurable Processor for Shading*

We utilize a proprietary low-power DSP core developed in our previous work [8][9]; it is called the SRP (Samsung Reconfigurable Processor). The SRP is very flexible for supporting full programmability; thus, various shaders (e.g. material and illumination) can be easily implemented. Unlike the conventional mobile GPU, the VLIW engine of the SRP can fully support control-flow such as recursion and branch, which make recursive ray tracing possible. In addition, the SRP is capable of highly parallel data processing. The coarse-grained reconfigurable array (CGRA) of the SRP makes full use of the software pipeline technique to allow loop acceleration. Therefore, the ray packet stream processing can be done in ray generation and shading kernels, which maximizes the utilization of the functional units.

C. *Parallelization Framework*

For scalable performance, we built a parallelization framework based on the Samsung Multi-platform Kernel (SMK) [10], a real-time operating system for embedded system. The SMK supports multi-tasking by systematic scheduling in the task queues, and it allows developers to create and use tasks easily. We define an individual task for each SGRT core that is responsible for different pixels (or pixel tiles), then the scheduler can distribute the next tasks to the idle SGRT core first, which results in dynamic load balancing. According to preliminary experiments, we could determine the performance scalability; 3.8x speedup on 4 SGRT cores compared to a single core.

III. Experimental Results

The validity of the SGRT is verified and its performance is evaluated during cycle accurate simulation. The Ferrari and Fairy has been thoroughly tested (Figure 2). Table 1 lists the performance results of ray tracing performed by the SGRT (4 cores), including shadow, reflection and refraction with WVGA (800x480) resolution at 1GHz clock speed. We achieve around 170M RPS (T&I engine), 255M RPS (SRP) and 87.82 fps (Fairy), which may be equivalent to the performance of recent desktop GPU ray tracers (~300M RPS).

TABLE I
Performance results of the SGRT architecture

Scene	# of tri.	# of ray	T&I engine				SRP	FPS
			Pipeline utilization	TRV $ hit ratio	IST $ hit ratio	MRPS*	MRPS*	
Fairy	170K	1.7M	87.27	93.83	96.53	171.32	255.72	87.82
Ferrari	210K	1.5M	79.75	92.56	92.92	122.48	319.56	67.83

*MRPS (Mega Rays Per Second)

IV. Conclusion

In this paper, we propose a novel mobile GPU based on ray tracing. This is a first approach to realize a real-time ray tracing in mobile environment, which has been impossible in state-of-the-art OpenGL-based mobile GPU due to the inadequate computational power and memory bandwidth. Furthermore, our system architecture is carefully designed to suit for mobile SoC platform. Simulation results show that our GPU can be a versatile graphics solution by presenting equivalent performance of recent desktop GPU ray tracers. We are now implementing the T&I engine at the RTL level, and we will release the complete GPU product targeted for future consumer electronics such as smart phone, tablet PC, and smart TV.

References

[1] I. Wald, W. Mark, J. Gunther, S. Boulos, and T. Ize, "State of the art in ray tracing animated scenes," *Computer Graphics Forum*, vol. 28, no. 6, pp. 1691-1722, 2009.

[2] H. Jim, "Ray tracing goes main stream," *Intel Technology Journal*, vol. 9, no. 2, pp. 99-108, 2005.

[3] P. Slusallek, "Hardware architectures for ray tracing," *ACM SIGGRAPH*, Course Notes, 2006.

[4] ARM Mali-T658 http://www.arm.com/products/multimedia/mali-graphics-hardware/mali-t658.php, 2012.

[5] S. Woop, J. Schmittler, and P. Slusallek, "RPU: a programmable ray processing unit for real-time ray tracing," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 24, no. 3, pp. 434-444, 2005.

[6] J.-H. Nah, J.-S. Park, C.-M. Park, J.-W. Kim, Y.-H. Jung, W.-C. Park, T.- D. Han, "T&I engine: traversal and intersection engine for hardware accelerated ray tracing," *ACM Transactions on Graphics (SIGGRAPH ASIA)*, vol. 3, no. 6, article 160, pp. 1-10, 2011.

[7] S. Laine, "Restart trail for stackless BVH traversal," *ACM Conference on High Performance Graphics*, pp. 107-111, 2010.

[8] W.-J. Lee, S.-O. Woo, K.-T. Kwon, S.-J. Son, K.-J. Min, C.-H. Lee, K.-J. Jang, C.-M. Park, S.-Y. Jung, and S.-H. Lee, "A scalable GPU architecture based on dynamically embedded reconfigurable processor," *ACM Conference on High Performance Graphics*, poster, 2011.

[9] W.-J. Lee, S.-Y. Jung, and S.-H. Lee, "An effective task scheduling scheme for multicore tile based rendering GPU," *ACM Conference on High Performance Graphics*, poster, 2012.

[10] Y. Shin, S.-W. Lee, M.-Y. Son, and S.-H. Lee, "Predictable multithread scheduling with cycle-accurate thread progress monitor," *ACM Symposium on Applied Computing (SAC '11)*, pp. 627-628, 2011.

The Portable Projection System Design Based on Light Emitting Diode Using Secondary Colors

Oh-Jin Kwon[1], Yongseok Chi[2], Youngseop Kim[2] , *Member IEEE* and Hack Youp Noh[2]
[1]Department of Electronics Engineering, Sejong University, Seoul, Korea
[2]Department of Electrical and Electronics Engineering, Dankook University, Yongin, Korea

***Abstract*--We propose a primary color overlapping method for increasing the brightness of projection system. It consists of a projected optical system based on 0.55 inch diagonal digital micro mirror device panel and a red, green, blue LED light source. This color overlapping method synthesized secondary colors of yellow and cyan instead of primary colors of red, green, and blue. By our method, the brightness of projected image was improved about 30 percent compared to a non-color overlapping method in a projection system with LED.**

I. INTRODUCTION

Light emitting diode (LED) has been known to be advantageous for portable projection systems. It provides a solid state light source. It is small in size, reliable, and does not need mercury. Its life time is longer than the lamp light source. Therefore, LED is very eco-friendly. Recently, the development of high bright LED makes LED to be adopted as the lighting source of projection systems [1]-[3].

Technically, LED is designed to operate with a forward driving current by the continuous waveform driving method and the pulse width modulation (PWM) driving method [1]. Figure1 shows the brightness and the efficiency characteristics of LED. It is noted that the brightness is not proportional to the current and the efficiency decreases as the current increases. This problem becomes serious when LED is used for the projection system where LED is used in the very high current mode [2][3].

In this paper, we design a new LED driving method. Whereas traditional driving methods use red (R), green (G), and blue (B) colors, our driving method uses R, G, B yellow (Y), and cyan (C) colors. So we did improve the driving efficiency of LED.

Technically, LED is designed to operate with a forward driving current by the continuous waveform (CW) driving method and pulse width modulation (PWM) driving method. But the brightness is not proportional to current because the efficiency decreases as current increases in Figure 1. This issue is very serious in projection system using LEDs(R, G, B), because LED is used in very high current mode. So we have designed new driving method. This new driving method algorithm used red, green, blue, yellow, and cyan. This is similar with multi primary color or Digital Light Processing (DLP)’s brilliant color [3][4] .

In this paper, by increasing the duty ratio of the secondary colors (Y, C and M) from color synthesis between primary colors (R, G and B), the driving efficiency of LED is improved. In our algorithm, the primary color overlap (synthesis) method is used to develop higher efficient and brightness in projection system. In section II, our proposed method is described. Experimental results are given in section III. And conclusions show in section IV.

II. PROPOSED SYSTEM

The fig 1 shows the brightness and efficiency characteristics of LED.

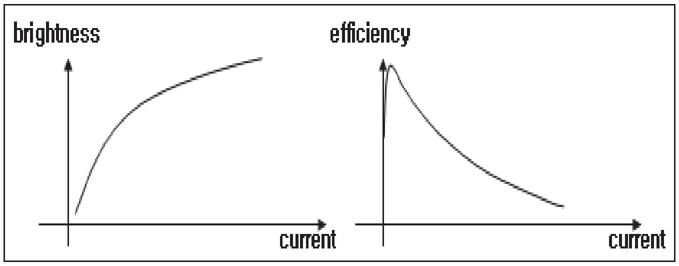


Fig. 1. Brightness and efficiency characteristics of LED.

The primary color overlap algorithm is used to develop a highly efficient and higher brightness than general LED projection (non-color overlap) system. The driving efficiency of LED is improved by increasing the duty ratio of the secondary colors (like this Y, C) earned from color synthesis between primary colors (R, G and B). In Figure 2, the driving pulse wave of LED (R, G, and B) is shown. Yellow is produced from red and green by synthesis, cyan is produced from green and blue by synthesis. The brightness (ANSI lumen) of the projector is measured as the duty ratio of the generated secondary color (Y, C and M) increases at 60Hz frequency (1frame). A frame is composed of several sequential colors for reduced the PWM noise and color sequence consists of primary color and secondary color (like a picture). For instance, when one frame is composed of 5 sequential colors in 60Hz, the available time of sequential color is 3.4msec. And secondary color period of the 3.4msec is called the overlap duty ratio.

The brightness (ANSI lumen) is measured by four synthesis methods that of yellow, yellow and cyan, yellow and magenta, yellow, cyan and magenta. A detail of the procedure is shown in Figure 3. The table is shown kinds of color overlap.

The more the duty ratio of secondary color from primary color synthesis is increased, the less the duty ratio of primary color is decreased due to the limits of frame frequency. This is the result of bad color linearity. Therefore, secondary color ratio is a very important factor in image quality.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

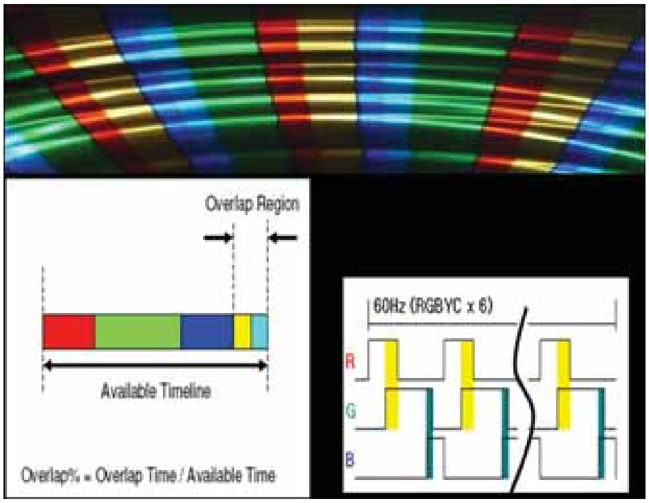


Fig. 2. Waveform of primary color overlap

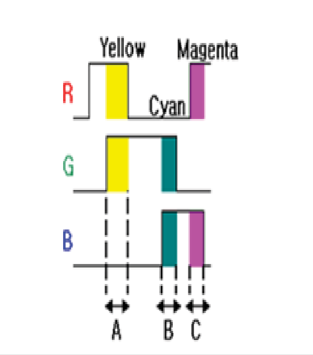
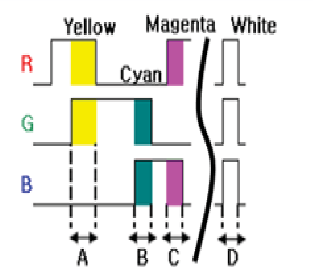
Generated color	Overlap ratio [%]	
No-overlap	A+B+C = 0%	
Overlap Y	A = 5% to 40%	
Overlap C	B = 5% to 40%	
Overlap M	C = 5% to 40%	
Overlap YC	A+B = 5% to 40%	
Overlap YCM	A+B+C = 15% to 35%	
Overlap YM	A+C = 5% to 40%	
Overlap W	D = 15% to 40%	
Overlap YW	A+D = 20% to 40%	
Overlap YCW	A+C+D = 25% to 40%	

Fig. 3. Overlap duty ratio of primary

III. Experimental Results

Figure 4 shows results that the color overlap method increases brightness (ANSI lumen) more than a non-color overlap method in a projection system. The more color overlap duty is increased, the more the rate of brightness is increased up to around 30 percent. When the rate of colors overlap duty is 40 percent, the rate of brightness is increased by 25 percent to 30 percent.

One of the best methods of increasing brightness is improving the efficiency of the LED. In order to improve the efficiency of the LED, several technical factors have been considered, these being a cooling system for the LED, primary color duty ratio for producing secondary color, driving forward current of a high-brightness LED and improvement of video image quality (quantization noise, color linearity, PWM noise).

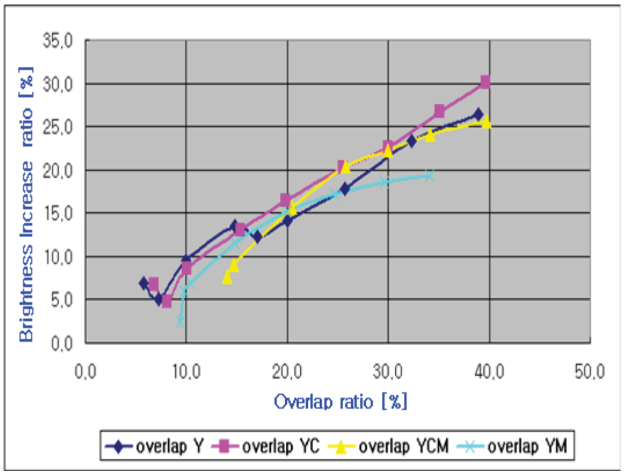


Fig. 4. Increasing rate of Brightness according to rate of color overlap

IV. Conclusions

This experiment created a high brightness LED projection system that used secondary color as the color overlap method for increasing brightness (ANSI lumen). The secondary color(Y, C, and M) increases the efficiency of LED in a high-brightness projection system that is compact which shows merit in a projector. This method improves brightness by 30 percent when primary color overlap is composed of red, yellow, green, cyan and blue sequential color.

Acknowledgment

This research was supported by the National Research Foundation of Korea (NRF), the Korean government grant funded by the Korea government (MEST) [No.2011-0010181, The Study of Green Video Compression and Joint Source and Channel Coding for Wireless Communication].

.

References

[1]. Takako Nonaka, “Additive Color Mixing Model Based on Human Color Vision for Bayer-type Pixel Structures.” IEEE Tenth International Symposium on.2006

[2]. K. Kurahashi, “Visual Color Shifts in Spatial Array of Three Primary colors.” Journal of Institute of Television Engineers of Japan, vol.40,5,pp.392-397,1986

[3]. W. Kunzman and G. Pettitt, "White Enhancement for Color Sequential DLP," Proc. Soc. for Information Display Conf., Soc. for Information Display, San Jose, Calif., 1998

[4]. Kawashima, M.; Yamamoto, K.; Kawashima, “Display and Projection Device for HDTV,” *IEEE Trans. Consumer Electron.*, vol. 34, pp. 100-110, 1988

Initial Direction and Speed Decision System for Auto Focus Based on Blur Detection

Quoc Kien Vuong, and Jeong-won Lee
DMC R&D Center, Samsung Electronics, Korea

***Abstract*—This paper proposes a new algorithm for deciding the starting lens focus direction and speed in aid of contrast auto-focus using only one initial image with blur detection. Simulation results show that the fundamental idea could support contrast auto-focus in reducing operation time by deciding appropriate starting direction in some certain circumstances.**

I. INTRODUCTION

There are two major types of auto focus (AF) methods commonly used in digital cameras: phase-detection AF (PAF), and contrast AF (CAF). While the former is mainly used in DSLR and late hybrid cameras with dedicated AF sensors, the latter, which works on actual image data, is applied in most compact devices and mobile cameras. Unlike PAF which can estimate the starting direction with any single image frame, CAF generally needs at least two images to judge whether the initial direction is correct or not. In the case of wrong direction, AF operating time may increase significantly. However, there is a lack of successful research on how to decide the initial direction with only one first image.

Originally, CAF aims at finding the AF lens position with maximum contrast value, or where the image region of interest (ROI) is least blur. From the second viewpoint, blur detection could be useful for CAF. There are a lot of blur detection algorithms in the literature, most of which are computationally not suitable for compact and mobile cameras. Furthermore, these methods aim at judging captured pictures of high quality with large ROI, while actual CAF only processes low quality live-view images of small size with even smaller ROIs. However, blur detection using Harr wavelet transform (HWT) [1] could be considered as an exception, owing to its lower complexity and computation compared to others. Besides, Fast HWT was verified to possess similar or better computational performance compared to the common 2-D FFT in relating areas such as image compression [2]. Therefore, blur detection using HWT could be used to support a CAF system.

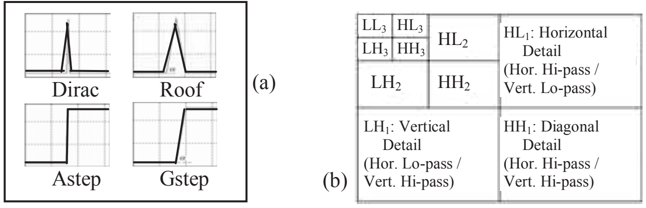
Section II of this paper introduces a modified blur detection algorithm using HWT. Section III illustrates the proposed method for deciding the initial AF lens direction and speed. Then, simulation results are presented in Section IV. Finally, conclusions are given in Section V.

II. MODIFIED BLUR DETECTION USING HARR WAVELET TRANSFORM

The blur detection scheme in [1] judges the blur extent of an image by analyzing different edges which are generally categorized into four types as shown in **Fig. 1**(a): Dirac-structure, A-step structure, G-step structure, and Roof-structure. After performing HWT on the original image up to the decomposition level of 3, three edge maps are constructed using (1) before partitioning with different window sizes. Local maxima are then identified using some small threshold.

$$Emap_i(k,l) = \sqrt{LH_i^2 + HL_i^2 + HH_i^2} \quad (i = 1, 2, 3) \tag{1}$$

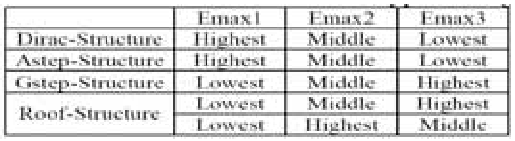
The effect of HWT on different types of edges varies; based on the observation listed in Table I, [1] proposes some rules to determine the edge type of each local maximum as well as the final total amount of edge points and numbers of each edge type. This set of rules results in the following indicators: total number of edge points N_{edge}, number of Dirac- and Astep-structure edge points N_{da}, number of Roof- and Gstep-structure edges N_{rg}, and number of Roof- and Gstep-structure edges that have lost their sharpness N_{brg}. The ratio $Per=N_{da}/N_{edge}$ is used to indicate whether an image is blur or not; the ratio $BlurExtent=N_{brg}/N_{rg}$ indicates the blur degree of that image.



LL$_3$ HL$_3$ LH$_3$ HH$_3$	HL$_2$		HL$_1$: Horizontal Detail (Hor. Hi-pass / Vert. Lo-pass)
LH$_2$	HH$_2$		
LH$_1$: Vertical Detail (Hor. Lo-pass / Vert. Hi-pass)			HH$_1$: Diagonal Detail (Hor. Hi-pass / Vert. Hi-pass)

Fig. 1. (a) Edge types; (b) Image with HWT sub-bands

TABLE I
EFFECT OF HWT ON DIFFERENT TYPES OF EDGES



	Emax1	Emax2	Emax3
Dirac-Structure	Highest	Middle	Lowest
Astep-Structure	Highest	Middle	Lowest
Gstep-Structure	Lowest	Middle	Highest
Roof-Structure	Lowest	Middle	Highest
	Lowest	Highest	Middle

However, the mechanism summarized above does not work well with live-view images captured with short exposure at high frame rate of 30fps (frame per second), 60fps, or even 120fps commonly used in cameras. Such images often possess smaller sizes and lower quality, especially due to much more noise caused by larger sensor signal gain (higher sensitivity). Meanwhile, [1] originally aims at high quality images with negligible noise. Therefore, this paper proposes a modified version of blur detection in consideration of all above issues.

Two major noise sources that affect live-view images are dark noise and shot noise which tend to add high frequency patterns to images. Besides, due to the short exposure and limited hardware resource, each live-view image and its ROIs have much smaller sizes than final captured images. All these factors make the HH component of each HWT decomposition level less meaningful to the process of judging edge points. Thus, (1) is modified with HH neglected as below:

$$Emap_i(k,l) = \sqrt{LH_i^2 + HL_i^2} \quad (i = 1, 2, 3) \tag{2}$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$Blur^{*} = BlurExtent - Per \qquad (3)$$

Equation (2) not only improves the accuracy when working with live-view images but also helps reduce the complexity and computation. *Blur** in (3) will be used as blur degree. The next modification step is noise threshold calibration which would result in a much higher value than that of [1]. In the future, with better platforms that can provide more accurate and sufficient statistics, detailed dark and shot noise analysis will be carried out to estimate noise level more precisely.

III. INITIAL LENS DIRECTION AND SPEED DECISION

A typical contrast curve is depicted in Fig. 2(a). At each lens position, a live-view image is used to calculate contrast value. If blur detection using HWT is applied to each image, the *Blur** curve would be illustrated as in Fig. 2(b) (blue solid curve). It can be seen that in defocus regions, *Blur** values are saturated at maximum value of 1; in focus region, *Blur** values are below 1 and *Blur** decreases very fast.

According to Fig. 2(b) and Fig. 3, modified blur detection curves show better performance. If the noise threshold is determined properly, *Blur** curve would have large enough on-focus region indicated by *w*. When *w* is ensured to be large enough, approximately one third of the whole full-scan lens range, *Blur** value could be used to decide initial direction with just one image. The rules are described as below:

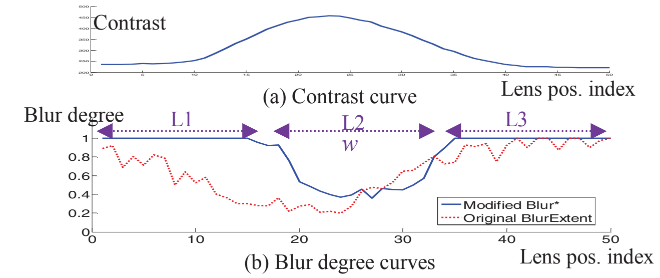
- Divide the whole lens range into three parts: L1, L2, and L3.
- If initial lens position is in L1 and the first *Blur** value is saturated or above a blur threshold close to 1, the contrast curve must have its peak located in L2 or L3. Thus, the lens can confidently move rightwards to L2 and L3. Similarly, if the same situation occurs with initial lens position in L3, the lens can move leftwards to L2 and L1. In these cases, the lens could move at a high speed (coarse search).
- For other cases, traditional direction decisions are invoked.

The proposed algorithm would be most suitable to cameras working with zoom lens at tele and macro modes when the subject usually locates in the middle of the lens range.

IV. SIMULATIONS

Simulations were carried out on our test camera that could send live-view image data to FPGA board. The proposed algorithm was tested with different objects and distances using luminance (Y) component; no color components were needed. When focus positions and initial positions both fell in either L1 or L3, normal CAF is conducted. In the case of focus positions falling in L2 as well as the case when focus positions were in L1 or L3 but starting positions were in L3 or L1 respectively, initial directions were decided correctly with high lens speed. The last case depicted in Fig. 3(d) is an extreme case with low contrast object under low luminance resulting in no quick decision and actual bad contrast curve. Original *Per* and *BlurExtent* [1] curves failed to work in all cases.

Blur* threshold = 0.9; Edge threshold [1] = 40 to 80;
$Per_{MinZero}$[1] = 0.15; $Epsilon_{Noise}$[1] = 10 to 80;



(b) Blur degree curves

Fig. 2. (a) Contrast curve, and (b) Blur degree curves

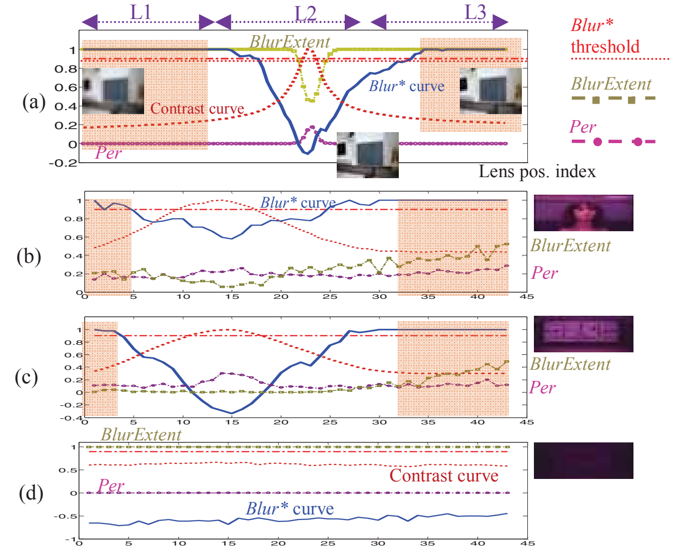


Fig. 3. Simulation: light brown rectangular areas cover initial lens positions that can quickly decide starting direction and speed for AF lens.

V. CONCLUSION

Simulation results show that the proposed method could assist AF system in deciding initial direction and speed for the lens with just one image in certain circumstances. The future research plan is to improve the accuracy as well as the range that is feasible to decide starting direction and speed. Furthermore, a more detailed noise analysis is expected to improve the overall performance in worse cases such as the case Fig. 3(d) discussed in the previous Section when the noise becomes significant due to the lack of contrast or due to higher sensor gain and longer exposure in low lighting conditions.

REFERENCE

[1] H. H. Tong, M. J. Li, H. J. Zhang, and C. S. Zhang, "Blur detection for digital images using wavelet transform," *IEEE Intl. Conf. on Multimedia and Expo*, vol. 1, pp. 17-20, Apr. 2004.

[2] S. T. Bow, Y. L. Sun, and L. H. Zhang, "Fast Harr wavelet transform for monochrome and color image compression," *Visual Information Processing VI (Proceedings of SPIE)*, vol. 3074, pp. 90-101, Jul. 1997.

[3] R. L. Lagendijk, "Basic method for image restoration and identification," *Academic Press*, 2000.

[4] X. Marichal, W. Y. Ma, and H. J. Zhang, "Blur determination in the compressed domain using DCT information," *Intl. Conf. on Image Processing*, vol. 2, pp. 386-390, Oct. 1999

[5] F. Rooms, and A. Pizurica, "Estimating image blur in the wavelet domain," *ProRISC*, pp. 568-572, 2001.

[6] T. H. Tsai, and C. Y. Lin, "A new auto-focus method based on focal window searching and tracking approach for digital camera," *Intl. Symp. on Comm., Control and Signal Processing*, pp. 650-653, Mar. 2008.

A Simulation Tool for Digital Autofocus Design

Dong-Chen Tsai*, Zuo-Min Tsai†, and Homer H. Chen*
* Graduate Institute of Communication Engineering,
National Taiwan University, Taipei, Taiwan
† Department of Electrical Engineering,
National Chung Cheng University, Chiayi, Taiwan
d96942024@ntu.edu.tw, zuomintsai@gmail.com, homer@cc.ee.ntu.edu.tw

Abstract—**To provide smooth viewing experience, the design of continuous autofocus of video cameras needs to take the scene dynamics into consideration in the algorithm development and testing cycles. However, this is often a time-consuming and tedious process because it requires a huge amount of tests on various real scenes. We propose in this paper a simulation method to simplify the digital autofocus design process. This design-by-simulation method significantly reduces the development and testing time of the search strategy for digital autofocus. Experimental results are shown to demonstrate the effectiveness of the proposed method.**

***Keywords*- Autofocus, simulation method, search strategy, still camera, video camera.**

I. Introduction

Focus measurement [1], [2] and search strategy [3], [4] are the two basic elements of autofocus (AF) for video cameras. The former measures the sharpness of an image and outputs a focus value, whereas the latter determines the direction and distance of lens movements to make the image captured by the camera in-focus. For the same scene, an image with higher focus value is considered sharper than the ones with lower focus value.

To provide good viewing experience, a search strategy of digital autofocus should take the dynamics of scenes into consideration and should work for various scenes. Conventional design methodology for the search strategy typically involves the steps of 1) developing a prototype of the search strategy, 2) implementing the search strategy on a camera, 3) testing the performance of the search strategy on various real scenes and recording the resulting video and the data (e.g. lens position, focus value, etc.) generated in the search process, and 4) modifying the search strategy through a careful analysis of the collected data and, if necessary, repeating Steps 2 to 4 for further refinement.

The iterative refinement process is time-consuming because it involves repetitive implementation and testing of the search strategy on various real scenes. Moreover, it allows us to only guess the cause of poor result because all available to us at the end of the process are the collected data and video. To address the problem, we propose a simulation method that allows us to evaluate the goodness of the search strategy without going through Steps 2 to 4. In other words, with this design-by-simulation method, we only need to implement the final search strategy on the camera; no other implementation is required. This significantly saves the development time of a search strategy.

This paper is organized as follows. Section 2 describes our simulation method. Section 3 verifies the effectiveness of our method, and Section 4 draws the conclusion of this paper.

II. The Simulation Method

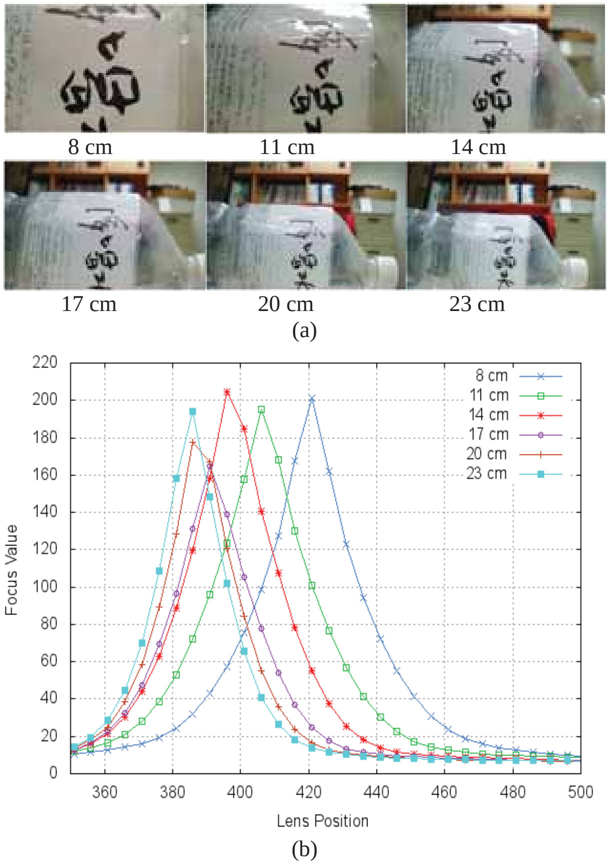
A search strategy determines the direction and distance of the next lens movement in the autofocus process. We decompose a dynamic scene into a number of still scenes. Each still scene has its own focus profile charactering the relation between the focus value of the image and the lens position. The focus data samples at the current and past sampling time needed for the search strategy are obtained from the corresponding focus profiles.

The basic idea of our design-by-simulation method for a dynamic scene is to generate the focus profiles of the static scenes that compose the dynamic scene and to use the resulting focus profiles for evaluating the performance of the search strategy. As an illustration, the sequence of a moving object at various distances to the camera is shown in Fig. 1(a). The corresponding focus profile of each simulated static scene is shown in Fig. 1(b). We can see that each focus profile has its own peak and that the lens position corresponding to each peak is different. This is exactly what it should be. The reciprocal focus profile representation [5] is used in our method to describe the focus profiles of the simulated static scenes. Each focus profile is modeled by a quadratic function, and the focus profile of the dynamic scene over time is represented by

$$\frac{1}{f(l)} = a(t)\left(l - m(t)\right)^2 + b(t), \tag{1}$$

where l denotes the lens position, $f(l)$ denotes the focus value measured at l, t denotes time, and a, b, and m denote the quadratic coefficients that vary with time. Therefore, we have different focus profiles at different time. In the example shown in Fig. 1(a), the focus profiles are obtained by setting

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a)

(b)

Fig. 1. (a) A moving object at various distances to the camera. (b) The corresponding focus profile of the dynamic scene at each object distance.

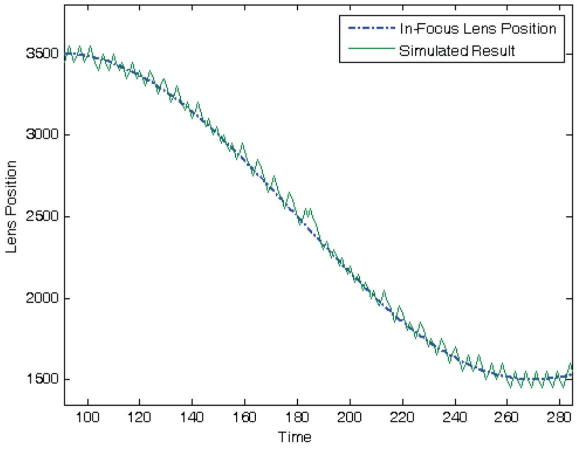


Fig. 2. Simulation result. The dash line represents the in-focus lens position, and the solid line represents the lens movements determined by (3).

$$m(t) = q \sin \omega t, \tag{2}$$

where the parameters q and ω determine the speed of the moving object, and letting $a(t)$ and $b(t)$ each be a constant.

III. Experimental Results

The effectiveness of our design-by-simulation method is tested by evaluating the performance of a commonly used search strategy, which determines the direction of lens movement from consecutive focus values, on the dynamic scene shown in Fig. 1(a). The details are as follows. Let the lens position and the focus value at time i be l_i and f_i, respectively. We determine the next lens position by

$$\hat{l} = \begin{cases} l_{i-1} + L \times \mathrm{sgn}(l_{i-1}\text{-}l_i), & \text{if } f_i < f_{i-1}, \\ l_i + L \times \mathrm{sgn}(l_i\text{-}l_{i-1}), & \text{otherwise.} \end{cases} \tag{3}$$

where L is a predetermined parameter. From the simulated result shown in Fig. 2, we may expect that a camera using this search strategy will exhibit bouncing, which refers to the behavior where the sharpness of a capture video sequence constantly changes between two different states.

We then evaluate (3) in real time by loading the code of the search strategy to a video camera provided by an unnamed manufacturer. As the code runs, the output video of the autofocus process is recorded. The resulting video is available for download from the website provided in Reference [6]. The bouncing behavior indeed occurs.

IV. Conclusion

In this paper, we have proposed a useful simulation tool for the design of digital autofocus. This tool significantly saves the development time of a digital autofocus design. It can be used to perform a preliminary evaluation of an autofocus search strategy on a computer before we actually implement the search strategy on a video camera.

References

[1] S. Y. Lee, Y. Kumar, J. M. Cho, S. W. Lee, and S. W. Kim, “Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, Sep. 2008.

[2] K. S. Choi, J. S. Lee, and S. J. Ko, “New autofocusing technique using the frequency selective weighted median filter for video cameras,” *IEEE Trans. Consumer Electron.*, vol. 45, no. 3, pp. 820-827, Aug. 1999.

[3] M. Gamadia and N. Kehtarnavaz, “A real-time continuous automatic focus algorithm for digital cameras,” *IEEE Southwest Symposium on Image Analysis and Interpretation*, Denver, March 2006.

[4] K. Ooi, K. Izumi, M. Nozaki, and I. Takeda, “An advanced autofocus system for video camera using quasi condition reasoning,” *IEEE Trans. Consumer Electron*, vol. 36, no. 3, pp. 526-530, Aug. 1990.

[5] D. C. Tsai and Homer H. Chen, “Effective autofocus decision using reciprocal focus profile,” *IEEE Int. Conf. on Image Processing*, Belgium, Sep 11-14, 2011.

[6] http://www.youtube.com/watch?v=AKpnkpZWYas&feature=youtu.be

Mirrorless Interchangeable-Lens Light Field Digital Photography Camera System

ByungJoon Baek, HyeongKoo Lee, YoungJin Kim, and TaeChan Kim
Samsung Electronics Co., Ltd

***Abstract*—Camera system is expected to be evolved merging computational photography technique for having rich image information while maintaining image quality. The proposed camera system shows a newly designed system architecture scenario to have both high quality image performance and computational photography technique by combining digital single lens reflex (DSLR) based architecture and light field concept into a single integrated system. It starts from conventional digital DSLR architecture and enhances system features and reduces the problems with previous techniques.**

I. Introduction

Camera has been an important tool for human to make image information while many other devices such as displays and many kinds of processors are focused on processing existing information. Recent cameras have evolved enhancing high quality imaging technology and improving convenience for use. High quality imaging technology includes the development of sensor, system architecture and precise mechanical control. Most of the current cameras systems also provide user convenient automatic mechanism such as auto-focus (AF), auto white balance (AWB) control and auto exposure (AE) control and user-friendly interface. DSLR camera and its family show well the trends described above.

Normal DSLR camera system has a swinging mirror to direct light to pentaprism viewfinder and phase detection AF sensor before exposure begins. Recent DSLR systems have gradually adopted the support of preview display and electrical viewfinder instead of optical viewfinder. There has been a tendency to incorporate more than one sensor such as AF sensor or second sensor for more functionality. In the near future, many emerging technologies of newly devised optical mechanism and corresponding computational photography techniques are expected to merge into pre-existing camera system, which means that camera system will be no longer a simple optical capturing system but complex tool for imaging and processing, accelerating camera revolution. It might provide richer image information and faithful imaging simultaneously.

This paper proposes architecture for near future camera system. The architecture incorporate light field camera concept into the system and replace AF sensor with it. Light field concept enables disparity evaluation [2] and therefore can carry out auto focusing similar to phase difference AF, the major focusing method of current DSLR system. The system reveals more enhanced features or potentials including single lens multi-view image generation and full frame based AF.

II. System Architecture

A. Conventional Architecture

Conventional DSLR has a mirror which direct light to the pentaprism viewfinder in order for a user to see what camera see. The mirror usually swings up for sensor exposure after focusing. Fig. 1 shows typical architecture of conventional DSLR.

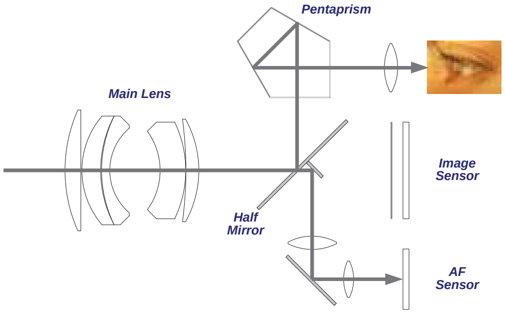
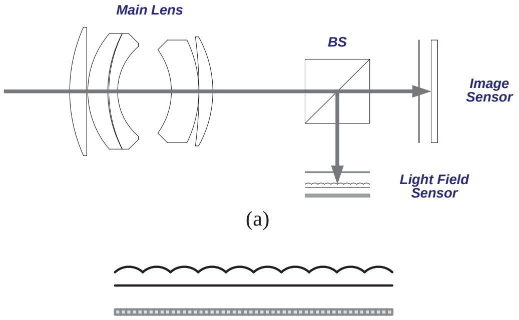


Fig. 1. Conventional DSLR camera system with swinging mirror mechanism and optical viewfinder

B. Proposed Architecture

Proposed architecture basically uses DSLR camera as a base architecture for high quality imaging and removes the mirror mechanism to reduce system size and weight. It uses beam splitter instead of mirror mechanism to separate and direct light to main and light field sensor which replaces an AF sensor. It needs a more processor or hardware logic for computational photography calculation. It is noted that in contrast to conventional DSLR, the proposed system doesn't mechanically separate composing/focusing and exposure time since both sensors receive the incoming light simultaneously.



(a)

(b)

Fig. 2. (a) Proposed camera system with a beam splitter (BS) and micro-lens array sensor (Light Field Sensor) (b) micro-lens array over pixel array in light field sensor

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. System Operation

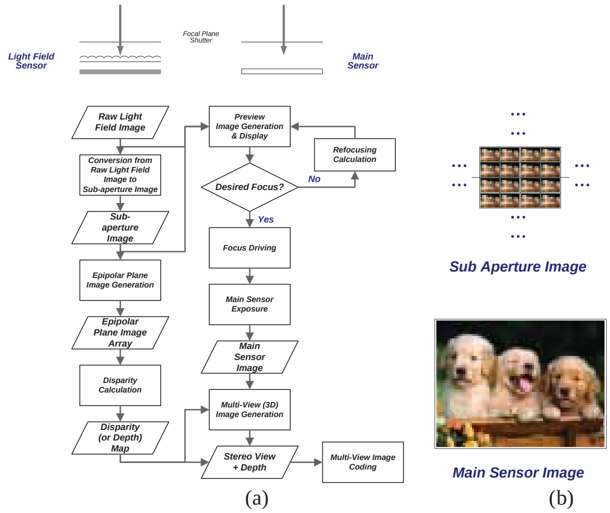


Fig. 3. System operational description. (a) Flow chart example (b) intermediated system outputs – sub-aperture image and main sensor image.

It is known that sub-aperture image can be extracted from raw light field image [2]. These sub-aperture images are interpreted as the multi-view images over the main lens aperture. Analyzing these images enables disparity calculation, which leads to the depth interpretation making it possible to replace AF sensor. According to depth information, focusing position is determined through user interface interaction, and then finally mechanical focus actuator is operated after the determination of focus value. Since focus is changed, re-capture of light field sensor might be needed in order to take focus-changed image. The combination of high quality image and light field image results in various possible computational photography application. As lots of researches have been done, multi-view image generation from depth and image is one typical example. It is noted that the proposed system makes it easy to generate multi-view image facilitating explosive 3D applications without using stereo view camera. The preview display can show not only what main sensor will capture but also intermediate or processed images such as multi-view thumbnail images, refocused and rendered images.

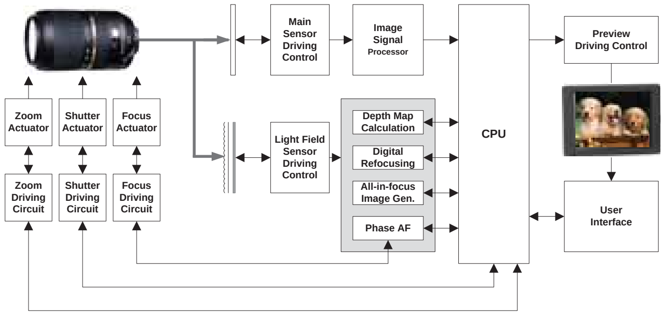


Fig. 4. Overall camera system for gathering rich image information and high quality images. The functions in shaded block can be performed on CPU.

IV. Potential Applications

The proposed system has many features compared with conventional DSLR camera system. Some are listed below.

A. Full frame based Phase-Detection AF

While AF performance of DSLR strongly depends on the shape and coverage of 1D AF sensors, the proposed one isn't affected by specific shape and coverage, since it cover whole image area at the cost of depth calculation. It provides more flexibility by removing re-aligning focus region to target.

B. Multi-view Image Generation

Typical 3D image format is stereo view image plus depth image [3]. Since scaled depth map could be calculated from the obtained images, the proposed system could be served as a tool for 3D image capture. Stereo view generation from single view and depth map is a prerequisite photography processing.

C. High Quality Image

Light field photography suffers from reduced image resolution since it uses large portion of pixels as storage for light direction information. Combining DSLR imaging mechanism compensates the reduced resolution.

D. Focusing Exploration without Mechanical Focusing

Digital refocusing ability of the system [1] enables focus change on preview screen without mechanical zoom and focus adjustment until exposure begins.

E. Compact Size & Light Weight

Removing swinging mirror mechanism significantly reduces the size and weight of overall system. Alternative to mirror, the system adopt beam splitter to direct light to multi-sensors.

V. Discussion and Conclusion

There are many points needed to be discussed. For example, separate exposure time control for multi-sensors may be required for application reasons. In addition, the performance and precision of phase detection could differ from each other.

In spite of computational complexity and difficult system design, upcoming camera system is expected not to stay as a faithful capture tool only but to serve as an integrated imaging machine merging digital computational photography and high quality imaging technology. The paper proposes one architectural candidate making both possible, which might facilitate camera evolution and discussion of emerging system.

References

[1] C.C. Chen, Y.C. Lu, and M.S. Su, "Light field based digital refocusing using a DSLR camera with a pinhole array mask," *Acoustics Speech and Signal Processing (ICASSP)*, pp. 754-757, 2010.

[2] R. Ng, *Digital Light Field Photography*, Doctoral Dissertation, Stanford University, 2006.

[3] R. Szeliski, "Scene Reconstruction from multiple cameras," *International Conference on Image Processing (ICISP)*, pp 13-16, vol.1, 2000.

[4] A.Kubota, "Reconstructing Dense Light Field From Array of Multifocus Images for Novel View Synthesis," *IEEE Transactions on Image Processing*, Vol. 16, No. 1, pp. 269-279, 2007.

Light field acquisition using wedge-shaped waveguide

Chang-Kun Lee[1], Taewon Lee[1], Hee-Jin Choi[2], Jae-Hyeung Park[3] and Sung-Wook Min[1]
[1]Department of Information Display, Kyung Hee University, Seoul 130-701, South Korea
[2]Department of Physics, Sejong University, Seoul 143-747, South Korea
[3]School of Electrical & Computer Engineering, Chungbuk National University, 410 SungBong-Ro, Heungduk-Gu, Cheongju-Si, Chungbuk, 361-763, South Korea

***Abstract*-- We proposed the light field acquisition system using the wedge-shaped waveguide. To obtain the light field reflected from the object, the lens array is mounted on the wedge-shaped waveguide. The light field information is forwarded through the waveguide and captured. In the experiment, we obtained the light field image variant with the object distance.**

I. Introduction

Recently, the attention for the interaction between the user and the state-of-the-art electronic devices, such as TVs, PCs, and mobile devices, arises rapidly. Many users require the experience not only watching the display but also controlling the sights against the instinctive gestures. Keeping pace with the trend, some kinds of technology are researched and developed actively. Touch sensing is typically applied in the mobile phones or the tablet PCs. Since the user can give the limited order on the two dimensional surface, however, only simple gesture is available not for the complicated one. Another method using time-of-flight (TOF) camera has the advantages over detecting the three dimensional (3D) motion and sensing without the direct touch. This technology has a potential in the future, but it is immature so far. For instance, the low resolving power, the external attachment and the limited detectable range are pointed out as the weaknesses in TOF camera method.

The concept of the light field can be adopted to resolve the problem as mentioned above. When the 3D object reflects the light, the reflected rays are filled all over the space. In this condition, the light field means the positional and the directional distribution of the reflected rays from the object at a standard plane. Because the light field perfectly expresses the lights from the object in terms of the geometric optics, we can suppose that the light field has the 3D information of the entire object. Therefore, we can obtain the 3D information, such as the shape, color, position, and depth, of the object from the light field. This unique property enables to improve the resolving power by designing the optimized lens array and to sense the near space where the TOF camera method cannot detect. Though the light field is easily obtained by capturing the object through the lens array in general, the whole system is too bulky and hard to be instrumented [1]. To reduce the size of capture system, we apply the wedge capturing system which consists of the thin wedge-shaped waveguide [2], [3]. As a result, the light field can be obtained in narrow space and the light field acquisition system is easy to be applied in the display devices.

In this paper, we designed the light field acquisition system based on the wedge-shaped waveguide and lens array and observed the correlation between the captured light field and the object distance from the system through the basic experiment.

II. Principle

In the light field pickup using the lens array, the reflected light rays from the object enter each elemental lens. Since the power of the elemental lens is limited, the angle to the optical axis of the rays affects the number of the lens accepting the light field information of the object. In other words, the object distance from the lens array can be calculated by counting the total sum of the lenses capturing the light field.

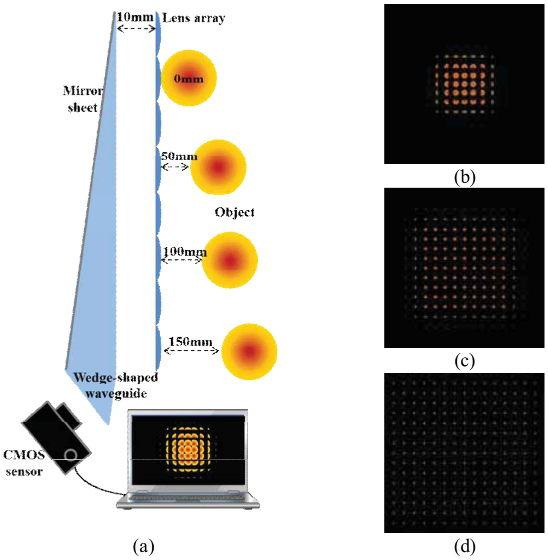


Fig. 1. Light field capturing systems scheme (a), Computer-generated elemental image from (b) 50mm (c) 100mm (d) 150mm

Fig. 1 shows the scheme of the light field capturing system using wedge-shaped waveguide in (a) and the computer-generated elemental image depending on each object distance

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

(b) 50mm, (c) 100mm, (d) 150mm. As the object distance from the lens array is increased, the number of each elemental lens obtaining the light field information is increased because the angle to the optical axis is decreased. On the contrary, the number of the light field information from the object in elemental image is minimal when the object is located near the lens array as shown in Fig. 1 (b).

Therefore, we can obtain the numerical correlation profile between the object distance and the light field information by counting the number of the lens accepting the light field. Applying the profile in real-time, it is possible to recognize the gesture of the object. Thus, the display system can be controlled by the 3D gesture as not only side to side motion but also back and forth without any direct touch on the screen. The light field acquisition is easy to apply in other devices because the size of the system is reduced by combining the lens array and the wedge capturing system.

III. Experiment

For the purpose of inspecting the feasibility of the proposed system, we set up the simple light field capturing system which consists of wedge-shaped waveguide made of acrylic, lens array and the CMOS camera. For a capturing device, USB camera is used for obtaining the elemental image from the object. The camera has the CMOS sensor area of 4.5 mm (width) by 2.8 mm (height), the number of pixel of 752 by 480 and the frame rate of 87 fps. The 15 by 15 lens array has the focal length of 10mm and single lens pitch of 5mm. The experimental setup is as shown in Fig. 2.

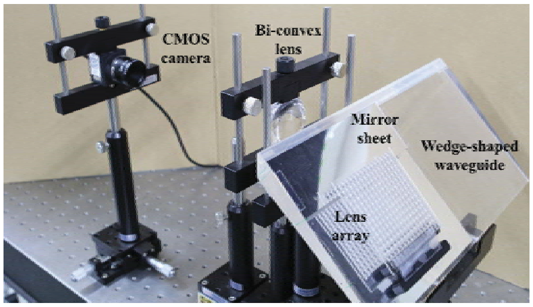


Fig. 2. Experimental setup of proposed system

The CMOS camera and the lens are located to capture the light field information focusing on the inclined surface of the waveguide. To image the light field on the inclined surface, the lens array is mounted away from the value of the focal length of 10mm. The mirror sheets are attached on the bottom side of the waveguide to improve the capture efficiency.

Fig. 3 shows the experimental results. The proposed system captures the size of 70 mm (width) by 40 mm (height) and the light field information from the object enters the 112 elemental lenses in Fig. 3 (a). The number of single lens accepting the light field information is 9 at 50mm, 14 at 100mm and 23 at 150mm, respectively. The number of single lens accepting the light field information in the experiment is smaller than that expected in the simulation. This is because the decrease in the light power is not considered in the computer-generated elemental image. In any case, we confirm the tendency between the object distance and the light field expected from the computer-generated elemental image.

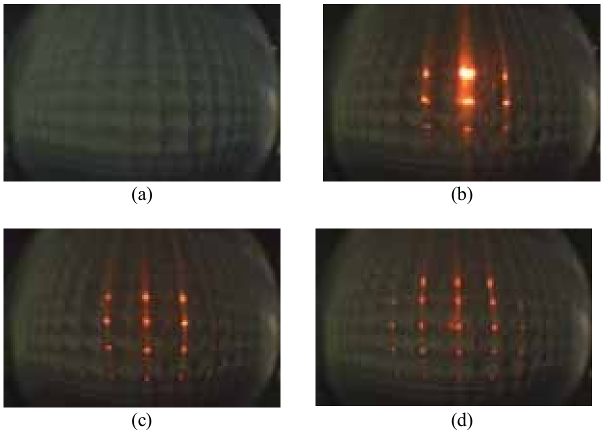


Fig. 3. Light field information of object captured in proposed system (a) lens array (b) 50mm (c) 100mm (d) 150mm

IV. Conclusion

We proposed the light field acquisition system based on the wedge-shaped waveguide to obtain the light field information depending on the factor of the distance from the lens array. By combining the wedge-shaped waveguide and the lens array, the light field information of the object near the system is captured in real-time. The experimental result shows that the captured elemental images are varied with the position of the object, likewise the computer-generated light field image. This consequence contributes to illuminate the depth position of any object by investigating the light field information. We expect that the proposed system could be applied in the 3D motion sensing and the user interface system.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0009225).

Reference

[1] J. H. Park, K. H. Hong, and B. H. Lee, “Recent progress in three-dimensional information processing based on integral imaging,” Appl. Opt. 48(34), H77-H94 (2009)

[2] A. Travis, T. Large, N. Emerton, Z. Zhu, and S. Bathiche, "Image capture via a wedge light-guide with no margins," Opt. Express 18 8453-8458 (2010)

[3] A. Travis, T. Large, N. Emerton, and S. Bathiche, "Wedge Optics in Flat Panel Displays," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE 2011), pp. 1-16

Parallel Pipelined Histogram Architecture Via C-slow Retiming

José O. Cadenas, *Member, IEEE*, R. Simon Sherratt, *Fellow*, IEEE, Pablo Huerta,
Wen-Chung Kao, *Senior Member*, IEEE, and Graham Megson

***Abstract*—A parallel pipelined array of cells suitable for real-time computation of histograms is proposed. The cell architecture builds on previous work to now allow operating on a stream of data at 1 pixel per clock cycle. This new cell is more suitable for interfacing to camera sensors or to microprocessors of 8-bit data buses which are common in consumer digital cameras. Arrays using the new proposed cells are obtained via C-slow retiming techniques and can be clocked at a 65% faster frequency than previous arrays. This achieves over 80% of the performance of two-pixel per clock cycle parallel pipelined arrays.**

I. INTRODUCTION

Image analysis based on histograms is abundant and well used in many consumer applications [1]. An array of cells to perform the computation of *m*-bin histograms that takes *k* pixels per clock cycle offers to gain a speedup factor of *k*. Such a design was proposed [2], but required a sensor or processor supplying four pixels per clock cycle to get a speedup of four. Many embedded microprocessors consist of 8-bit data buses and consequently are able to supply one pixel per clock cycle [3,4]. In order to exploit this property, a histogram solution using C-slow retiming to create two sub streams of computation derived from a dataset arriving at one pixel per clock cycle is proposed here.

This paper briefly explains the principle of C-slow retiming and applies C-slow to fully develop the proposed cells in section II before presenting final conclusions. The essential result is that the proposed design provides speed-up while also facilitating easier interfacing to camera sensors or microprocessors compared to other designs.

II. C-SLOWING RETIMING

C-slow retiming is a method used to reduce the critical path delay in digital circuits especially when feedback loops exist [4]. Every register in the datapath is replaced by *C* registers and then all registers are moved around on the critical data paths using a retiming algorithm. C-slow retiming separates the calculation performed in the original datapath into *C* instances. Fig. 1 shows an excerpt of the datapath of a histogram cell previously presented [2] that includes a feedback path (left), its C-slow version by a *C* factor of two (center) and after retiming to get a C-slow retimed version (right). A simple example using Fig. 1 illustrates the principle of retiming. For input sequence u = 3, 5, 4, 1 the left diagram in Fig. 1 produces r = 0, 3, 8, 12, 13; the leading zero reflects the register delay with output r being the running accumulation on input u. The diagram on the right of Fig. 1 gives r = 0, 0, 3, 5, 7, 6 for the same input u. The output corresponds to the accumulation as if there were two separate input streams: u_0 = 3, 4 and u_1 = 5, 1 and as such the output has been separated into r_0 = 3, 7 and r_1 = 5, 6; and the two interleaved into output r. In general C-slow retiming creates *C* interleaved streams of computation and as such also requires *C* input data streams. For practical reasons related to the design, only the factor C = 2 is considered in the rest of the discussion.

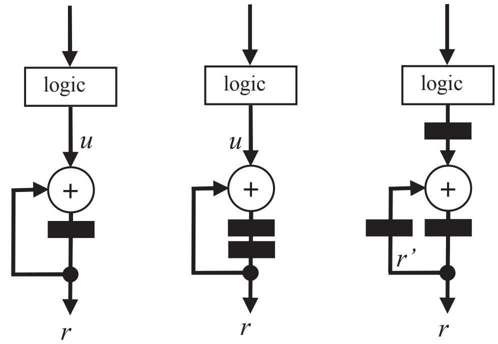


Fig. 1. Pipelined datapath with feedback (left), C-slow with C = 2 (center) and C-slow retimed (right).

A. Discussion on the C-slow effects

Fig. 1 demonstrates that the process of re-timing reduces the critical path delay from the cost of a binary adder and the associated logic to being either the time of the binary adder or the logic time whichever is longer. The downside is that the register count may increase significantly (by a factor of *C*). For example, compare the diagrams in Fig. 1 as retiming proceeds from left to right. The final architecture is also influenced by the specific places within the datapath where the registers are finally placed (due to datapath widths.) So, r' = 0, 0, 0, 3, 5, 7 (Fig. 1 right) and $r + r'$ = 0, 0, 3, 8, 12, 13 implies the cost of an extra adder is required to merge the two streams; this is unavoidable in the context of the example and also applied to computation of histograms.

B. C-slow retimed histogram processing cell

A C-slow retimed (*C*=2) processing cell for the computation of histograms is presented in Fig. 2. This follows straightforwardly from the above discussion and the histogram cell presented [2]. The new registers introduced by C-slow retiming are shown in gray. The mechanism to read bins out from the cell in a pipelined fashion has been omitted for simplicity.

The cell structure above the Logic block has been preserved except for the fact that C-slowing by a factor of two replicates the pipeline registers moving data left to right in the original

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

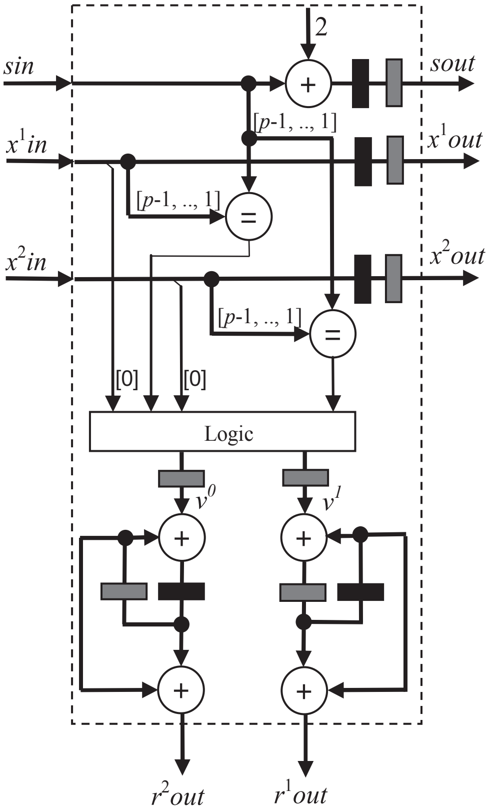


Fig. 2. C-slow retimed cell internal structure processing two data items while computing two histogram bins.

design. It should be appreciated that the structure looks very much like an instance of Fig. 1. It follows that, the separation of the computation into two streams does require the use of the extra adders as seen at the very bottom of Fig. 2. The critical path delay for the cell is now either the comparison followed by the block of Logic or the adder. Without C-slow retiming the critical path is due to the compare-logic-accumulation chain.

C. Results and analysis

A design was tested using ASIC technology of 35 microns giving the results in Table 1. Although the C-slow cell is only around 25% faster than the standard pipelined cell [2] the real advantage comes when the cells are arranged as an array. A pipelined array accepting 2 data items per clock cycle computes the histogram in $n/2 + m/2$ clock cycles with each cell processing two bins; $m/2$ is the latency. The C-slow cell in Fig. 2 requires two data items per clock cycle. Assume the cell of Fig. 2 is fed with every other data item (from an input dataset of n items) every clock cycle: half the items go into the array stream piped through x^1in input and the other half into through x^2in input. As a result an array processes a single data item per clock cycle. Thus, the histogram is computed in $n + m$ clock cycles (the latency is m even for cells computing two bins since each cell in Fig. 2 has a latency of two clock cycles.) As $n >> m$ for typical image sizes, latency can be ignored for a quick analysis.

TABLE I
HISTOGRAM ARRAY FREQUENCY AND AREA

	MHz	No. gates
Cell [2]	226	562
C-slow retimed Fig. 2	282	1366
Histogram array [2]	144	86336
Histogram array of C-slow cells of Fig. 2	238	194840

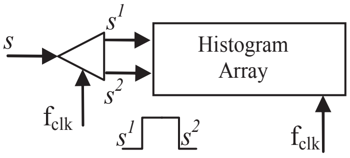


Fig. 3. Principle of operation of a double data rate to create two input streams out of a single input stream.

Arrays of C-slow retimed cells can be clocked 65% faster than the histogram arrays previously proposed [2]. In fact, from Table 1, ratio $T_{\text{pipe}}/T_{\text{C-slow}} = 1.65$ between the pipelined array and the C-slow array, then the time to compute the histogram for any dataset of size n with the C-slow array (one data item per clock cycle) reaches over 80% of the throughput delivered by a parallel (of two data items per clock cycle) pipelined array. The separation into two streams from a single dataset can be accomplished using a double data rate arrangement [5]. The principle of operation of dual data rate is shown in Fig. 3. The single stream s is distributed into two sub streams s^1 and s^2 by a de-multiplexer operating at both edges of the clock, so streams s^1 and s^2 are both output at a frequency f_{clk}.

III. CONCLUSIONS

A new array of cells computes m-bins histograms on streams of one pixel per clock cycle at over 80% of the performance of a pipelined array, working on streams of two pixels per clock cycle. This is due to arrays of C-slow cells achieving 65% faster clocks than previous pipelined arrays. The proposed array is consequently better suited for when camera sensors or microprocessors are limited to supply one pixel per clock cycle.

REFERENCES

[1] H.-C. Huang, F.-C. Chang and W.-C. Fang, “Reversible data hiding with histogram-based difference expansion for QR code applications,” *IEEE Trans. Consumer Electron.*, vol. 57, no. 2, pp. 779-787, 2011.

[2] J. O. Cadenas, R. S. Sherratt, P. Huerta and W. C. Kao, “Parallel pipelined arrays for real-time histogram computation in consumer devices,” *IEEE Trans. Consumer Electron.*, vol. 57, no. 4, pp. 1460-1464, 2011.

[3] K. Yoon, C. Kim, B. Lee and D. Lee, “Single-chip CMOS image sensor for mobile applications,” *IEEE J. on Solid State Circuits*, vol. 37, no. 12, pp. 1839-1845, 2002.

[4] Available: www.mipi.org/specifications/camera-interface#CPI

[5] C. Leiserson, F. Rose and J. Saxe, “Optimizing synchronous circuits by retiming,” 3rd Caltech Conf. on VLSI, 1993.

[6] R. S. Sherratt and Oswaldo Cadenas, “A double data rate architecture for OFDM based wireless consumer devices,” *IEEE Trans. Consumer Electron.*, vol. 56, no. 1, pp. 23-26, 2010.

Energy-Efficient Data Synchronization for Cooperative Context-Aware Computing

Wonjong Noh, Tae-suk Kim, Jaehoon Kim, Changyong Shin and Kyunghun Jang

Samsung Advanced Institute of Technology, Samsung Electronics, KOREA

Abstract— **In cooperative context-aware computing, the synchronized context information among devices is critical to the system performance. In this work, we propose an energy efficient context synchronization scheme using the reference-credit concept. Simulation results show that it improves energy efficiency by approximately 50% compared to the legacy random synchronization scheme which is popular in use. The proposed scheme can be employed as a key context synchronization function of the context-aware devices and computing platforms.**

I. Introduction

As communication and network technologies evolve, context-aware computing is attracting much interest. The context-aware computing refers to a general class of mobile systems that can sense their physical environment, and adapt their behavior accordingly [1]. For effective context-aware computing, context synchronization (or context consensus) among context-aware devices is one of the most important issues. There are some studies [2],[3] on the context synchronization. They employed a central controller such as a central database or gateway for the synchronization. However, the assumption on the existence of the central controller is not applicable to distributed or ad-hoc context-aware environments. Therefore, some studies [4],[5]worked on the context synchronization under distributed network environment. They proposed random synchronization schemes using device mobility. They provide good performance in terms of the convergence speed, but not in terms of energy efficiency. Thus, this paper proposes an energy-efficient context synchronization scheme for the cooperative context-awaring services in distributed network environments.

In Section II, we explain system model. In Section III, we present our proposed scheme. Sections IV and V give performance evaluation and conclusion, respectively.

II. System Model

As shown in Fig. 1, there are N mobile sensing/actuating devices which cooperate together for a context-aware service. Each device senses its context data every T slots. The sensed context data are expected to be fully exchanged and synchronized within the measurement interval T. Each device then takes its proper context-aware action according to the synchronized context data. For example, we simply consider a temperature as a required context information. Then, according to the agreed temperature information, one device can send a fire alarm signal and other devices can sprinkle water toward the high temperature region in a cooperative manner. We assume that all the devices measure and exchange the sensed data at the same time. Whenever they exchange the sensed data, the accuracy of the context information gets better while the energy consumption for the exchange increases. Thus, it is obvious that there exists a trade-off between the context accuracy and the energy consumption. In the following, we propose a reference-credit based context synchronization scheme to maximize the system energy efficiency.

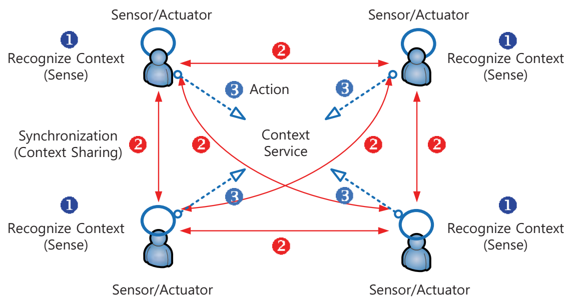


Fig. 1. System model: cooperative context-aware service

III. Proposed Synchronization Scheme

Initially, device i has its initial measured context data $x_i(0)$, consumed energy $e_i(0)$ and credit $c_i(0) = 0$. Whenever devices in the network share their context data to make a context synchronization at time t, they are given some credit $\tilde{c}(t)$. The credit value $\tilde{c}(t)$ comes from the following credit function.

$$\tilde{c}(t) = \left(1 - \frac{t}{T}\right) \cdot \delta^t, \quad 0 < \delta \leq 1, \tag{1}$$

where $\tilde{c}(t)$ has following characteristics.

$$(a) \quad \text{a monotone decreasing function of time, } \frac{d\tilde{c}}{dt} < 0$$

$$(b) \quad \tilde{c}(0) = 1 \text{ and } \tilde{c}(T) = 0$$

The credit value given as a reward decreases linearly or exponentially depending on δ. We define the *reference-credit* $c^{ref}(t)$ as the total credit which a device is expected to keep at time t.

$$c^{ref}(t) = \alpha \sum_{i=1}^{t} \tilde{c}(i) \tag{2}$$

In (1) and (2), two control variables α and δ can be controlled on an application basis by the system operator. Then, every

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

time slot t, the device i decides whether they join in context data sharing using (3).

$$\text{Join} * I_{[c_i(t) < c^{ref}(t)]} + \text{Sleep} * I_{[c_i(t) \geq c^{ref}(t)]} \qquad (3)$$

where $I_{[\cdot]}$ is a function having 1 or 0 depending on whether the condition $[\cdot]$ is satisfied or not, respectively. In (3), $c_i(t)$ is the total credit of the device i, which has been accumulated up to time t. If $c_i(t)$ is higher than the reference-credit $c^{ref}(t)$, the device i sleeps for that time-slot to save its energy because it has already contributed to the context synchronization process more than expected. Otherwise, the device i randomly chooses either sending mode or receiving mode. In sending mode, the device i chooses one of its neighbor devices to share its context data considering its channel status. In receiving mode, the device i observes several devices requesting the context sharing. Then, the device i chooses a device j having the highest credit, where the credit information of the devices in sending mode is sent together in their request message. Considering the total credit as a selection criteria in receiving mode encourages the devices to join in context sharing process, and can prevent devices from waiting to the last minute for free riding. The context sharing at time τ is carried out as follows.

$$x_i(\tau) = x_j(\tau) = \frac{c_i(\tau-1)x_i(\tau-1) + c_j(\tau-1)x_j(\tau-1)}{c_i(\tau-1) + c_j(\tau-1)}$$

After the context sharing, the device i and the device j update their total credit and consumed energy.

IV. EVALUATION

We first observe the effect of the credit function and the reference-credit, and then compare proposed scheme with the legacy random synchronization. In our evaluation, there are 10 devices which need to share their context information in $500m \times 500m$ distributed network. Devices can exchange their context data with one of neighbor devices within its transmission range $20m$. The experiment used 200 independent runs.

Fig. 2 shows the convergence, error and energy consumption of the proposed synchronization process with different reference-credit coefficient α under a linear credit function with $\delta = 1$. As the reference coefficient increases, the average context error decreases and the energy consumption increases. There is almost no difference beyond $\alpha = 0.7$.

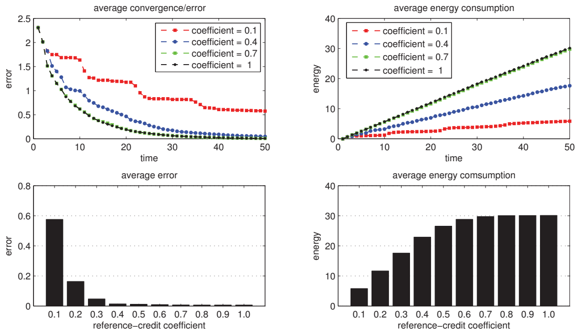


Fig. 2. The effect of the credit function and the reference-credit coefficients

Fig. 3 compares the random synchronization with our reference-credit based scheme in terms of convergence, energy consumption and energy efficiency. The reference-credit coefficient α is set to 0.3 with a linear credit function of $\delta = 1$. In Fig. 3 (a), the random scheme converges slightly faster than the proposed one. However, their is almost no difference since after the 35^{th} time slot. In Fig. 3 (b), the random scheme consumes approximately 100% more energy than the proposed scheme. In Fig. 3 (c), the random scheme has higher energy efficiency up to before the 13^{th} time slot. However, eventually, the proposed scheme achieves 50% higher energy efficiency than the random scheme. In addition, our proposed scheme can control the context data synchronization behavior using the credit function and the reference credit.

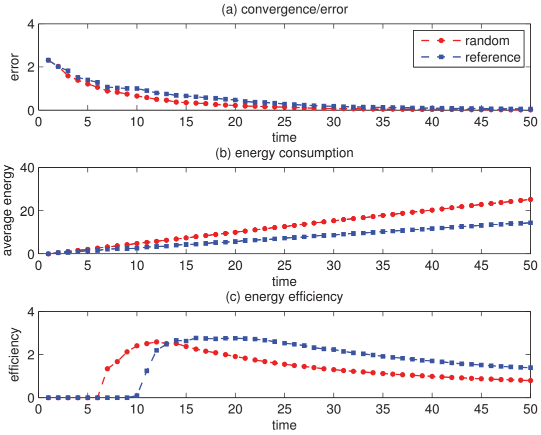


Fig. 3. Comparison of random and reference-credit based synchronizations

V. CONCLUSION

In this paper, we proposed an energy-efficient context synchronization scheme for the cooperative context-aware computing. The proposed approach is based on the concept of the reference-credit. Through the simulation, we confirmed that the proposed approach can offer approximately 50% better energy efficiency than the legacy random synchronization. The proposed scheme can be implemented as a key function in sensor devices, machime-to-machine devices and consumer devices in context-aware computing environments.

REFERENCES

[1] S. Loke, Context-Aware Pervasive Systems: Architectures for a New Breed of Applications, Auerbach Publications, Dec. 2007

[2] N. Miller, G. Judd, U. Hengartner, F. Gandon, P. Steenkiste, I-H Meng, M-W Feng and N. Sadeh, Context-Aware Computing Using a Shared Contextual Information Service, Pervasive 2004, Vienna, April 2004

[3] A. Malik, A. Manzoor, and S. Dustdar, Context-Aware Privacy and Sharing Control in Collaborative Mobile Applications, Engineering Science Reference, May 2012

[4] S. Boyd, A. Ghosh, B. Prabbakar, and D. Shah, Randomized Gossip Algorithms, IEEE/ACM Transactions on Networking, Vol.14, No.SI, pp.2508-2530, June 2006.

[5] E. Choi, C. Bae and J. Lee, Data Synchronization Between Adjacent User Devices for Personal Cloud Computing, IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, USA, Jan. 2012

Efficient High Throughput Rate Cross-Correlation Logic Design for Sign-Bit Reference Waveforms

Seok Joong Hwang, Donghoon Yoo, Soojung Ryu, and Jeongwook Kim
Samsung Advanced Institute of Technology

***Abstract*— This paper presents an efficient design method for high throughput rate cross-correlation logic using sign-bit reference waveforms which is one of widely equipped signal processing units in embedded consumer electronic devices. The proposed method minimizes resource usages by efficiently sharing the first level adders in adder trees of parallel sub-logics.**

I. Introduction

Cross-correlation is a standard method of measuring similarity of two waveforms as a function of a time-lag applied to one of them. It has been widely applied in consumer electronic devices which carry out digital signal processing, such as a baseband modem which detects a packet based on cross correlation of a received signal and a known preamble.

Computing cross-correlation is one of highly computational tasks in such systems. It requires a large number of multipliers and adders to carry out dot-product at a signal sampling frequency. This paper focuses on efficient implementation of cross-correlation logic particularly for high throughput rate one which is essential to deal with high speed signal sampling rate. High throughput rate cross-correlation logic generates multiple outcomes in a single clock cycle. Conventional high throughput cross-correlation logic consists of multiple parallel cross-correlation sub-logics which are identically designed while each sub-logic *w* processes distinct lagged signal windows as follow [1],

$$C_w(t) = \sum_{i=0}^{N-1} I(t+w+i) \cdot R(i), \quad 0 \le w < W \tag{1}$$

where $C_w(t)$ is a cross-correlation result generated by sub-logic *w* among *W* sub-logics at time *t*, *I* is a received signal waveform to which a time-lag is applied, and *R* is a reference waveform having *N* samples.

In cases that the resulting accuracy loss is tolerable, dramatic cost saving is possible by taking only sign-bits from a reference waveform, e.g., preamble, to replace multiplication with selective negation; it was reported that it incurs just 0.778 dB loss in cross-correlation results when applied to a packet detector [2]. And Hwang et al. proposed an efficient design concept for high throughput rate cross-correlation logic which takes sign-bit reference waveforms [3]. Their approach shares the first level adders in adder trees of cross-correlation sub-logics, which simplifies the design of sub-logics significantly. The shared adders simply provide summation results of two consecutive received signal samples. However, straightforward adoption of the idea cannot optimally share the first level adders, because odd-numbered and even-numbered sub-logics require summation results computed by each other due to the input alignment; there is no overlap between even-numbered and odd-numbered sub-logic inputs, since sub-logic *w* takes the following set of summation results,

$$\{I(t+2i+w) + I(t+2i+w+1) : 0 \le i < \left\lfloor \frac{N+W}{2} \right\rfloor\} \tag{2}$$

The contribution of this paper is to develop the proposed idea and present the detailed design method to optimally share the first level adders across all cross-correlation sub-logics. We evaluate our approach by implementing a cross-correlation logic which supports throughput rate 8 and reference waveform length 128. The evaluation results indicate that our approach could reduce 30.8% gate count compared to a conventional one like [1] which consists of identical parallel sub-logics without computational resource sharing.

The rest of this paper is organized as follows. Section II presents the detailed design method. Section III provides the evaluation. Finally, this paper concludes in Section IV.

II. Design Method

Equation (1) implies that cross-correlation sub-logic basically requires *N* selective negation logics and *N*-1 adders when only sign-bits of a reference waveform are taken. We observed that about a half of the adders in the sub-logic can be saved as shown the following equations that reorganize Eq. (1),

$$I_w(t, j) = I(t+w+j) \tag{3}$$

$$R_w(j) = R(j) \tag{4}$$

$$g_w(t,j) = \begin{cases} I_w(t,j) + I_w(t,j+1) & \text{if } R_w(j) \ge 0 \text{ and } R_w(j+1) \ge 0 \\ I_w(t,j) & \text{elif } R_w(j) \ge 0 \\ I_w(t,j+1) & \text{elif } R_w(j+1) \ge 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$B_w(t) = \sum_{i=0}^{N} I(t+w+i) \tag{6}$$

$$C_w(t) = B_w(t) - 2 \sum_{i=0}^{\lfloor N/2 \rfloor - 1} g_w(t, 2i), \quad 0 \le w < W \tag{7}$$

because summation results in Eq. (5) can be shared among the sub-logics and Eq. (6) can be cheaply implemented by buffered summation logic which requires just one register and two adders.

However, as addressed in Section I, summation results in Eq. (5) cannot be shared between even-numbered and odd-numbed sub-logics when a time-lag is directly applied to a received signal waveform in Eq. (3) and all sub-logics use the same reference waveform in Eq. (4) in the same way to the conventional high throughput rate cross-correlation logics [1].

In order to optimally share the first level adders, Eq. (3) and

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

(4) need to be rewritten as follows,

$$I_w(t,j)=\begin{cases} I(t{+}w{+}j) & if\ \ j\geq w \\ I(t{+}w{+}j{+}N) & otherwise \end{cases} \tag{7}$$

$$R_w(j)=\begin{cases} R(j) & if\ \ j\geq w \\ R(j{+}N) & otherwise \end{cases} \tag{8}$$

Equations (7) and (8) change coordination of the cross-correlation sub-logic inputs in order to equally align the inputs of all the sub-logics. But, this simple input coordination yields the optimal sharing of the first level adders.

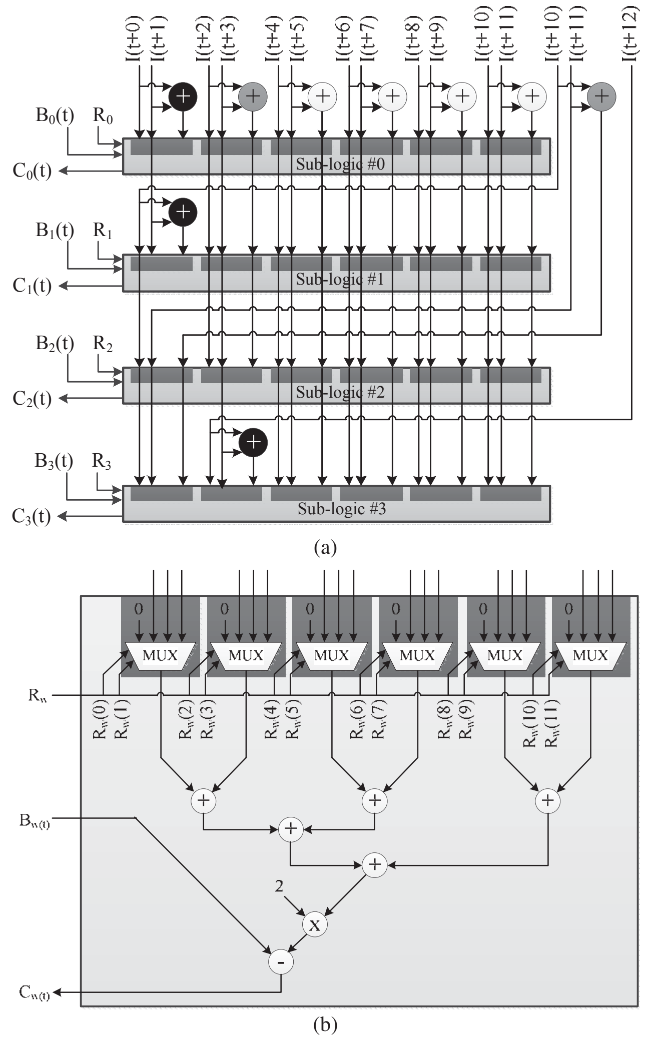


Fig. 1. A design example of high throughput rate cross-correlation logic; throughput rate 4 and reference waveform length 12. (a) The overall structure of cross-correlation logic. (b) The internal of the sub-logic.

Fig. 1 shows a design example of cross-correlation logic based on Eq. (5)-(8). Adders shown in Fig. 1(a) are the first level adders, where each of them is either fully (white), partially (gray), and non-shared (black). Our approach maximizes white colored adders, i.e., fully shared while obviously reducing gray and black ones. Thanks to the sharing of the first level adders, the amount of adders in sub-logics decreases about in half compared to a conventional one as shown in Fig. 1(b), e.g., 6 out 11. Selective negation logics (usually, XORs) are replaced with 4-to-1 multiplexers. Table I compares the number of logic components as a function of throughput rate and reference waveform length, denoted by *W* and *N*, respectively). Although our approach can be extended to multi-level adder sharing, it does not further reduce resources in a practical range of throughput rates since lots of shared adders are required for providing various combinations of summation and multiplexer inputs increases exponentially.

TABLE I
THE NUMBER OF REQUIRED LOGIC COMPONENTS

Component	The conventional	The proposed
Adder	$W\cdot N$	$W\cdot(N/2+1)$
XOR	$W\cdot N$	-
4:1 MUX	-	$W\cdot N/2$

III. EVALUATION

The proposed design method significantly reduces the hardware costs in high throughput rate cross-correlation logic by replacing selective negation logics and about a half of adders in cross-correlation sub-logics with 4-to-1 multiplexers and shared first level adders, respectively. In order to evaluate the benefit of the replacements, we implement a high throughput rate cross-correlation logic which supports reference waveform length 128, throughput rate 8, and 8-bit two channels (I/Q) for a packet detector of an MBOFDM baseband modem [3] with condition of 0.18-um CMOS process and 1.62 V supply voltage, and 528 MHz sampling frequency (8 times of 66 MHz logic clock frequency). As shown in Table II, our approach reduces 30.8% and 18.0% gate count in the cross-correlation logic and the entire packet detector which includes shift registers to hold received signals, respectively, from conventional ones.

TABLE I
GATE COUNT COMPARISON

Component	The conventional	The proposed	Reduction
Cross-correlation logic	116,391	80,595	30.8%
Packet detector	164,251	134,692	18.0%

IV. CONCLUSION

This paper proposes an efficient design method for high throughput cross-correlation logic using sign-bit reference waveforms. While computation resources in conventional ones are direct proportional to throughput rate due to the parallel structure, our approach mitigates such heavy resource usages by optimally sharing the first level adders in adder trees among sub-logics.

REFERENCES

[1] W. H. Wu, Y. W. Wu, and H. P. Ma, "A 480 Mbps MB-OFDM-based UWB baseband inner transceiver," *Proc. IEEE Asia Pacific Conf. Circuits Syst*., Hsinchu, pp. 164-167, Nov.-Dec. 2008.

[2] W. J. Lai, A. Wu, and W. Chen, "A systematic design approach to the band-tracking packet detector in OFDM-based ultrawideband systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3791-3806, Nov. 2007.

[3] S. J. Hwang, Y. Han, S. W. Kim, J. Park, and B. G. Min, "Resource Efficient Implementation of Lower Power MB-OFDM PHY Baseband Modem With Highly Parallel Architecture," *IEEE Trans. Very Large Scale Integration Syst.*, vol. 20, no. 7, pp. 1248-1261, Jul. 2012.

A Novel Readout Chip with Extendability for Multi-Channel EEG Measurement

Yi-Chung Chen, Chung-Han Tsai, Zong-Han Hsieh and Wai-Chi Fang, *Fellow*, *IEEE*
Department of Electronics Engineering and Institute of Electronics, National Chiao Tung University,
Hsinchu, R.O.C.

***Abstract*-- This paper proposes an extendable front-end readout chip (EFRC) for electroencephalography (EEG) measurements. An EFRC is developed for EEG measurement with features including low power consumption, a high signal-to-noise ratio, and highly efficient chip area usage. A chopper-stabilized differential difference amplifier (CHDDA) is used in the first stage to amplify signals and then during another adjustable amplification stage and filter are used to process biomedical signals. A 10-bit successive approximation register analog-to-digital converter (SAR-ADC) then links to the back-end for digital signal processing. In the last stage, shift-register pairs are used to transmit data to the next chip and receive data from the previous chip. The shift register design allows the number of channels to be extended. A TSMC 0.18 um CMOS process is used to design the EFRC and it operates with a 1.8 V supply voltage. The results shows that the total power consumption for the EFRC chip is approximately 80.268 uW and the chip area is approximately 944 x 863 um^2.**

I. INTRODUCTION

For biomedical research to advance, that analog front-end (AFE) IC features improve. Some studies have focused on designing entire AFE circuit. For example, [1] recorded electroencephalograph (EEG) signals through dry-contact electrodes using a micro-electrical-mechanical system, but power consumption exceeded 3mW. [2] proposed an eight-channel AFE chip for EEG, electrocardiography (ECG) and diffuse optical tomography (DOT). However, this system encountered real-time monitoring issues. A study indicated that EEG signal of approximately 40~50Hz [2], means that circuit delay cannot be more than 25 ms. In the method proposed by [2], implementing 64 channels would not be allow real-time monitoring because of multiplexer delay. Current EEG equipment is large, especially system with more than 16 channels. EEG systems rarely contain several channels in a chip.

Pin numbers create another difficulty for an EEG acquired chip with several channels. [3] showed that three electrodes are required for measurement. They are signal, reference, and common ground electrodes and this number cannot be reduced. Therefore, three pins are required for a channel. To perform back-end digital chip computing, an EEG signal must through an analog-to-digital converter (ADC). Even if an ADC is designed as a serial output, a pin is still required. A 64-channel EEG acquired chip contains 192 input pins and 64 output pins. In this design, shift registers can be used to reduce the number of output pins to one. Fewer pins mean less external noise and lower chip-to-chip connection complexity. (Fig.1)

This concept can also be used to measure other biomedical signals with certain amplifier magnification and cut-off frequency modifications.

II. SYSTEM ARCHITECTURE

Fig.2 shows the system architecture which includes a chopper-stabilized differential difference amplifier (CHDDA), low-pass filter (LPF), amplifier, shift-registers, and successive approximation register analog-to-digital converter (SAR-ADC). The system measurement data can be transmitted by the wireless communication module.

Acquiring the accurate biomedical signals is necessary, but fabrication mismatch affects the common mode rejection ratio (CMRR). The differential difference amplifier (DDA) can decrease this defect, because CMRR of the DDA is only related to the input port mismatch. Biomedical signals are weak and easily affected by external noise. Most biomedical signals are low frequency, and chopper circuit can transform low frequency noises to high frequency ones which can be filtered by the LPF. This can accurately detect biomedical signals.

The EFRC uses a 10-bit SAR-ADC [6] which consumes less power and uses space more efficiently. The SAR-ADC consists of four parts. A sample-and-hold (S/H) circuit, wherein the state of the transmission gate follows the voltage level acquired by the sample and holds circuit, and it is based on a cross-coupled charge pump mechanism. In addition, a clock comparator circuit is employed wherein the common mode range is expanded through a parallel configuration in order to reach a better signal-to-noise ratio (SNR).

To decrease power consumption, the charge-redistribution digital-to-analog converter uses a spilt capacitor array to avoid significant capacitor scaling.

The successive approximation registers uses two sets of D flip-flops that control the split capacitor array.

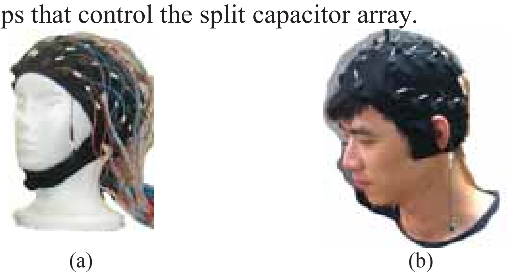

(a) (b)

Fig.1 EEG Hats (a) traditional design (separate outputs for each channel) (b) the novel design presented in this paper (one output)

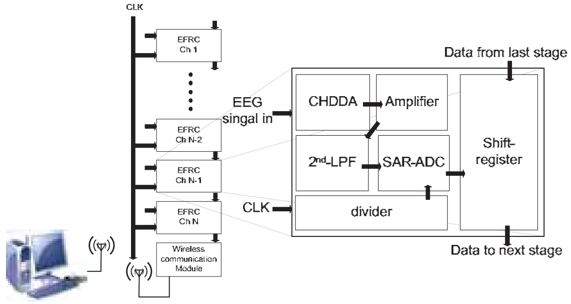


Fig. 2 System Block Diagram

III. RESULTS

The proposed EFRC acquires EEG signals with low power consumption. Noise is sufficiently low to ensure that the EEG signal is accurately acquired. The cascode architecture detects multiple channels and uses space efficiently.

Fig. 3 shows the chip layout which uses a one-channel EFRC. The proposed EFRC was fabricated using TSMC 0.18 um CMOS technology and the chip area was 944 x 863 um^2. Table I. summarizes the specifications of the proposed EFRC. The power consumption of the readout channel is approximately 60.286 uW with a 1.8 V supply voltage and 10 KHz frequency. The power consumption of the SAR-ADC is approximately 8.27 uW at a 100 KHz sampling rate and 1.8 V supply voltage. The total power consumption of the EFRC is 80.268 uW.

Table I.
EFRC SPECIFICATIONS AND COMPARISIONS

	This work	[2]	[7]
Supply Voltage	1.8V	1.8V	2.65V
Process Technology	0.18μm	0.18μm	0.35um
current	33.49u	39.54u	485u
channel	1	8	1
CMRR	74dB	75dB	x
PSRR+（dB）	115.41	113	x
PSRR-（dB）	107.84	105	x
Power Dissipation(uW)/CH	60.286	71.159	476
Area(mm 2)	0.8146	3.003	0.9261
extendable	yes	no	no

IV. CONCLUSION

This paper proposes a highly extendible EFRC. The IC features a low noise CHDDA and accurate cut-off frequency and is easy to use.

This paper develops a well-designed device which acquires and transfers multiple-channel EEG signals as accurately as traditional EEG equipment. This EFRC has the advantage of portability for convenient, real-time diagnoses.

ACKNOWLEDGMENT

This work was supported in part by the National Science Council of Taiwan, R.O.C., under grant NSC101-2220-E-009-049 and NSC101-2221-E-009-169-MY2. The authors would also like to express their sincere appreciation to the National Chip Implementation Center for chip fabrication and testing service.

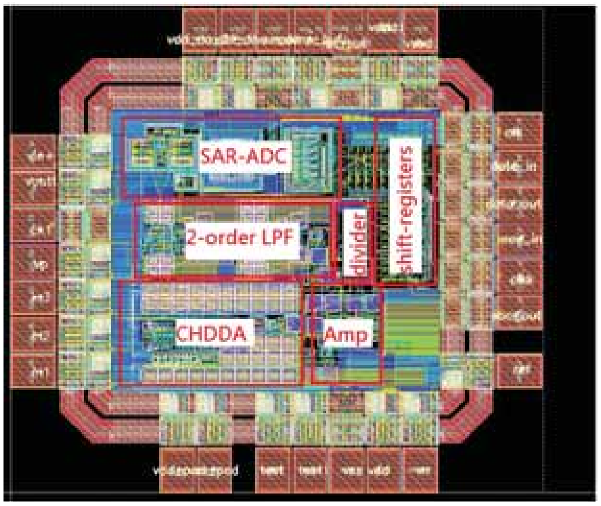


Fig.3 EFRC Layout

REFERENCES

[1] Sullivan, T. J., S. R. Deiss, Tzyy-Ping, Jung, Cauwenberghs, G, “A brain-machine interface using dry-contact, low-noise EEG sensors. Circuits and Systems,” 2008. ISCAS 2008. IEEE International Symposium on.

[2] Chung-Han Tsai; Zong-Han Hsieh; Wai-Chi Fang; , "A low-power low-noise CMOS analog front-end IC for portable brain-heart Monitoring applications," Life Science Systems and Applications Workshop (LiSSA), 2011 IEEE/NIH , vol., no., pp.43-46, 7-8 April 2011

[3] A. J ames Rowan & Eugene Tolunsky (2003). Primer of EEG with Mini-Atlas. US :Elsevier

[4] Harrison, R. R. and C. Charles (2003). "A low-power low-noise CMOS amplifier for neural recording applications." Solid-State Circuits, IEEE Journal of 38(6): 958-965.

[5] Chun-Chieh, H., H. Shao-Hang, Jen-Feng, Chung, Van, L. D., Chin-Teng, Lin. “Front-end amplifier of low-noise and tunable BW/gain for portable biomedical signal acquisition.” Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on.

[6] Y.-J. Chen, K.-T. Tang, W.-C. Fang, “An 8uW 100kS/s successive approximation ADC for biomedical applications,” in Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH, 2009, pp. 176-178.

[7] Tiong Lim; Yong Ping Xu; , "A low-power and low-offset CMOS front-end amplifier for portable EEG acquisition system," Biomedical Circuits and Systems, 2004 IEEE International Workshop on , vol., no., pp.S1/1-17-20,1-3Dec.2004doi: 10.1109/BIOCAS.2004.1454086

[8] Ng, K. A. and P. K. Chan. "A CMOS analog front-end IC for portable EEG/ECG monitoring applications." Circuits and Systems I: Regular Papers, IEEE Transactions on 52(11): 2335-2347.

GPGPU Implementation of an Improved Nonparametric Background Modeling for Moving Object Detection Strategies

Carlos Cuevas, Daniel Berjón, Francisco Morán, and Narciso García
Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, Spain

***Abstract*—A GPGPU-based real-time nonparametric modeling strategy for moving object detection is proposed. By dynamically estimating bandwidth matrices and by using a selective update mechanism, previous approaches are outperformed in quality while preserving their computational requirements.**

I. Introduction

Many computer vision tools have been recently developed to be used by the increasing number of consumer electronic devices endowed with camera [1]. To perform high level analysis tasks, these tools include moving object detection strategies [2]. Among these strategies, nonparametric modeling methods have shown to be those providing the best quality detections in a very large variety of scenarios [3]. However, the great computational cost of these strategies hinders their integration in tools requiring real-time processing [4]. To solve this limitation, algorithms proposing efficient implementations of nonparametric strategies have been recently developed [5]. However, to provide real-time detections, these algorithms carry out some simplifications that decrease the quality of the detections [3]: they use fixed kernel bandwidth matrices and blind update mechanisms.

In this paper, we propose a very efficient implementation of an improved nonparametric background modeling for moving object detection on a general-purpose graphical processing unit (GPGPU). This modeling, by dynamically estimating the bandwidth of the kernels and by using a selective update mechanism, improves the quality of previous approaches while maintaining the computational requirements. Therefore, the proposal can be perfectly integrated into the computer vision applications used by the latest generation of consumer electronic devices.

II. Nonparametric Moving Object Detection

Let p^n be a pixel in the current image, at time n, defined by a $(D+2)$-dimensional vector, $\mathbf{x}^n=((\mathbf{a}^n)^T,(\mathbf{s}^n)^T)^T$, where $\mathbf{a}^n$ is a vector containing appearance characteristics of p^n and $\mathbf{s}^n=(h^n,w^n)$ is a vector containing its coordinates. The probability of p^n to belong to the sequence foreground, ϕ, can be computed [5] as

$$\Pr(\phi \mid \mathbf{x}^n) = \frac{\Pr(\phi)p(\mathbf{x}^n \mid \phi)}{\Pr(\phi)p(\mathbf{x}^n \mid \phi) + \Pr(\beta)p(\mathbf{x}^n \mid \beta)},$$

where $p(\mathbf{x}^n|\beta)$ is the probability density function (pdf) that p^n belongs to the sequence background, β, $p(\mathbf{x}^n|\phi)$ is the pdf that p^n belongs to the foreground, and $\Pr(\beta)$ and $\Pr(\phi)$ are the background and foreground prior probabilities. $p(\mathbf{x}^n|\beta)$ is nonparametrically estimated from a set of N_β reference samples, $\{\mathbf{x}^i_\beta\}$, as in [5]. However, as here we focus on the background modeling, $p(\mathbf{x}^n|\phi)$ has been set as a constant pdf and the prior probabilities as $\Pr(\beta)= \Pr(\phi)=½$.

A. Background bandwidth estimation

To estimate the bandwidth matrices, $\mathbf{\Sigma}$, the strategy proposed in [6] has been taken as starting point. However, to maintain the computational requirements, instead of computing the median of thousands of distributions of sample differences, we propose to directly use their typical deviations. Then, the elements of $\mathbf{\Sigma}$ are obtained from weighted sums of typical deviations computed as

$$\sigma_\beta(h,w) = \left(\sum_{i\in\psi} \Pr\left(\beta \mid x^i_\beta\right)\right)^{-1/2} \left(\sum_{i\in\psi} \Pr\left(\beta \mid x^i_\beta\right)\left(\Delta a^i_\beta\right)^2\right)^{-1/2},$$

where $\Pr(\beta|\mathbf{x}^i_\beta)$=1- $\Pr(\phi|\mathbf{x}^i_\beta)$ is the probability of the reference sample $\mathbf{x}^i_\beta$, to belong to the sequence foreground and ψ is the set of background reference samples at (h,w).

B. Selective update mechanism

To achieve an additional improvement in the quality of the detections, in contrast to previous works [4], the background model is updated by using a novel and efficient selective mechanism. Then, the background pdf is estimated as

$$p\left(\mathbf{x}^n \mid \beta\right) = K_w \sum_{i=1}^{N_\beta} w_i \prod_{j=1}^{D+2} \exp\left(-\frac{\left(\mathbf{x}^n(j) - \mathbf{x}^i_\beta(j)\right)^2}{2\mathbf{\Sigma}(j,j)}\right),$$

where k_w is a normalization factor and w_i is a weighting factor, proportional to $\Pr(\beta| \mathbf{x}^i_\beta)$, that is obtained without additional computational effort.

III. GPGPU Implementation

In this algorithm, the classification of each pixel in a given input image is completely independent of that of its neighbors; this makes it an obvious choice for parallel implementation on a GPGPU, mapping an execution thread to each pixel. However, memory access operations on a GPGPU are slow compared to arithmetic operations and must be therefore minimized in order to maximize throughput.

Modern GPGPUs have several different tiers of memory, each with its own scope (global, block, or thread-local) and characteristics (read-only, cached, etc.) and, as it often happens in engineering, trade-offs must be made that result in the fastest tiers of memory being also the smallest or most restrictive in their usage. The host computer can only write the

This work has been partially supported by the Ministerio de Economía y Competitividad of the Spanish Government under the project TEC2010-20412 (Enhanced 3DTV).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

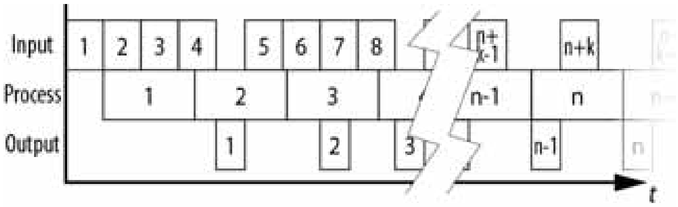


Fig. 1. Timeline (not to scale) showing the overlapping of input/output transfer operations (top and bottom rows) with processing; numbers indicate sequence number of each image. On the right side, after the break, all the buffers (size *k*) are full and the steady processing state has been reached. Priority is given to the output in order to reduce latency.

input images into the global memory, which is the largest and slowest of the memory tiers of the GPGPU. For each input pixel, the analogous pixel and its surroundings must be read from each reference image. This implies that every pixel from each reference image is input to several execution threads. Therefore, it is crucial to use a strategy so that each datum is read as often as possible from a fast memory tier rather than from the slower global memory.

In [5] an explicit cache management strategy is described: tiles of the reference images are copied into block-scope shared memory. However, due to strict memory access pattern requirements and/or linear hardware caching, this approach suffers from overfetching, which hinders its performance. Hardware texturing units use a different strategy: they map cache lines to 2D regions in a texture using space-filling curves such as Z-order or Hilbert [7], so that a high cache hit rate is achieved when execution threads in the same execution group (warp) read from locations that are close to each other in 2D. For our algorithm, this means that we want to pack the threads in each warp (32 threads/warp in our GPGPU) as compactly as possible, in sub-blocks of 8x4 pixels. Recent advances in GPGPU hardware and software have allowed us to use the hardware texturing units not only for reading the reference data but also to write all the intermediate data as textures, resulting in a significant speedup.

Finally, three different host threads manage input, processing, and output operations. This allows the GPGPU to simultaneously perform memory transfers and data processing (see Fig. 1), yielding a higher utilization of the GPGPU and an increased end-to-end throughput.

IV. RESULTS

Fig. 2 displays some images comparing the quality of the detections provided by the proposed background modeling with those obtained in [3] and [5]. Thanks to the proposed improvements, the amount of false detections decreases while the moving objects are better detected.

Table I presents the achieved mean processing times for sequences with different spatial resolutions. These data prove that despite the proposed improvements in the background modeling, with the proposed GPGPU-based implementation we barely increase the computational cost corresponding to previous implementations [5] of the simplest spatio-temporal background modeling [3].

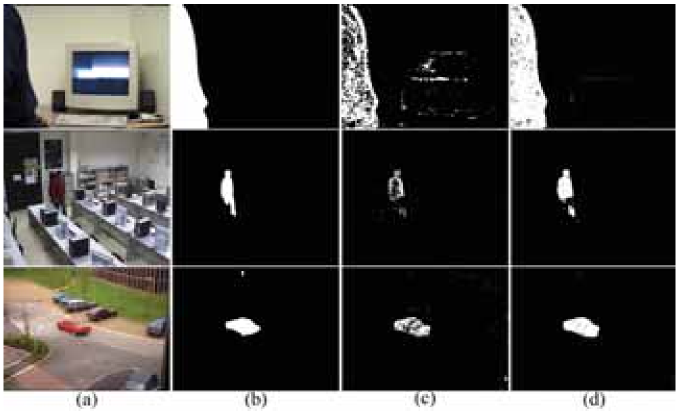


Fig. 2. (a) Original images. (b) Ground truth. (c) Detections with the background modeling in [3] [5]. (d) Detections with the proposed strategy.

TABLE I
MEAN PROCESSING TIMES FOR DIFFERENT IMAGE RESOLUTIONS

Image resolution (*H*×*W*)	128×160	288×352	576×768
Background modeling in [5]	9 ms	38 ms	145 ms
Proposed background modeling	11 ms	42 ms	152 ms

V. CONCLUSIONS

A very efficient GPGPU-based implementation of an improved spatio-temporal nonparametric background modeling for moving object detection has been proposed. An innovative dynamical kernel bandwidth estimation method and a novel background selective update strategy significantly increase the quality of the detections provided by previous strategies. The full utilization of the GPGPU resources, based on hardware texturing units and overlapping memory transfers and data processing, allows the implementation of all these improvements while retaining the real-time performance required by the computer vision tools used by last generation of consumer electronic devices endowed with camera.

REFERENCES

[1] G. Hua, Y. Fu, M. Turk, M. Pollefeys, and Z. Zhang, "Introduction to the special issue on mobile vision,*" Int. Jour. Computer Vision*, vol. 96, pp. 277-279, 2012.

[2] E. Komagal, A. Vinodhini, A. Srinivasan, and B. Ekava, “Real time background subtraction techniques for detection of moving objects in video surveillance system,” *IEEE Int. Conf. Communication and Applications*, pp. 1-5, 2012.

[3] Y. Sheikh and M. Shah, “Bayesian modeling of dynamic scenes for object detection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1778-1792, 2005.

[4] S. Y. Elhabian, K. M. El-Sayed, and S. H. Ahmed, “Moving object detection in spatial domain using background removal techniques – state-of-art,” *Recent Patents on Computer Science*, vol.1, no. 1, pp. 32-54, 2008.

[5] C. Cuevas, D. Berjón, F. Morán, and N. García, “Moving object detection for real-time augmented reality applications in a GPGPU,” *IEEE Trans. Consumer Electronics*, vol. 58, no. 1, pp. 117-125, 2012.

[6] C. Cuevas, R. Mohedano, and N. García, “Kernel bandwidth estimation for moving object detection in non-stabilized cameras,” *Optical Engineering*, vol. 51, no. 4, 3 pages, 2012.

[7] M. Doggett, “Texture Caches,” *IEEE Micro*, vol. 32, no. 3, pp. 136-141, 2012.

The Acceleration of Various Multimedia Applications on Reconfigurable Processor

Minwook Ahn, Donghoon Yoo, Soojung Ryu, Jeongwook Kim
Samsung Advanced Institute of Technology

Abstract— **Modern consumer electronics require efficient and versatile processors as their heart for various functions. This paper shows how our reconfigurable processor (RP) accelerates various multimedia applications even with different features. The RP provides an acceleration mode powered by modulo scheduling algorithm and intrinsic specialized to each application. By exploiting these two, RP can enhance the performance up to 71.54x for audio, video, image signal processing (ISP), 3D graphics and major communication channel applications (DVBT2).**

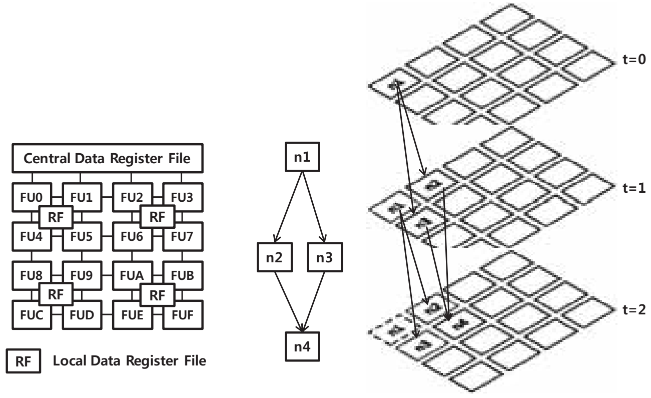
I. INTRODUCTION

Modern consumer electronics provide the capability to process various functions. For example, with cellar phones, we can not only call but also enjoy many applications such as listening to music, watching and recoding a video, playing interesting games, taking a photograph, and etc. Their application processors should be able to do those functions. This versatility had been based on the complex ASICs specialized to each application. Now as the complexity of the target applications grows, processor based solutions begin to take the places of ASICs due to their programmability in spite of its inherent overhead compared to the ASIC solutions. In this paper, we show how our reconfigurable processor accelerates various multimedia applications even with inherently different features. In order to maximize the performance of the processor based solutions, the performance gap between ASIC and processor should be minimized. To make the processor based solutions be comparable to ASIC solutions, there are two major obstacles to be overcome. (1) What is a good method to do the repeated tasks like in ASIC? (2) How can the application specific tasks be accelerated well like ASIC? To address the first, RP has an acceleration mode, called Coarse Grain Reconfigurable Architecture (CGRA) mode, for the repeated tasks in the target applications. Loops, which are the repeated tasks in the target applications, can be mapped onto the array of processing elements (PEs) consisting of the functional units (FUs) and registers. They are connected by dedicated wires for fast communication. The loops mapped onto CGRA can be accelerated by pipelining the different iterations of the loops on the array. To address the second, RP supports many application specific instructions (ASIs) in the form of intrinsic. An intrinsic is a special instruction as it is used just like a function call in the application but directly translated into the corresponding instruction by a compiler. With ASIs, many cycles can be saved otherwise consumed by the corresponding many basic instructions such as arithmetic, logical and memory instructions.

--

II. CGRA: ACCELERATION MODE FOR THE REPEATED TASKS

As briefly noted in Section 1, CGRA mode accelerates the loops in the applications. In CGRA mode, the array of multiple PEs is used as shown in Figure 1(a). The array consists of 16 FUs and 5 register files (RFs). Each RF is shared by the surrounding four functional units.



(a) Array of Processing Elements (b) Data Flow Graph of a loop (c) Execution in CGA mode

Figure 1 An example explaining CGA mode

If an application developer wants to accelerate a loop having the data flow graph (DFG) shown in Figure 1(b) in CGRA mode, the loop should be mapped onto the array of PEs. This mapping is automatically done by our compiler that implements a modulo scheduling algorithm [1]. A *kernel* of a loop, which is the repeated part of a schedule of a loop generated by the modulo scheduling algorithm, is stored in the code memory. In the case of the DFG in Figure 1(b), the kernel of the DFG is generated and excuted as shown in Figure 1(c).

III. INTRINSIC INSTRUCTION ACCELERATING THE APPLICATION SPECIFIC TASKS

RP supports various ASIs. For example, we can think of a function doing ‘shift and saturate’. Even though this function is not typically supported in the general programming languages like C as an operator, it is widely used in many multimedia applications. Executing the function requires several basic operations. However, the same function can be implemented as a single ASI with a unit latency without much effort on hardware implementation.

Generally the ASIs are designed from the commonly used patterns or sequence of the basic instructions in the target applications. Among the candidates for the ASIs, the processor architects choose the ones giving the maximum performance increase without much hardware cost such as the increase of

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

gate counts and critical path length. Our compiler helps this process by providing an easy way for evaluating the effect from using such ASIs. Even before there is a real HW implementation of such ASI, the core architecture simulator can show the statistics related to the cycle counts of the processor core. It is possible if there is the emulation code of the ASI implemented by c language and estimated clock cycles for the ASI. This information is used for generating the processor core simulator of RP as shown in Figure 2. Likewise, the processor architects easily plug-in and plug-out their ASIs and evaluate them within a short time. As a result, various ASIs are designed within a short time. Some ASIs like 'shift and saturate' have a unit latency and they are commonly used in almost all multimedia application domains. On the other hand, some ASIs have long latencies and they are used only in some application domains. As an example, we can think an ASI for an application in ISP. The ISP application requires a lot of calculations especially related to trigonometry. Those calculations generally consume more than several tens of cycles and frequently used in only the ISP application domain. The application developers can find a common pattern among the trigonometric calculations and make it an ASI for acceleration. Unfortunately, this kind of ASIs is specific to an application domain. In the case of the trigonometric operation explained above, it is not used in other application domains like audio and video.

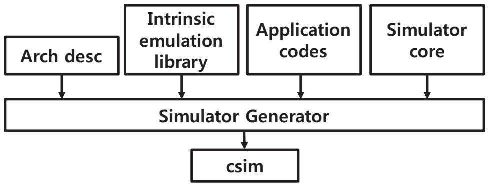


Figure 2 Simulation framework diagram for SRP

As explained, in order to find a tradeoff between the programmability and the direct use of the ASI in the form of assembly for performance optimization, all ASIs are used in the form of *intrinsic*. An intrinsic is the same as the function call in the application developer's perspective except there is no function implementation of the intrinsic. Our compiler identifies which function calls are intrinsic. Our compiler directly translates it into an ASI. As the intrinsic is used as the same as the function call, there can be the codes for managing function call such as passing arguments and retrieving a return value. These codes can be eliminated by compiler optimizations such as common subexpression elimination, copy propagation, dead code elimination, and so on.

IV. Experimental results

In this section, we show how much RP can enhance the performance of the target applications by the use of intrinsic and CGRA mode. The experimental results are shown in Figure 3. Our evaluation is done using five different multimedia application domains: audio(ac3), video(h.264 baseline profile), ISP (2D panorama / alignment), 3D graphics (pixel shader) and channel (DVB-T2). We measure the performance of the applications in three cases. *Baseline* means the one without the use of intrinsic and CGRA mode. In this case, only two issue VLIW processor works. *Baseline+intr* means the applications with intrinsic but without the use of CGRA mode. *Baseline+intr+CGRA* means the one with both intrinsic and CGRA mode. The performances of *baseline+intr* and *baseline+intr+CGRA* are normalized to that of baseline. When only intrinsics are exploited, 6.39x speed-up is gained on average. In case of the channel application, its performance can be increased by 24.59 times as the channel application uses lots of intrinsics similar to Single Instruction Multiple Data (SIMD). This ASI has great impact on the performance of the channel application. The further use of CGRA mode can increase the performance more to 15.79x on average. Especially, in the case of a pixel shader from 3D graphics application, the performance is increased by 71.54x times when both the intrinsic and CGRA mode are exploited. It is because a shader in the 3D graphics domain consists of a single large loop with much parallelism. On the other hand, CGRA mode cannot increase the performance much in the channel application. It is as there are not many loops which can be mapped onto the array of PEs.

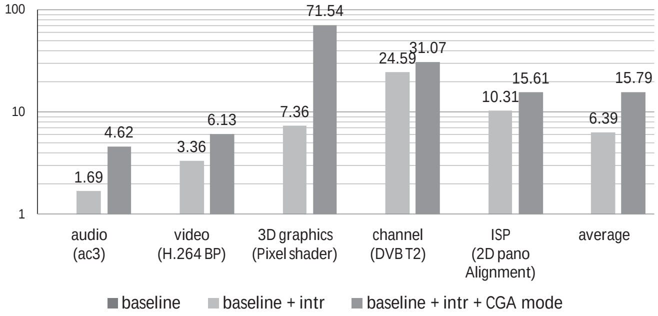


Figure 3. Speed ratio compared to baseline case, using CGA mode and intrinsic instructions

Conclusion

RP can minimize the performance gap between ASIC and processor based solution by exploiting CGA mode and intrinsic instructions. In average, the performance of the applications is increased by 15.79x compared to the baseline without the use of intrinsic and CGA mode.

Acknowledgment

We would like to thank all ACT team members: Heejin Ahn, Taisong Kim, Wonsob Kim, Jinseok Lee, Choonki Jang, Seokjoong Hwang, Junkyoung Kim and Seonggun Kim.

References

[1] Taewook Oh, Bernhard Egger, Hyunchul Park, Scott A. Mahlke, Recurrence cycle aware modulo scheduling for coarse-grained reconfigurable architectures, in the Proceedings of the 2009 ACM SIGPLAN/SIGBED Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES '09), pp 21-30, Dublin, Ireland, June 2009

Real-time Realistic 3D Facial Expression Cloning for Smart TV

Jung-Bae Kim, Youngkyoo Hwang, Won-Chul Bang, Heesae Lee, James D.K. Kim, and ChangYeong Kim

***Abstract*--This paper suggests a novel technology that can clone a user's facial expression to his avatar in 3D virtual worlds realistically using only one color camera on a smart TV. To do this, the user's 3D head movement information and 3D position of facial feature points are needed in real time. We propose two novel approaches to achieve this. Firstly, we use a personalized 3D and 2D facial expression model to deal with head movement and various expressions. Secondly, we use a facial muscle model to generate natural motion of facial feature points located in cheeks and forehead which are difficult to be tracked using a camera. Experimental results demonstrate that the proposed method would be an efficient technique to perform realistic 3D facial expression cloning.**

I. Introduction

When a user's facial movement is mimicked by a virtual avatar in 3D within a smart TV, it gives him not only interest but also a sense of immersion that he has entered and lives in a virtual world. There are two representative methods in this technology. One is based on the motion capture device called mocap-based system; the other is based on a single camera, called a vision-based system.

The mocap-based system attaches almost one hundred markers on the user's face and uses more than seven well-calibrated IR cameras. While the system can track subtle 3D expression, it needs post-processing to reduce internal noise. As a result, it is not easy for general users to use because of several problems: inability to perform real-time processing, large system size, cumbersome markers, high cost, etc.

The vision-based system would be desirable for most users at home since a camera device is cheap and easy to use, requires no markers on the face, and is simple to set up. Unfortunately, it is very challenging to enable a vision-based system to capture a user's 3D expression using a monocular camera, and to track subtle facial movement without marker.

In this paper, we propose a novel vision-based expression cloning system having two major features: 1) cloning 3D facial expression in real time and 2) cloning a subtle movement on cheeks and forehead that traditional vision-based system cannot achieve [1].

To achieve the first feature, we use a personalized 3D and 2D facial expression model which can track head motion and facial expression together. And we use a facial muscle model to generate natural motion of the cheeks and forehead. An overview diagram is shown in Fig. 1. We will explain two methods in detail in section II and III. The experimental results are explained in section IV.

The research in this paper was performed by the Advanced Media Lab in SAIT, Samsung Electronics Co.

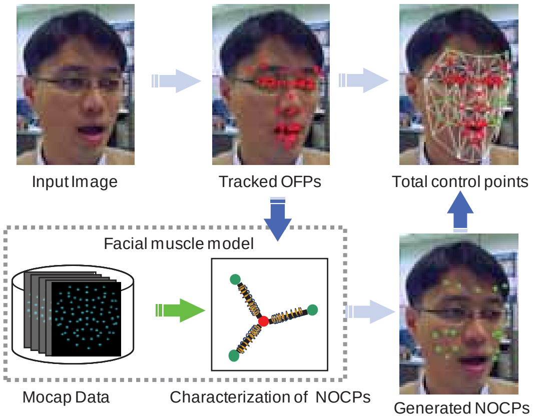


Fig. 1. Overview diagram of the real-time realistic 3D facial expression tracking system. Outstanding Control points (OCPs), shown in red, are tracked from computer vision techniques. The behavior of Non-Outstanding Control points (NOCPs) is modeled as facial muscle model using Mocap data offline. The movement of NOCPs is generated in real-time based on facial muscle model parameters and position of OCPs.

II. Tracking of Outstanding Control Point's Motion

Our system is able to track 56 outstanding control points (OCPs) and 3D head motion with six degrees of freedom (three rotational and three translational) under the condition of using only one camera. We exploit the 3D face model Candide-3 [2], a well-defined parameterized face model, to track and obtain 3D locations of pre-defined expression control points. This 3D face model is deformable to match face shapes of different people as well as varying facial expressions of a specific person. We defined key points for facial animation as control points to be tracked. The 3D face model is linearly parameterized and is written as:

$$\mathbf{F} = \overline{\mathbf{F}} + \boldsymbol{\tau}_s \cdot \mathbf{S} + \boldsymbol{\varepsilon} \cdot \mathbf{E} \qquad (1)$$

where $\overline{\mathbf{F}}$ is the standard shape of the model, $\mathbf{S}$ is the shape basis, $\mathbf{E}$ is an expression basis, $\boldsymbol{\tau}_s$ is a vector of shape parameters, and $\boldsymbol{\varepsilon}$ is a vector of expression parameters. A shape basis and its parameters account for different shapes among human faces, while an expression basis and its parameters account for different expressions for a specific person.

Our proposed system automatically localizes control points by fitting the 3D face model on the user's face image. To meet the requirement of inter-person variability, the AAM (Active Appearance Models) method is exploited [5]. We have collected more than 5,000 faces and made a generic model. In order to provide robustness to pose variation, we have brought the concept of 2D+3D AAM which is the key point localization technique. It imposes 3D shape fitting constraints on the conventional 2D AAM. After we have localized key

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

points from 2D+3D AAM, we relocate each point along strong edges [3] to overcome the lack of robustness to illumination variation present in the original AAM. Thus, the system is very fast and robust to in the presence of eyeglasses and localizes feature points very accurately. In addition, it is possible to accurately align a personalized 3D model on the face image.

In order to obtain 3D locations of outstanding control points, we bring Dornaika's approach [4]. We have modulated the expression bases as follows: 1) jaw dropper, 2) lip stretcher, 3) lip corner depressor, 4) upper lip raiser, 5) nose wrinkler, 6) left eyebrow raiser, 7) right eyebrow raiser, 8) eyebrows lower, and 9) eye closer. These expression bases are enough to cover a user's common facial expressions.

A sequential Monte Carlo algorithm [6] is used to track the 3D head pose in noisy environment. It estimates the current pose of 3D face model by considering the correspondence of the control points' location on the 3D face model between the previous image frame and the current image frame.

III. Generation of Non-Outstanding Control Points' Motion

In order to achieve more realistic facial animation, the system should include supplementary facial expression control points in addition to the aforementioned OCPs. These additional control points, called non-outstanding control points (NOCPs), are located on the cheeks and forehead. We suggest a new method to generate motion in these points by activation from neighboring points. The neighboring points are tracked by the vision-based method suggested in the previous section. The tracked points have a conspicuous edge and are located over eyebrows, eyes, nose, mouth and chin. Our muscle model is defined by stiffness between adjacent points [7]. The stiffness values are determined by training from Mocap data offline. The model adds 17 points on the cheeks and forehead. Consequently, the entire system can create motion in 73 facial points. The muscle model is shown in Fig. 2.

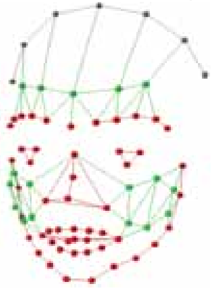


Fig. 2. Muscle model: 75 points used for training from Mocap data. Red points are OCPs. Grey points are static points. Green points are NOCPs whose motions are generated by neighboring points and their links.

IV. Experimental Results

To test the system's performance, we collected videos of six people that include their facial expressions and head motions. The facial expressions are composed of movements in 11 different facial components that can be corresponded to our expression parameters. The movements include raising left or right eyebrow, lowering both eyebrows, wrinkling nose, raising upper lips, pouting, stretching lip, raising/lowering lip corner, opening mouth, and blinking both eyes. The videos are composed of 2,250 frames. Our whole system can track control points on normal facial expressions with changes in head pose. Points on the forehead move realistically according to the movement of eyebrows, and points on cheek move realistically according to the movement of nose and mouth.

As a result, our tracking system has a 96.96% of success rate. It takes just 36ms/frame on 900Mhz CPU which should be suitable for real-time processing on a smart TV. One of the results is shown in Fig. 3.

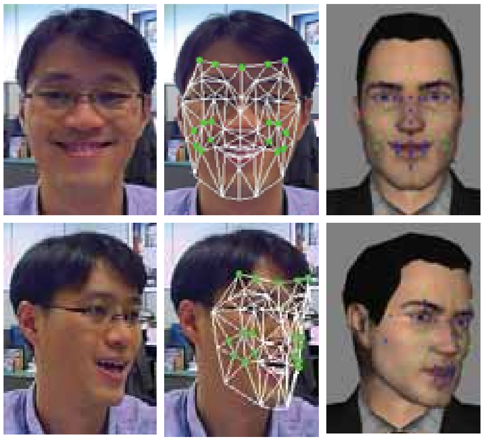


Fig. 3. Experimental results. Blue dots are the points tracked in real-time by vision method (OCPs), and green dots are the points generated from muscle model (NOCPs). The results show that the green dots (NOCPs) respond well to facial movements.

V. Conclusion

In this paper, we proposed a novel method to track a user's facial expression for 3D face animation with only one camera. Since the system tracks outstanding points on eyebrows, eye, nose, mouth and chin in real-time, it can show the user's genuine expression immediately. Furthermore, since the system generates 3D motion of non-outstanding points on the cheeks and forehead by using a muscle model trained by Mocap DB, the system can show more realistic and detailed expressions. We expect that this system will be very useful for smart TVs to show subtle movement on an avatar's face which will impress and interest a user.

Reference

[1] J.X. Chai, J. Xiao, and J. Hodgins, "Vision-based control of 3D facial animation," SIGGRAPH/Eurographics Symposium on Computer Animation (2003)

[2] J. Ahlberg, "Candide-3 – an updated parameterized face," http://www.icg.isy.liu.se/candide/ (2001)

[3] I. Matthews and S. Baker, "Active appearance models revisited," IJCV, vol.60, no.2, pp.135-164 (2004)

[4] Fadi Dornaika and Franck Davoine, "On appearance based face and facial action tracking," IEEE Trans. on Circuits and Systems for Video Technology, vol.16, pp.1107-1124 (2006)

[5] T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham, "Active shape models–their training and application," CVIU, pp.38-59 (1995)

[6] A. Doucet, N. De Freitas and N.J. Gordon, Sequential Monte Carlo Methods in Practice (2001)

[7] X. Provot, "Deformation constraints in a mass-spring model to describe rigid cloth behavior," Proc. of Graphics Interface, pp.147-154 (1995)

Discovering Unusual Behavior Patterns from Motion Data

Kai-Lin Pang, Guan-Hong Chen, and Wei-Guang Teng*, *member, IEEE*
Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan

***Abstract*-- As there are more and more surveillance cameras installed in public places, a challenging problem is to discover unusual behavior patterns from a huge amount of video data. However, this task is currently only feasible for human beings because both object recognition and intention detection are still difficult for computer vision. Recently, the development of low-cost depth cameras significantly improves the efficiency and effectiveness of capturing motion data. We thus propose in this work an algorithmic scheme that extracts unusual behavior patterns from motion capture data. Specifically, feature extraction and data clustering techniques are applied in our scheme so as to detect such outlier patterns. Example applications of our scheme include public area surveillance and home healthcare.**

I. Introduction

Research works in 3D image acquisition have proposed two well-known techniques, i.e., stereo vision (SV) and Time-of-Flight (TOF), in past decades. Nevertheless, specific devices of high price but poor quality, e.g., laser scanners, are usually required to realize these techniques. In 2010, the invention of the low-cost Microsoft Kinect sensor, high-resolution visual and depth (RGB-D) sensing [3] has become available. This Kinect sensor also opens up new opportunities to solve fundamental problems in computer vision such as detecting and identifying objects/humans in real-world situations.

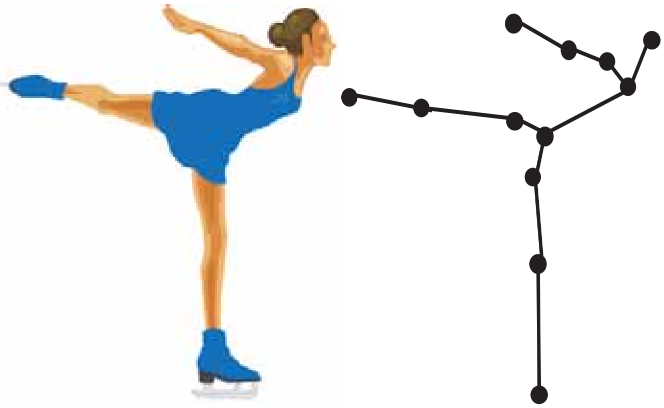


Fig. 1. (a) An example posture; and (b) the corresponding skeleton model

In general, the additional depth information is the key difference between video data and motion data. With depth cameras, motion capture denotes the process of recording object movements and translating the movements onto a digital model [4]. As shown in Fig. 1, human motion is commonly modeled using a kinematic chain, which may be regarded as a simplified copy of the human skeleton. A human skeleton model is composed of many major joints, i.e., elbow, knee, ankle and so on. Using motion capture techniques, an actor's motion can be derived to be a time-dependent sequence of 3D joint coordinates as well as joint angles with respect to some fixed kinematic chain [7][8].

One's usual behavior such as gait and rate of walking can be extracted from the history of his or her motion data. For thousands of years, people often use body characteristics such as face, height, body shape, and gait to recognize each other [5]. In other words, biometrics can be used to facilitate the purposes of user identification and unusual behavior detection [9]. Note that the premise of unusual behavior detection is to effectively conduct human motion recognition [1]. In prior works, there are generally two ways to recognize unusual human motion behavior. One approach is to build a unified model to represent all normal modes of human behavior in an observed scene [2]. In contrast, the other approach is to build a model which represents all abnormal modes [1]. Nevertheless, a common limitation of these two approaches is that all modes of (normal or abnormal) human behavior have to be predefined. This limitation may greatly reduce the feasibility of these approaches when applied in practical applications.

By properly modeling the problem of behavior detection as an unsupervised learning problem in this work, data clustering techniques can then be utilized to separate unusual patterns from usual ones. Specifically, unusual patterns may form outliers that are quite different from or inconsistent with the remaining set of data [6]. In other words, an outlier, or outlying behavior pattern, is one that appears to deviate markedly from other observations of the sample in which it occurs.

II. Usage Scenarios

In this work, we use Kinect to capture the body motion of a user. As shown in Fig. 2, when Kinect is installed in the example scenario of a train/airplane cabin, behavior patterns such as people sitting on their seats or walking slowly may occur frequently. However, patterns of running and falling down on the floor occur very infrequently that are considered as unusual behavior patterns in this case. This usage scenario can be also applied to other environments such as monitoring in a factory and healthcare in home places. In summary, a key observation is that usual patterns occur frequently. Thus, it is not necessary to define usual patterns before the identification process.

The authors are supported in part by the National Science Council, Project No. NSC 100-2221-E-006-261-, Taiwan, R.O.C.

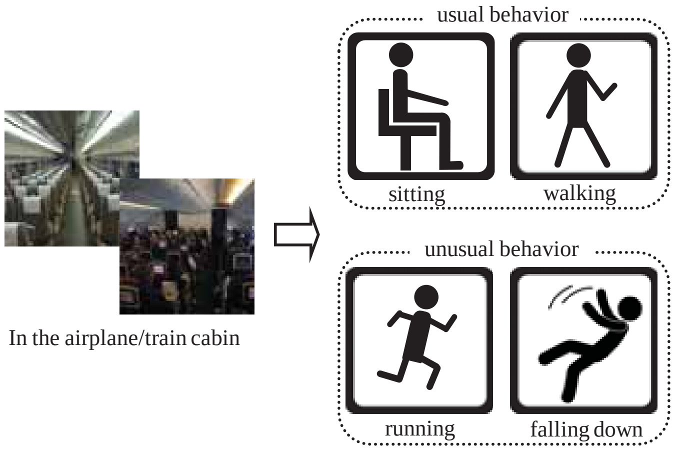


Fig. 2. An example scenario showing that some behavior patterns are usual whereas some others are unusual.

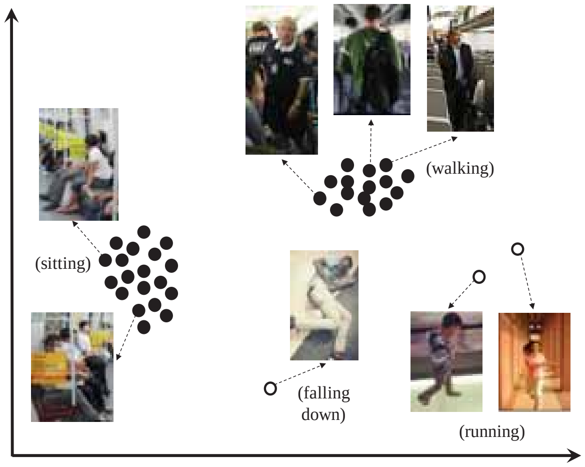


Fig. 3. Outliers (i.e., white nodes) detected after conducting the data clustering technique are possibly unusual behavior patterns.

III. The Algorithms

In the empirical studies, our algorithmic scheme is trained and evaluated using the Carnegie Mellon University Motion Capture Database (http://mocap.cs.cmu.edu) and the database provided by Motion Capture Lab of the Ohio State University (http://accad.osu.edu/research/mocap/mocap_data.htm). These datasets contain time-dependent sequences of 3D joint coordinates and several well-defined motion categories that facilitate the development process of our proposed scheme. All datasets were transformed to the quaternion representation without the information of the bone length. Also, the translation and orientation information of the root joint is discarded.

The flow of our scheme is illustrated as follows. With the skeleton model as shown in Fig. 1, all motion capture data are represented as 3D coordinates. Data segmentation and feature extraction techniques are then applied to map significant motion segments into the feature space. When conducting the step of feature selection, features that have high correlation are combined together so as to reduce the dimensionality and to retain the independence of selected features. Note that some features are meaningful, e.g., 3D coordinates, motion speed, body movement range, whereas many others are not interpretable. Without loss of generality, a well-known clustering technique *k*-means is then utilized to cluster these motion segments and obtain motion clusters. Then, the resulting outliers may correspond to unusual behavior patterns.

By following the example usage scenario in Fig. 2, an illustrative result is depicted in Fig. 3. Two major clusters, i.e., sitting and walking segments, are found in the clustering result. Also, a few outliers, i.e., white nodes in Fig. 3, that stand for the motion segments of running and falling down are identified. Note that a feature space may include more than a few dimensions. Also, not all resulting clusters can be interpreted as meaningful motion patterns when applying our scheme in a complicated environment.

IV. Conclusions

The development of low-cost depth cameras has significantly improved the efficiency and effectiveness of capturing motion data. In this work, we have presented an algorithmic scheme that extracts unusual behavior patterns from motion capture data. Feature extraction and data clustering techniques are applied to detect such outlier patterns for possibly practical applications including public area surveillance and home healthcare.

References

[1] H. Bao, Y. Shi, and B. Xu, “Video Based Abnormal Behavior Detection,” *Proceedings of the 2011 International Conference on Innovative Computing and Cloud Computing*, Wuhan, China, pp. 32–35, August 2011.

[2] P. Dickinson and A. Hunter, “Using Inactivity to Detect Unusual Behavior,” *Proceedings of the IEEE Workshop on Motion and Video Computing*, Copper Mountain, CO, USA, pp. 1–6, January 2008.

[3] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, “RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments,” *Proceedings of the 12th International Symposium on Experimental Robotics*, Delhi, India, December 2010.

[4] Y. Hu, S. Wu, S. Xia, J. Fu and W. Chen, “Motion Track: Visualizing Variations of Human Motion Data,” *Proceedings of the 3rd IEEE Pacific Visualization Symposium*, Taipei, Taiwan, pp. 153–160, March 2010.

[5] A. K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, January 2004.

[6] J.-G. Lee, J. Han and X. Li, “Trajectory Outlier Detection: A Partition-and-Detect Framework,” *Proceedings of International Conference on Data Engineering*, Cancun, Mexico, pp. 140–149, April 2008.

[7] M. Müller, T. Röder and M. Clausen, “Efficient Content-Based Retrieval of Motion Capture Data,” *ACM Transactions on Graphics*, 24(3):677–685, July 2005.

[8] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-Time Human Pose Recognition in Parts from Single Depth Images,” *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, USA, pp. 1297–1304, June 2011.

[9] T.-H. Yu and Y.-S. Moon, “Unsupervised Real-Time Unusual Behavior Detection for Biometric-Assisted Visual Surveillance,” *Proceedings of the Third International Conference on Advances in Biometrics*, Alghero, Italy, pp. 1019–1029, June 2009.

Sentiment Diffusion in Large Scale Social Networks

[1]Jie TANG and [2]Acm FONG
[1]Tsinghua University, China
[2]Auckland University of Technology, New Zealand

Abstract—**Popularity of online social networks provides the chance to make sentiment analysis on every user instead of every document or sentence. And relations between users on social media sites often indicate correlation (negation) between users' opinions. In this work, we study how user's opinion spread in social networks. We employ the data from Tencent.com, the largest social network of China to empirically study the problem. Our work focuses on six different topics including policy, products, brand, sports, movie and politician. We study the distributions of peoples' opinions on different topics and how users' opinions are influenced by those he is following. We propose a graphical model to capture the essence of social network as well as an algorithm to perform semi-supervised learning. The learning algorithm can be used to accurately predict users' sentiment in the social network.**

I. INTRODUCTION

The success of many online social networks has empowered users to actively interact with each other and freely publish user-generated content online. Analyzing the user-generated content thus becomes a fundamental issue to understand the social Web, e.g., how do people think about a new product? And how does a user's sentiment spread in the social network?

Indeed, people nowadays are strongly influenced by social users' opinions. For example, users usually want to first refer to the others' comments, before purchasing a product, booking flights, and selecting restaurants. However, with the extremely rapid increase of the user-generated content, it is really difficult for the user to digest the large amount of information.

Recently, quite a few work has been conducted for sentiment analysis, which aims to classify the polarity (e.g., positive or negative) of each user's comment. Such as, Turney et al. [5] employ the search engine to determine the polarity of each word, and accumulate them together to classify the polarity of the whole sentence. Pang et al. [3] study the performance of different machine learning methods and various features for sentiment analysis. Their results show that machine learning methods outperform the basic lexicon-based methods. However, most existing methods ignore one underlying factor: in many applications, the polarity should be on the user instead of a single document. In fact, the quality of a single document (e.g., a product review) may vary largely, even for the same user. Some work tries to evaluate the quality of each user's comment. Lu et al. [2] exploit social context for predicting the quality of user's review. However, they do not consider how users' sentiment spread in the social network.

Now, the problem is how to model users' sentiment and its diffusion. The problem is non-trivial and poses a set of unique challenges. First, how to simultaneously capture the content and the social network information? The user's sentiment may be hidden in her social context (social networks and published tweets). Second, the labeled data for training a sentiment prediction model is very limited. On Twitter, there are 50 million new tweets every single day. It is impractical and also impossible to annotate sufficiently labeled data to train an accurate prediction model. Finally, users' sentiment may quickly spread in the social network. It is important to develop an efficient algorithm to capture the diffusion of sentiment.

In this study, we try to conduct a systematic investigation on this problem of sentiment diffusion in large networks. We formulate the problem as a semi-supervised learning problem and propose a probabilistic factor graph model, named SSFG, to solve this problem. Specifically, each user's polarity and social correlation between users are modeled together as a joint probability. An efficient algorithm based on Metropolis-Hasting is presented to optimize the factor graph model. The algorithm supports both supervised learning and semi-supervised learning, thus the proposed model is able to learn a sentiment prediction model with only a small number of labeled data and a large number of unlabeled data. We validate the proposed model on a data set from Tencent Weibo[2]. Experimental results show that our method can clearly improve the prediction accuracy (+3-10%) over the baseline methods.

II. RELATED WORK

There have been extensive work focusing on document-level or sentence-level sentiment analysis. There approaches include rule-based method. Pang et al. [3] study the performance of different features with different learning methods and show the performance of SVM is similar to that of Maximum Entropy Classification, better than that of Naive Bayes. Turney et al. [5] employ an interesting method which makes use of the search engine to determine the polarity of each word. Some researchers also consider employing networks to help, including the min-cut framework [6], and the network between participants [7]. However, most existing studies do not consider the diffusion of users' sentiment.

With the growth of online social network, more and more researches have been devoted to social network analysis. Crandall [8] identifies both influence and homophily effects. Leskovec et al. analyse the positive and negative relationships in social network [9]. Tang et al. employ factor graph models to analyse influence between users in different topic level [10]. Lu et al. [2] study the problem of determining review quality using social context information. They propose a semi-

[2] Tencent is the largest social network in China. Tencent Weibo is a microblogging system, similar to Twitter.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

supervised approach based on author-consistency, trust consistency, co-citation consistency and link consistency. Tan et al. study the problem of user-level sentiment analysis using social networks [11].

III. OUR APPROACH

In this section, we will describe the proposed model and the semi-supervised learning algorithm for the model.

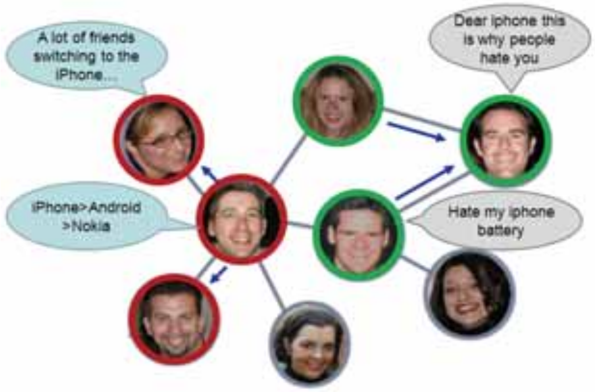

A lot of friends switching to the iPhone...
Dear iphone this is why people hate you
iPhone>Android >Nokia
Hate my iphone battery

Figure 1. Example of sentiment diffusion for "iphone". Red circle indicates positive sentiment, while green one indicates negative. Blue arrows indicate the diffusion of sentiment.

The essence of our model is to make use of the network properties to help classification. In our problem definition, we have two kinds of unlabeled data to predict, namely, unlabeled users and unlabeled tweets. As we mentioned in the introduction part, the label of users is much more important for potential application. Thus, we propose to use the state-of-art method to predict the label of unlabeled tweets (multiclass-SVM) and then design a graphical model based on the prediction results on unlabeled tweets. Using multiclass-SVM involves handling the imbalanced data, which we will give details in the experiment section.

With similar idea of condition random field, we employ the assumption that the influence between the nodes only occurs within distance of 1. Thus a user's sentiment is only influenced by himself and his followees. Then the distribution of a node or a user v_i is as follows

$$P(y_i \mid \mathbf{t}_i, \mathbf{y}) \propto \exp(\sum_{t \in \mathbf{t}_i, k} \mu_k f_k(y_i, x_t) \\ + \sum_{u \in following_i, k} \lambda_k h_k(y_i, y_u))$$

where $\mathbf{t}_i$ is the set of tweets posted by user v_i, $following_i$ denotes the users that v_i is following, y_i denotes the sentiment of user v_i, x_t represents the label of t and y_u denotes the sentiment of user u.

Based on the proposed framework, our task is to estimate the parameters in the model. The challenge lies in how to make use of the unlabeled data in the learning process. Traditional learning algorithm for graphical model is hard to apply in semi-supervised learning setting. Randomized algorithm can be helpful in solving such model. Thus we employ an algorithm similar to Metropolis-Hasting Algorithm. In this algorithm, we can avoid the solution of normalization factor Z and employ the unlabeled data naturally. While for the prediction part, with the learned model, loopy belief propagation is more accurate than randomized algorithm.

IV. EXPERIMENTS

The whole data set we crawled from Tencent contains 100,302,600 users, their tweets from Oct., 2011 to Dec., 2011 and their following relationships (more than 1 billion relationships). In the Tencent network, the user may post her/his popularity on a product in a survey. This information is used as ground-truth to evaluate the proposed approach and to compare with baseline methods. For baselines, we use Support Vector Machine (SVM) and Naïve Bayes (NB). The performance of our model is better (+3-10%) than several baseline methods.

V. CONCLUSION

We study the problem of sentiment diffusion in the social networks. We propose a framework for modeling and predicting users' sentiment and its influence on friends. Our experimental validate the effectiveness of the proposed model.

REFERENCES

[1] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Micro-blogging as online word of mouth branding. In CHI '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, pages 3859–3864, 2009.

[2] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi. Exploiting social context for review quality prediction. In WWW '10: Proceedings of the 19th international conference on World wide web, pages 691–700, 2010.

[3] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002.

[4] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In International AAAI Conference on Weblogs and Social Media, 2010.

[5] P. Turney. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the Association for Computational Linguistics (ACL), pages 417–424, 2002.

[6] B. Pang and L. Lee. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In ACL, 2004, 271—278.

[7] R. Agrawal, S. Rajagopala, R. Srikant and Yirong Xu. Mining Newsgroups Using Networks Arising From Social Behavior. In WWW, 2003, 529—535.

[8] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In KDD'08, pages 160–168, 2008.

[9] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In WWW'10, pages 641–650, 2010.

[10] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In KDD'09, pages 807–816, 2009.

[11] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li. User-level sentiment analysis incorporating social networks. In KDD'11, 1397-1405.

Novel Approach of Device Collaboration Based on Device Social Network

Kyuchang Kang, Dongoh Kang, and Changseok Bae, *Member, IEEE*

***Abstract*—This paper presents device collaboration based on device social network. In this work, we introduce the relationships between human social network and device social network and how to build the device sociality. The device sociality between devices is calculated by criteria of 'strength' and 'frequency'. The 'strength' means the connected time of two devices. The 'frequency' represents how many times two devices are connected. As a case study, we describe the I/O connection based on device sociality while we use virtual desktop.**

I. INTRODUCTION

People are living in society by forming human social relationships. In recent years, people particularly use smartphones as means of new communication methodologies. In the near future, people will get most of information through smartphones. Furthermore, they may use smartphones like their avatar performing their own commands and only reacting with them. Eventually, the device's sociality reflecting human social relationships may induce reverse flow of information and make it possible to close communication between people and devices.

According to the Pew Internet & American Life Project [1], social networking is popular and still growing. While only 8% of adult Internet users used social networking sites in 2005, that number had grown to 65% by 2011. Social networking has had a huge impact on how we communicate and interact.

Currently, we are developing zero-configured device interaction technology based on social networking of devices. This work is composed of two procedures; Firstly, we build social relationships among devices such as desktops, notebooks, tablets and smartphones used by individual user. Secondly, based on built devices' sociality, we connect and link resources or information among devices without particular configuration process while providing the functions of resource sharing, data exchanging or application collaboration. Current developing status of this work is on initial phase.

The concept of devices' relationship represented as 'device social network' is expected to contribute to the automation of mutual interface among user's own devices or friendly devices.

The remainder of this paper is organized as follows. In section II, we present conceptual approach and building of device sociality. The section III describes the case study of utilizing device sociality on SoD (System on-Demand) [2-3] service environment. Finally, we summarize this work and discuss the future works in section VI.

This work was supported by the IT R&D program of MKE/KEIT[12VS1810, Zero Configuration Type Device Interaction Technology using Device Sociality between Heterogeneous Devices]

II. METHOD

A. Conceptual Approach

As a paradigm shift, we may infer social phenomenon: from the concept that tracking of information stream used by smart devices is enabling to build human social relationships to the concept that the tracking of human social relationships and device's collaborating action are enabling to build device social network.

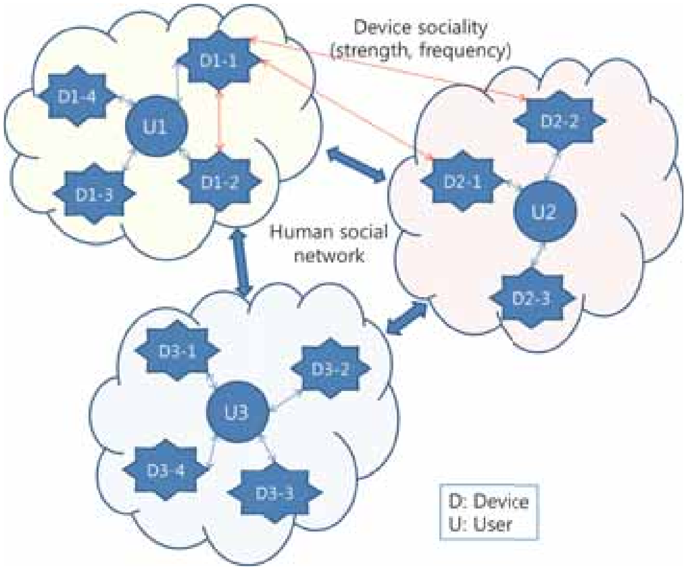


Fig. 1. Conceptual architecture of device social network

Fig. 1 shows the conceptual architecture of device social network. Basically, device social network is based on human social network. Firstly, the relationships among distinct users are analyzed by common social networking services [4] such as Facebook, Twitter and Google+. Consequently, the devices which are not owned by users consisting human social network are excluded. Secondly, by targeting only those devices, device social network is analyzed. The device sociality between devices is calculated by criteria of 'strength' and 'frequency'. The 'strength' means the connected time of two devices. The 'frequency' represents how many times two devices are connected.

B. Building of Device Sociality

As a prerequisite, we exclude the devices which are not owned by users of human social network. To analyze human social network, we can leverage conventional social network services such as Facebook, Twitter, Google+ and so on. Most of them provide open APIs enabling access to contents of social relationship information.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

After excluding non-target devices, every connection and invocation action between each device is monitored and recorded in device sociality mapping table shown in Table I.

Table I
DEVICE SOCIALITY MAPPING TABLE

Device Sociality		U1		U2		
		D1-1	D1-2	D2-1	D2-1	D2-3
U1	D1-1	-	(S, F)	(S, F)	(S, F)	(S, F)
	D1-2	(S, F)	-	(S, F)	(S, F)	(S, F)
U2	D2-1	(S, F)	(S, F)	-	(S, F)	(S, F)
	D2-2	(S, F)	(S, F)	(S, F)	-	(S, F)
	D2-3	(S, F)	(S, F)	(S, F)	(S, F)	-

S: strength, F: frequency

In Table I, ‘S’ represents the strength of connection between devices a day and ‘F’ describes the frequency of connection between devices a day respectively.

$$S = \text{time [second] of connection} / (24 \times 60 \times 60) \quad (1)$$

$$F = \text{number of connection} / 24 \quad (2)$$

III. RESULT

This section describes a simple case study of applying device sociality while we connect I/O (Input/Output) devices based on SoD (System on-Demand) technology [2-3].

The SoD service provides users with customized computers through network. The customized computers are comprised with user-preferred operating system, applications and peripheral devices through their dynamic integration.

The SoD service is to provide users in an on-demand way with instantaneous computing environment to enable users to use various services and software conveniently anytime and anywhere by connecting distributed smart devices through network and utilizing functional cooperation among the devices.

Fig. 3 shows an example of SoD service and I/O device connection. In previous work of leveraging SoD service, we connected I/O devices by selecting each device as input and output explicitly. However, by applying device sociality while we are selecting any device, we can connect each device implicitly.

Table II describes a simple example of device sociality between my phone and colleague’s smart pad.

Table II
EXAMPLE OF DEVICE SOCIALITY BETWEEN PHONE AND SMART PAD

Device Sociality		Mine	Colleague
		Phone	Smart Pad
Mine	Phone	-	(0.0417, 0.2083)
Colleague	Smart Pad	(0.0417, 0.2083)	-

In Table II, the strength between my phone and colleague’s smart pad is ‘0.0417’. This means two devices are connected 1 hour in a day. The frequency between my phone and colleague’s smart pad is ‘0.2083’. This means 5 times connection history in a day. Currently, we do not allow for the direction of connection so that two (S, F) is the same.

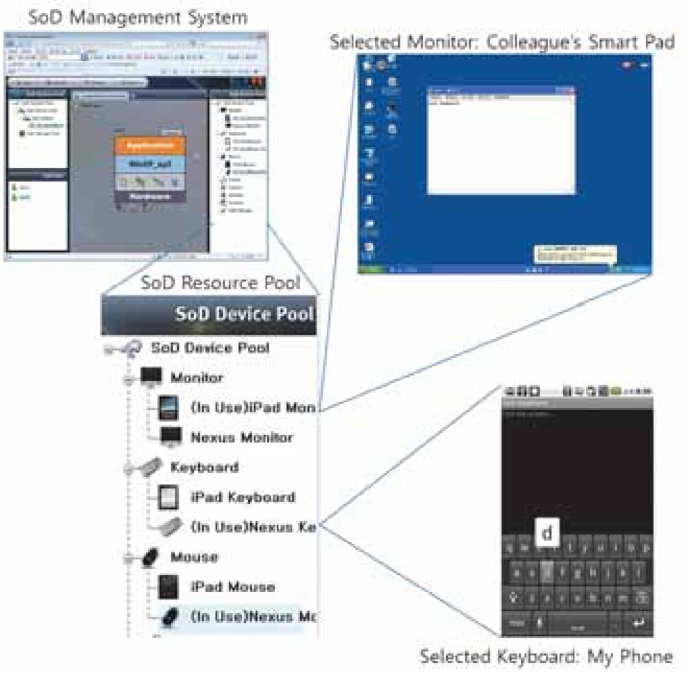


Fig. 3. Operation example of SoD service and I/O device connection

IV. DISCUSSION

In this paper, we propose new device collaboration scheme such as I/O connection based on device sociality described by the strength and frequency of connection between devices.

In current case study, we do not apply long-term device sociality information. However, if the strength and frequency information between devices are accumulated for long time, we may provide with intuitive I/O connection interface.

Additionally, we make it possible to find and link optimal applications from collaboration perspective, if the profiles of applications running on each smart device are managed by a management server. Comparing with human social network service, the management server has a role of service providers’ server.

At next stage of this work, we also allow for location information to find which device is optimally located and available.

REFERENCE

[1] Mary Madden and Kathryn Zickuhr, “65% of Online Adults Use Social Networking Sites,” Pew Research Center’s Internet & American Life Project surveys, Aug 26, 2011.

[2] Kiryong Ha, Dongho Kang, Hyungjik Lee, Kyuchang Kang, and Jeunwoo Lee , "SoD: Framework for on-demand computing in home environment," *Consumer Electronics (ICCE), 2011 IEEE International Conference on* , pp.577-578, 9-12 Jan. 2011

[3] Kyuchang Kang, Kiryong Ha, and Jeunwoo Lee, "Android-based SoD client for remote presentation," *Advanced Communication Technology (ICACT), 2011 13th International Conference on,* pp.1162-1167, 13-16 Feb.

[4] Wikipedia, “Social Networking Service”, http://en.wikipedia.org/wiki/Social_networking_service.

Development and Evaluation of Myoelectric Driving Interface

Jaesung Oh, Minsuk Kwon, Youngwon Kim, Jungsoo Kim, Sungyoon Lee and Jaehyo Kim
Department of mechanical & control engineering, Handong Global University, Pohang, Korea

***Abstract*-- This study proposes a myoelectric driving interface and its evaluation. The interface uses two wrist angles to accelerate, brake and steer. Two tasks were carried out to investigate the learnability, steering and speed control performance of the interface by comparing with the actual Wheel/Pedal interface. The proposed interface showed better learnability than Wheel/Pedal interface but was behind in driving performance.**

I. INTRODUCTION

An EMG(Electromyogram) is a bioelectric signal generated by muscle use which represents the muscle's voluntary activation level. Various movement characteristics such as joint stiffness, joint torque, time-continuous angle of joint, posture can be estimated EMG signals [1]-[5] transformed into quasi-tension [6], [7]. These characteristics can be used to develop intuitive interfaces using body parts as the interface itself. For the control interface using biosignals, training and characteristics estimation procedure should be simple. In the previous research, we have verified that movement direction, speed, difficulty, time-continuous angle and halt intention can be estimated solely from EMG signals during PTP(Point-to-Point) flexion/extension wrist movement using simple mathematical methods with reliable accuracy [8]. These methods can be applied to develop a myoelectric interface.

This study proposes a myoelectric driving interface which uses two wrist angles estimated from EMG signals to accelerate, brake, and steer. Two tasks are carried out to investigate the learnability and the driving performance of the interface by comparing with the existing wheel/pedal interface.

II. MATERIALS & METHODS

A. Experimental Setup

The subject is seated on a chair and asked to drive the car on the monitor screen using both myoelectric driving interface(MyoDrive) and the wheel/pedal interface. The MyoDrive uses two wrist angles to control the steering and speed as shown in Fig. 1. The horizontal angle of the right wrist controls the steering, and the vertical angle of the left wrist controls the speed. The angles are estimated from EMG signals measured at carpi radialis and carpi radilais longus which engage in the flexion/extension movement of the wrist. The signals are sampled with 1kHz sampling rate, transformed into quasi-tension [6], [7] and normalized with MVC(Maximum Voluntary Contraction). The wrist angle is estimated from the difference of the stiffness of both muscles calculated from EMG signals [7] using least square fitting. The reference angle is measured with AHRS(Altitude Heading Reference System) sensor attached on both hands

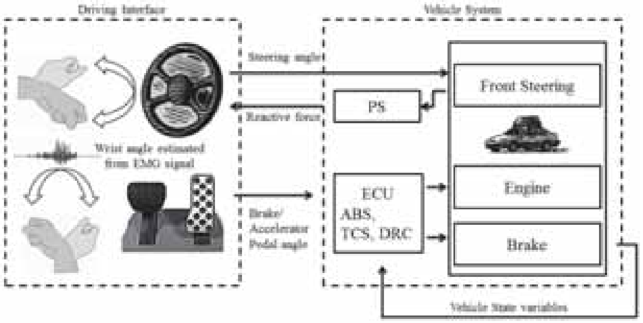


Fig. 1. Schematic of the myoelectric driving interface. The horizontal angle of the right wrist controls the steering, and the vertical angle of the left wrist controls the steering. Bending the wrist upward is acceleration, and bending the wrist downward is brake.

B. Experimental Procedure

Five male subjects, ranging in age from 22 to 25 participated in this study. Subjects were asked to conduct two tasks using both interfaces. Vehicle dynamics are applied to provide similar driving condition [9].

Task 1 was conducted to investigate the learnability and the steering performance of the interface. Subjects were asked to take a slalom test in which the subject zigzags between obstacles. Task difficulty was varied by increasing the number of obstacles – 4, 5, 6. Distance traveled and lap time were recorded during the task.

Task 2 was conducted to investigate the speed control performance of the interface. Subjects were asked to follow a target which changes its speed from 40 km/h to 60 km/h and then slows down back to 40 km/h. Speed profile of the subject was recorded.

III. RESULTS

A. Learnability and steering performance

Fig. 2 shows the convergent characteristic of the lap time recorded during task 1 as the number of trials increases. The number of trials taken until the lap time converges to a certain value represents the learnability of the interface. MyoDrive required average 6 trials to learn the interface and Wheel/Pedal required average 5.33 trials. The traveled distance is average 353.14±60.35 for MyoDrive and 341.94±26.97 for Wheel/Pedal as shown in Table 1. The number of trials required to learn Wheel/Pedal is less than that of MyoDrive, but considering that the subjects are already used to Wheel/Pedal interface through games or actual driving experience and that the lap time is usually less than that of Wheel/Pedal, MyoDrive has better learnability. However, Wheel/Pedal has better steering performance as it has less traveled distance during the task.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

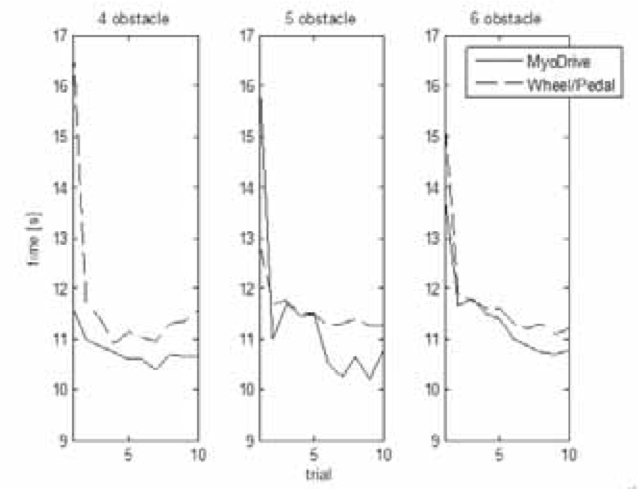


Fig. 2. Lap time according to trials and task difficulty. Both interfaces convergent characteristic in all difficulties.

TABLE I
DISTANCE TRAVELED DURING TASK 1

	4 obstacles	5 obstacles	6 obstacles	Average
MyoDrive	381.70±34.18	344.12±19.17	369.29±26.85	353.14±60.35
Wheel/Pedal	333.26±31.80	347.18±25.86	346.91±17.00	341.94±26.97

B. Speed control performance

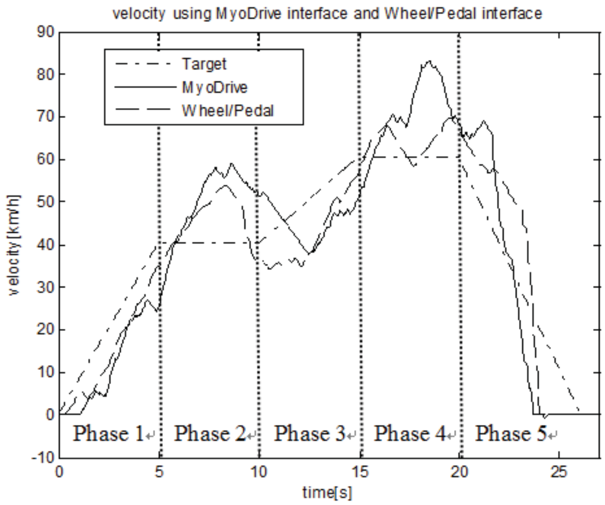


Fig. 3. Speed profile of the Target and the Subject during task 2.

TABLE II
SPEED DIFFERENCE BETWEEN THE TARGET AND THE SUBJECT

	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
MyoDrive	-9.74±2.02	8.72±5.17	-7.64±2.43	8.70±4.81	1.56±4.68
Wheel/Pedal	-11.06±3.28	5.12±2.83	-6.70±4.46	5.79±2.22	0.29±6.56

Figure 3 shows the speed profile of the target and the subject during task 2. The task consists of 5 phases — two acceleration(phase 1 and 3), two constant velocity(phase 2 and 4), and one deceleration(phase 5). The speed difference between the target and the subject is compared to investigate the speed control performance. Wheel/Pedal showed less difference than MyoDrive in all phases except for phase 1, and therefore has better speed control performance.

IV. CONCLUSION

This study proposed a myoelectric driving interface which uses two wrist angles to accelerate, brake and steer and also evaluated the interface by comparing with the existing Wheel/Pedal interface. Two tasks were carried out to investigate learnability and speed control performance of the interface. MyoDrive showed better learnability than Wheel/Pedal, but lacked both steering and speed control performance compared to Wheel/Pedal.

The reason for the better learnability of MyoDrive is because the internal model of our body is already acquired as we use our body daily. On the other hand, when using Wheel/Pedal interface, subjects should learn the interface to use it.

The lack of driving performance is because of the EMG signal's trembling characteristic and unrestrained movement of MyoDrive interface. EMG signals are not maintained but decrease even if same posture is maintained, and this causes trembling in the signal, which causes unstable input. In addition, MyoDrive does not restrain the movement of the wrist whereas Wheel/Pedal restrains the movement with the motor of the wheel and spring of the pedals, which works as the feedback to the subject and therefore allows precise control.

If the interface is designed considering such characteristics, MyoDrive will have both better learnability and driving performance. Further studies are needed for the interface improvements.

REFERENCE

[1] Y. Koike, M. Kawato, "Trajectory formation from surface EMG signals using a neural network mode,l" In Proceedings of annual international conference of the IEEE engineering in medicine and biology society 15, pp. 1628-1629, 1993.

[2] H. Gomi, M. Kawato, "Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement," Science, vol. 272, pp.117-120, 1996.

[3] Gomi H., Kawato M., "Human arm stiffness and equilibrium point trajectory during multi-joint movement.," Biol Cybern, vol.76, pp.163-171, 1997.

[4] Osu R., Gomi H., "Multi-joint muscle regulation mechanisms examined by measured human arm stiffness and EMGsignals," J. Neurophysiol, pp.1458-1468, 1999.

[5] J. Kim, "Study of human's motor control model using 2 degree of freedom target tracing", Advanced Science Letters, vol 13, pp. 347-350, 2012.

[6] Y. Koike, M. Kawato, "Estimation of arm posture in 3D-space from surface EMG signals using a neural network model," IEICE Transactions on Fundamentals, E77-D 4, pp.368-375, 1994.

[7] Y. Koike, M. Kawato, "Estimation of dynamic joint torques and trajectory formation from surface electromyography signals using a neural network model," Biological Cybernetics, vol.73, p. 291-300, 1995.

[8] Y. Kim, J. Oh, M. Kwon, J. Kim, "Wrist joint movement characteristic estimation from EMG signal", iNFORMATION, vol. 15, No. 6, pp. 2487-2498, 2012.

[9] Thomas D. Gillespie: Fundamentals of Vehicle Dynamics (SAE International, USA, 1992.

A Novel Anti-Vignetting Method for Color Shading Artifact Suppression

Ja-Won Seo*† and Jong-Hyub Lee†
*Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea
†DMC R&D Center, Samsung Electronics, Suwon, Korea

Abstract—**In this paper, we present a novel automatic anti-vignetting method which effectively alleviates color shading artifacts regardless of ambient color temperature. The proposed method incorporates both a Color Temperature Metric (CTM) and a Vignetting Gain Control (VGC) algorithm in a commercial Image Signal Processor (ISP) which is embedded in mobile phone camera modules. Experimental results validate that our proposed method explicitly addresses the color shading problem in a wide variety of color temperatures.**

I. Introduction

People who have captured homogeneous achromatic-colored subjects (e.g., paper, wall, etc.) using commercial phone cameras may experience that a captured image becomes parti-colored as shown in Fig. 1. In general, we call this prevalent artifact a *color shading*, which stems from the so-called *vignetting effect*. The vignetting effect indicates the gradual falloff of light intensity towards the image periphery mainly due to the small aperture of inner lens system. Therefore, a robust anti-vignetting method is an essential component in the ISP to prevent the color shading artifact. In general, previous anti-vignetting approaches are grouped into two categories: 1) functional approximation based methods [1], [2], and 2) LUT (Look-up Table) based methods [3], [4]. The former and the latter utilize a polynomial and a LUT respectively to compute pixel-wise anti-vignetting gains.

According to [5], the color shading artifact is highly influenced by an illumination spectrum (i.e., color temperature of an illuminant). However, most previous approaches have not taken it into account but have assumed a specific illuminant in establishing their algorithms. Therefore, we propose a new anti-vignetting method which robustly operates irrespective of color temperature changes by incorporating the aforementioned functional approximation and LUT based methods.

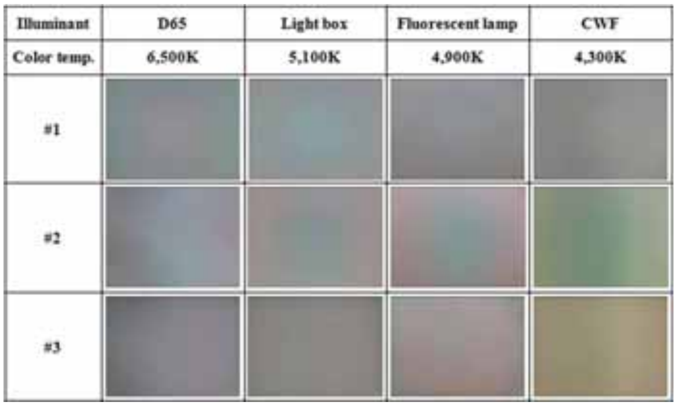


Fig. 1. The gray subject images captured with different three phone cameras (#1, #2, and #3) under four different illumination conditions.

II. Proposed Method

Figure 2 illustrates an overall framework of the proposed anti-vignetting method and a photograph of our camera module and evaluation board in which the framework is embedded. The framework consists of two blocks: one is the CTM which estimates the color temperature ($\hat{R_n}$) of a current illuminant from the vignetting-compensated RGB channels (I_{XY^C}), and the other is the VGC which computes the pixel-wise RGB anti-vignetting gains (S_{XY^C}) for image sensor output (RAW_{XY^C}). The set of pixels in image sensor which is exampled in Fig. 2 (a) is defined as (1) for image width (W) and height (H).

$$\begin{aligned} XY^C &= \left\{(x,y)^T \,\middle|\, 1 \le x \le W,\ 1 \le y \le H\right\} \\ &= XY^B \cup XY^{G1} \cup XY^{G2} \cup XY^R \qquad (1) \end{aligned}$$

First, the CTM operates based on color analysis of various gray subject images which are collected from various indoor and outdoor illumination conditions. As shown in Fig. 3, each gray subject image is plotted on R_n against B_n plane which indicates the normalized red and blue components of the gray subject in RAW_{XY^C}. Fortunately, indoor illuminants are evidently distinguishable from outdoor ones, so they can be parameterized by the GMM (Gaussian Mixture Model). Also, various illuminants in each Gaussian model are discriminable along the regression lines (i.e., f_{in} and f_{out}). Therefore, for a given scene ($X \in \mathbb{R}^2$) under the gray world assumption,

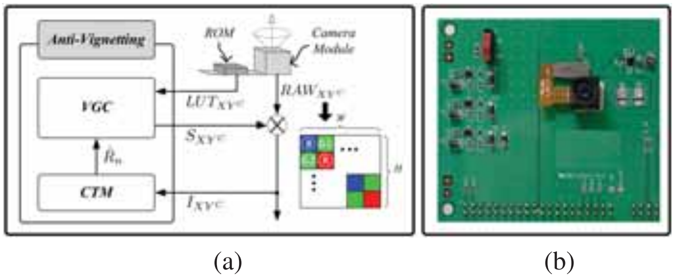


Fig. 2. The proposed anti-vignetting method. (a) overall framework of the proposed method. (b) a camera module on evaluation board.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

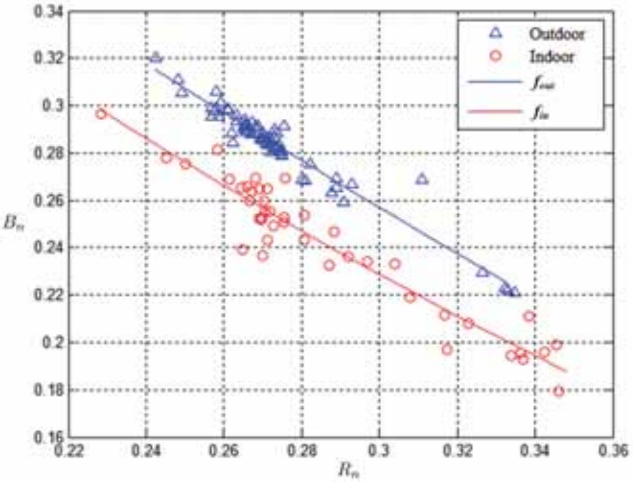


Fig. 3. Distribution of gray subject images on R_n against B_n plane which are collected under various both indoor and outdoor illumination conditions.

the color temperature is estimated as follows : 1) measure the Mahalanobis Distances (MD) to both distributions, then decide the class the scene belongs to as (2), and 2) find the closest point ($\hat{X}$) on the regression line to X as (3).

$$k = \arg \min_{i \in \{in, out\}} (\mathrm{MD}_i) \tag{2}$$

$$\left(\hat{R}_n, \hat{B}_n\right) = \arg \min_{\hat{X} \in f_k} \left\| \hat{X} - X \right\|_2 \tag{3}$$

Finally, we can utilize the $\hat{R}_n$ as an index of the color temperature thanks to its good discriminability.

Second, the VGC algorithm utilizes the $\hat{R}_n$ to adjust the canonical RGB anti-vignetting gains, i.e., LUT_{XY^C}, which are stored in an off-chip memory (i.e., ROM). The LUT_{XY^C} is the pixel-wise predetermined gains which make each channel profile flatten out under the user-specific illumination condition. In order to bend each channel profile smoothly, we can compute new pixel-wise gains, i.e., S_{XY^C}, as (4) by utilizing a hyperbolic cosine function, $L^C(D)$.

$$\begin{aligned} L^C(D) &= \cosh(\kappa^C D) \\ S_{XY^C} &= LUT_{\mathbf{XY}^C} \oslash L^C(D) \ , \end{aligned} \tag{4}$$

where $\oslash$ denotes the element-wise division operator, $D = \|XY - XY_o\|_2/W$, $XY_o = (W/2, H/2)$, and κ^C is a curvature control parameter for RGB channels, which can be computed as (5). The required RGB anti-vignetting gains at horizontal image edge (i.e., γ^C) are experimentally decided to maintain constant RGB luminance ratios between image center and corners, and then a LUT that exists in the VGC stores the $\hat{R}_n$ against γ^C relations.

$$\kappa^C = 2 \cdot \cosh^{-1}\left(1/\gamma^C\right) \tag{5}$$

Consequently, the vignetting-compensated RGB channels, i.e., I_{XY^C}, are computed as (6).

$$I_{XY^C} = S_{XY^C} \otimes RAW_{XY^C} \ , \tag{6}$$

where $\otimes$ denotes the element-wise multiplication operator.

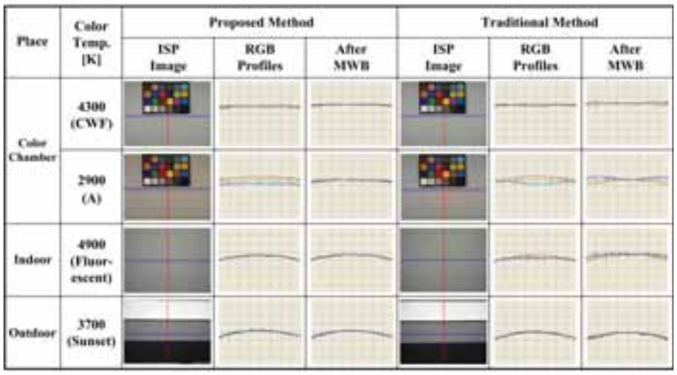

Place
Color Temp. [K]
Proposed Method
Traditional Method
ISP Image
RGB Profiles
After MWB
ISP Image
RGB Profiles
After MWB
Color Chamber
4300 (CWF)
2900 (A)
Indoor
4900 (Fluorescent)
Outdoor
3700 (Sunset)

Fig. 4. Comparison of the experimental anti-vignetting results between the proposed method and the traditional method under four different illumination conditions.

III. Experimental Results

Figure 4 compares the horizontal RGB profiles of ISP output images when the traditional anti-vignetting block is replaced with our proposed method. The proposed method utilizes the S_{XY^C} which adaptively varies as the ambient color temperature changes, but the traditional method uses only the LUT_{XY^C} for anti-vignetting gains. (Refer to the Fig. 2 (a)). Additionally, we add the so-called After MWB (Manual White Balance) results whose red and blue levels are moved to green level at image center to accentuate the RGB level differences of the two methods. For four different illumination conditions in different places, the proposed method shows negligible differences among RGB profiles. However, the traditional method shows gradual difference towards the image periphery, thus ring type color shading artifacts are noticeable in the ISP images.

IV. Conclusion

In this paper, we present a novel anti-vignetting method which remarkably suppresses the color shading artifacts. The experimental results demonstrate that the proposed CTM estimates the ambient color temperatures effectively, and then the VGC algorithm computes the pixel-wise anti-vignetting gains precisely based on the estimated color temperature. As a result, our proposed anti-vignetting method shows robust performance against the ambient color temperature changes.

References

[1] W. Yu, "Practical Anti-vignetting Methods for Digital Cameras," *IEEE Trans. on Consumer Electronics*, vol. 50, no. 4, pp. 975–983, 2004.

[2] K. He, P.-F. Tang, and R. Liang, "Vignetting image correction based on gaussian quadrics fitting," *Proc. of International Conference on Natural Computation*, vol. 5, pp. 158–161, 2009.

[3] I. Dinstein, F. Merkle, T. D. Lam, and K. Y. Wong, "Imaging system response linearization and shading correction," *Proc. of IEEE International Conference on Robotics and Automation*, vol. 1, pp. 204–209, 1984.

[4] P. Muralikrishna, S. Prakash, and B. Subbarya, "Digital processing of spacelab imagery," *Advances in Space Research*, vol. 2, no. 7, pp. 107–110, 1982.

[5] T. Tajbakhsh, "Color lens shade compensation achieved by linear regression of piece-wise bilinear spline functions," *Proc. of SPIE 7537*, 75370P, 2010.

Half-Face Detector for Enhanced Performance of Flash-Eye Filter

Peter M. Corcoran[1], *Fellow, IEEE,* Petronel Bigioi[1,2], *Snr. Member, IEEE,* Florin Nanu[3].

***Abstract*-- Red-eye and flash-eye defects in still photography continue to cause problems for digital imaging devices. New variants of flash-eye defects have appeared as cameras and cameras sub-systems get smaller in size. This paper describes advanced techniques to improve the detection of flash-eye defects using a novel half-face detector. In addition to improving the detection rate, the inherent symmetries of many of the haar-like classifiers used for such a face detector allow compression of the classifier chain providing benefits for resource constrained consumer electronics devices.**

I. INTRODUCTION

In a recent article Corcoran et al. have provided a detailed overview of a wide range of techniques employed in the detection and correction of red-eye and flash-eye defects in digital images [1]. In particular these authors have examined techniques that can be adapted within a digital camera, primarily to improve the detection of such flash defects. Flash defects are also extended to cover classes of non-red or part-red (hybrid) defects.

The analysis, detection and confirmation of these classes of flash-eye defects is invariably more onerous than simple red-eye artifacts and thus it is important to reduce the number and size of regions in an image where such algorithms are applied, particularly if the goal is to achieve a real-time detection in a consumer imaging device.

II. FACE DETECTOR WITH RED-EYE FILTER

One approach takes advantage of the fact that the majority of non-red artifacts typically occur in a pair with a conventional red-eye artifact. Thus, after a basic red-eye algorithm is applied to find all standard flash defects it is likely in a state-of-art consumer camera that a face detection result for the currently imaged scene will be available at the end of each preview frame [3]. This combination of (real-time) face tracking with basic red-eye filter enables a determination of faces that have a paired set of red-eye, and those which have a single, unpaired, red-eye.

By applying a more inclusive filter, or using a non-red algorithm it is practical to determine "missing" eye artifacts [2] as the areas of the image that must be scanned are very significantly reduced. Thus many undetected eye defects can be found by application of more sophisticated analytic techniques and additional confirmation steps in the detection process. A simple flowchart is provided in *Figure 1.*

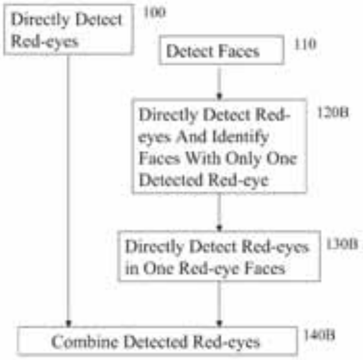

Directly Detect Red-eyes
100
Detect Faces
110
Directly Detect Red-eyes And Identify Faces With Only One Detected Red-eye
120B
Directly Detect Red-eyes in One Red-eye Faces
130B
Combine Detected Red-eyes
140B

Figure 1: Eye-pair technique to optimize the workflow tasks [2].

III. HALF-FACE DETECTOR TECHNIQUES

Another interesting use of incomplete face regions for red-eye detection arises from the symmetries that occur in the classifiers employed in many state-of-art face detectors. Some classifiers only apply to one side of the face - a selection of left-hand face classifiers are shown in the two left-hand side columns of *Figure 2* below. The first column shows the classifier located within the scanning window used to transverse the main image scene; the second illustrates its relationship to the detected face region.

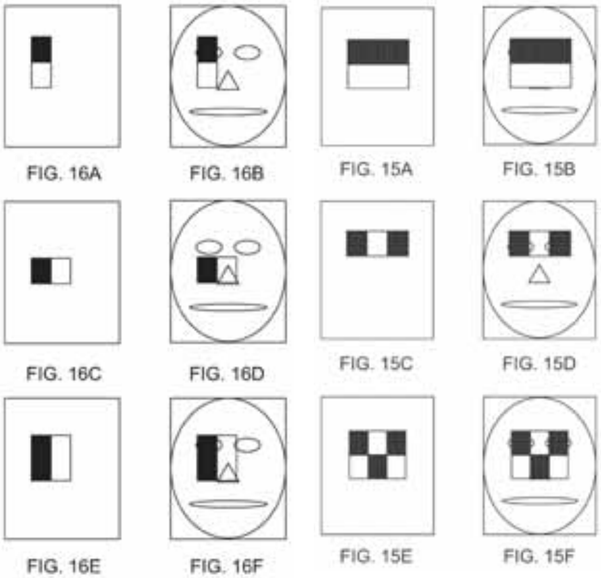


Figure 2: Predicted face candidate regions for next preview frame [4-7].

Note that each of these three example *half-face* classifiers can also be applied to right-hand face regions if they are flipped horizontally within the scanning window. Thus, although these classifiers are *asymmetric* within their scanning window they can be applied within a classifier cascade to detect either left-hand, or right-hand face regions through a

[1]*College of Engineering & Informatics, National University of Ireland Galway;* [2]*Digital Optics Corporation, Galway;* [3]*Digital Optics Corporation, Romania*

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

simple horizontal flipping transformation as explained by Nanu, Petrescu, Gagnea, Capata, Ciuc, Zamfir et al [4], [5], [6] and [7]. This implies that the number of stored classifiers can be reduced.

The second category of face classifier is a *symmetric* classifier shown in the two right-hand columns of *Figure 2*. Again the third column shows the classifier located within its scanning window, while the fourth column of images shows the classifier applied to a face region. Note that these *symmetric* classifiers apply to an entire face region and will return an error if a complete face region is not present.

Figure 3 shows the very first classifier of *Figure 2* as applied to (a) a left-hand half face; (b) a full-face, and (c) a right-hand half face. Clearly this particular, *left-face* classifier will successfully detect both left-face and full face regions, but will reject the right-face region. This observation leads to the concept of a *left-face* classifier chain that positively detects both left-face and full-face regions, a *right-face* classifier chain that detects right-face and full-face regions and a *full-face* classifier chain that detects only complete face regions.

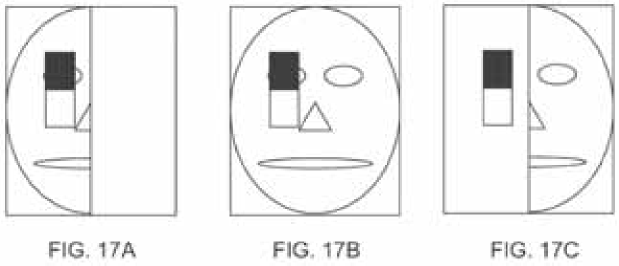


Figure 3: Predicted face candidate regions for next preview frame [5].

Note that a conventional face detector cascade will contain essentially the same classifiers but they will be ordered randomly according to their detection rates as deduced from the training process. Our training process differs as it trains independently for half-faces and full faces. The independently determined classifier sets are subsequently re-organized, removing unnecessary *left-* or *right-face* classifiers from the full face chain.

By ordering these classifiers according to their *asymmetric* or *symmetric* nature we retain the same capabilities as the conventional detector, with the additional benefit of being able to also detect *right-* and *left-faces*. An example workflow is shown in *Figure 4* that implements a combined half-face and full-face detector.

IV. Conclusions

Now the advantage of this approach can be appreciated - in addition to confirming eye-pairs using detected *full-faces* it is now also possible to confirm single-eyes using detected *half-faces*. This ensures that we avoid rejecting single-eye flash defects where they occur in partial face regions. It also may suggest additional half-face regions of the image which should have a more thorough analysis applied to ensure that a difficult eye-defect has not been overlooked.

In addition to its uses for enhanced red-eye and flash-eye defect detection this partitioning of the face detection process has many potential applications in image enhancement of scenes containing face regions. These and additional details on half-face detector techniques will be presented in an expanded version of this paper.

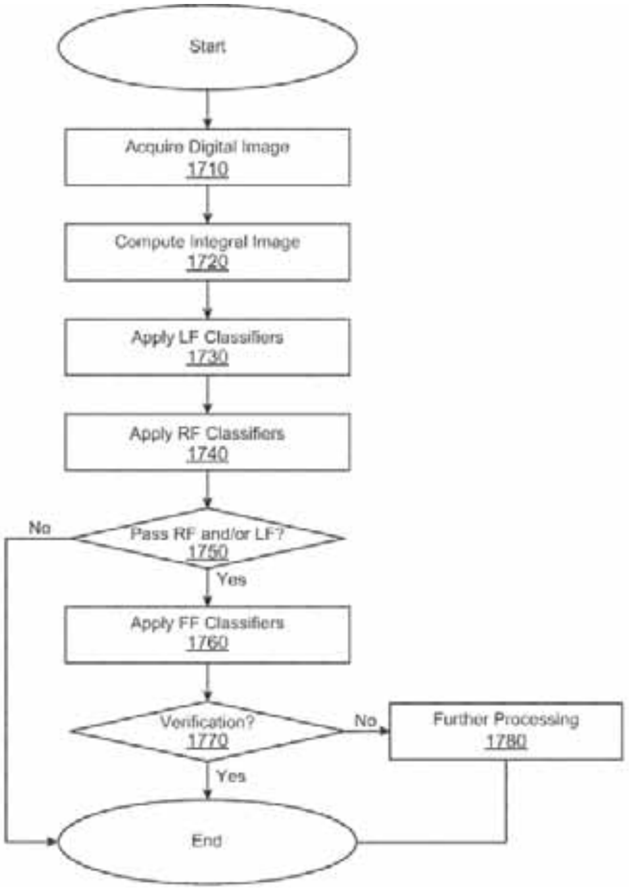


Figure 4: Predicted face candidate regions for next preview frame [5].

References

[1] P. Corcoran, P. Bigioi and F. Nanu, "Advances in the Detection & Repair of Flash-Eye Defects in Digital Images - A Review of Recent Patents", *Recent Patents on Electrical Engineering,* Vol 5 (1), pp. 30-54, April 2012.

[2] F. Nanu, P. Corcoran, A. Capata, A. Drimbarean, and E. Steinberg, "*Method of Detecting Redeye in an Image," US Patent Application 2*0080112599, May 15, 2008.

[3] P. Corcoran, E. Steinberg, S. Petrescu, A. Drimbarean, *F. Nanu, A. Pososin, and P. Bigioi, "Real-time face tracking in a digital image acquisition device*," US Patent 7,315,631, January 1, 2008.

[4] F. Nanu, S. Petrescu, M. Gangea, A. Capata, M. Ciuc, A. Zamfir, *E. Steinberg, P. Corcoran, P. Bigioi, and A. Pososin, "Partial Face Detector Red-Eye Filter Method and Apparatus*," US Patent Application 20100053362, March 4, 2010.

[5] F. Nanu, S. Petrescu, M. Gangea, A. Capata, M. Ciuc, A. Zamfir, E. Steinberg, P. Corcoran, P. Bigioi, and A. Pososin, "*Partial Face Tracker for Red-Eye Filter Method and Apparatus*," US Patent Application 20100053367, March 4, 2010.

[6] F. Nanu, S. Petrescu, M. Gangea, A. Capata, M. Ciuc, A. Zamfir, E. Steinberg, P. Corcoran, P. Bigioi, and A. Pososin, "*Analyzing Partial Face Regions for Red-Eye Detection in Acquired Digital Images*," US Patent Application 20100054592, March 4, 2010.

[7] F. Nanu, S. Petrescu, M. Gangea, A. Capata, M. Ciuc, A. Zamfir, E. Steinberg, P. Corcoran, P. Bigioi, and A. Pososin, "*Analyzing Partial Face Regions for Red-Eye Detection in Acquired Digital Images*," US Patent Application 20110063465, March 17, 2011.

Moving Object-High Dynamic Range Imaging (HDRI) for Artifact-free Digital Camera

Wonhee Choe, *Member, IEEE*, Sungchan Park, Hyunhwa Oh, and SeongDeok Lee, *Member, IEEE*
Samsung Advanced Institute of Technology (SAIT), Yongin, Republic of Korea

***Abstract* — We present a new artifact-free HDRI technology for a consumer digital camera that is based on de-ghosting with a dual-brightness mapping. The proposed approach reduces motion artifacts and the number of capturing images.**

I. Introduction

A number of researchers have studied to expand a dynamic range with combining multiple low dynamic range (LDR) images [1][2]. Lately some of the researchers have proposed approaches to correct motion artifacts which are the common problems in using sequentially captured images [3][4][5][6]. However, it is difficult to introduce them to actual products, because the High Dynamic Range Imaging (HDRI) technologies require the precise detection of misaligned regions as well as sequentially captured several images with highly overlapped dynamic ranges. In real HDR capturing conditions, these technologies may produce motion artifacts at the boundaries of saturation regions or generate a HDR image with a little expanded dynamic range.

This paper presents a new HDRI technology for an artifact-free digital camera that is based on de-ghosting with a dual-brightness mapping. This technology allows us to optimize capturing time and to remove motion artifacts by a camera motion and moving objects. To optimize the capturing time, the key component of our approach is the dual-brightness mapping to detect and compensate the artifacts with just two images.

In this paper, we describe our newly developed algorithm and the artifact-free HDRI system in which the algorithm has been implemented. Then the experimental results show that our approach has no artifact in moving object-HDR images.

II. Proposed Alogrithm

There are three problems when de-ghosting is applied to HDRI in digital still camera. First, the misaligned region has to be estimated precisely. Second, motion areas around saturation regions can generate some ghost artifacts due to inaccurate detections. Third, three or more images have to be captured with sequentially different exposure time. To solve these problems, we propose a new algorithm with dual-brightness matching. It consists of three functions as follows:

A. Dual-brightness mapping based Image Registration

The misaligned ghost region is detectable if the ghost boundary of reference region is discernible. However, some object motions around under- or over-saturation region still produce severe artifacts. We assume that the ambiguous ghost region is caused by the dynamic range limitation of a reference image (I_{AET}) as Fig. 1. We propose a dual-brightness mapping method to solve these problems. According to [7], Image Processing Chain (IPC) compresses the dynamic range of a raw image. Therefore, we use two uncompressed raw images with different exposures. That is the shorter exposed image (I_{SET}) has the wider dynamic range as a reference image. And the longer exposed image (I_{LET}) has the dynamic range without discontinuity of scene irradiance (Fig. 2(a)). The first brightness mapping is accomplished from I_{SET} to I_{LET} to express low luminance details according to Fig. 2(b), $I_{SET'2LET}$. Then, I_{LET}, compared with $I_{SET'2LET}$, is processed by motion correction and de-ghosting. The other brightness mapping is from I_{SET} to $I_{SET''}$ to protect high luminance details. $I_{SET''}$ is fused with de-ghosted I_{LET} to produce a final HDR image. That is, the dual-brightness mapping generates two different images from one input image.

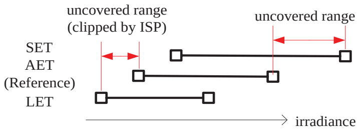
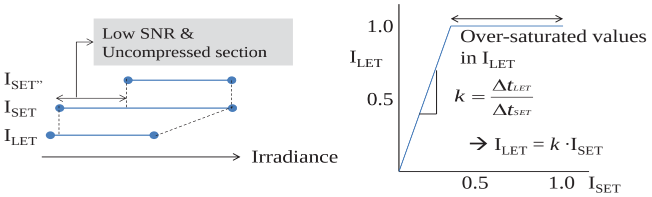


Figure 1. Irradiance range limitations at each image



(a) Irradiance range (b) Brightness matching function

Figure 2. Brightness matching with I_{SET} and I_{LET}.

B. Ghost Region Detection

Although motion estimation and compensation applied on the images, unexpected object motions can still induce the artifacts. The compensated images are re-searched with dual thresholds for detecting robustly any ghost. That is, the high and low threshold values find the intensity-differences between I_{LET} and $I_{SET'2LET}$. This process generates two ghost maps. And the maps are fused to generate a final detected ghost image (I_G). The two threshold values are adjusted by the noise level.

C. Ghost Reduction

An oversaturated region including a walking person and background is hard to determine the location where the real ghost boundary of the person is. In other words, if there is a misaligned region around saturation area, most of HDRI technologies are difficult to separate exact regions from the misaligned or aligned regions. To avoid these problems, our proposed method consistently replaces the pixel values belonging to the detected

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

ghost image in I_{LET} with those in $I_{SET'2LET}$. Then the brightness errors might be generated around the switched region boundary. The errors are treated with a spatial blending method using adaptively controlled weights [8].

III. System Implementation

Our approach is shown in Fig. 3. On the shutter button pressing, both shorter and longer exposure images, I_{SET} and I_{LET} respectively, are captured as [9]. Before the matching process, detail-preserving noise reduction [10] is accomplished only for I_{SET}. The brightness matching block generates two different brightness images, $I_{SET'2LET}$ and $I_{SET''}$. I_{LET} is accomplished by motion registration and de-ghosting function compared with $I_{SET'2LET}$. Then $I_{LET''}$ and $I_{SET''}$ are compressed by IPC, and produce a HDR image (I_{HDR}) with the HDR blending [9][11].

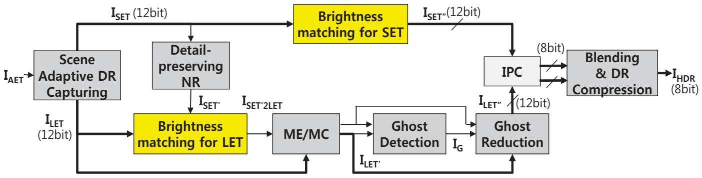


Figure 3. Block diagram of proposed method

IV. Experimental Results

We evaluated the performance of the Active Motion HDRI system with a Samsung GX-20 camera. Sequentially two images were captured under the handheld conditions. The average difference of the exposure time was 3-stop. The results showed as follows:

A. De-ghosting Performance

For natural HDR scene conditions, we took 30 pictures including people who move at a normal speed. Then we evaluated the results of ghost detection and reduction with our proposed method (Table 1). The detection was evaluated the rates of sensitivity and specificity with the ground truths. The reduction was evaluated as the artifact-free ratio on the ghost regions of the references. The detection and reduction were determined by their presence on the image from subjective tests.

TABLE I. Performance of Ghost Detection and Reduction

	Sensitivity[a]	Specificity[b]	Artifact-free[c]
Ratio	90.3%	88.5%	99.6%

a. Sensitivity = True Positive /(True Positive + False Negative)
b. Specificity = 1-False Positive/(False Positive + True Negative)
c. Artifact-free = the number of removed artifacts / the number of ghost regions

B. The other Performance

Figure 4 is the results with a Stouffer T4110 chart to test the dynamic range. The test was accomplished with the test camera, the proposed method, and a commercialized HDR camera (Sony A550). The dynamic ranges of generated images were about 9EV, >13EV, and 11EV respectively. Our result was beyond the dynamic range of the chart (>13EV). And, the de-ghosting test with metronomes was under the HDR condition as Fig. 5. The proposed method showed the clear image.

Lastly, overall image quality was also evaluated by subjective test with the 30 pictures which were captured for de-ghosting performance. Figure 6 shows the example of the test images. 10 people measured their satisfaction on 10-point scales compared with I_{AET}, I_{SET}, and I_{LET}. The average satisfaction score was 9.2.

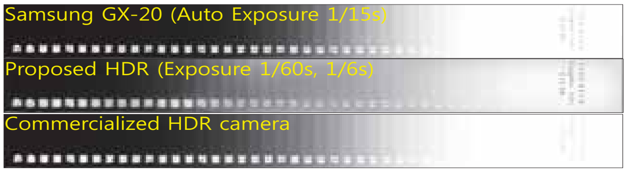

Samsung GX-20 (Auto Exposure 1/15s)
Proposed HDR (Exposure 1/60s, 1/6s)
Commercialized HDR camera

Figure 4. DR comparison of the test camera, proposed method, and commercialized HDR camera

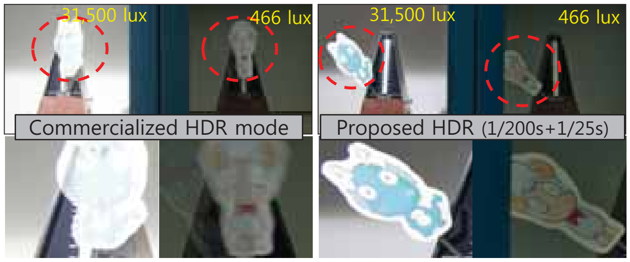


Figure 5. Artifact comparison of the proposed method and the commercialized HDR camera

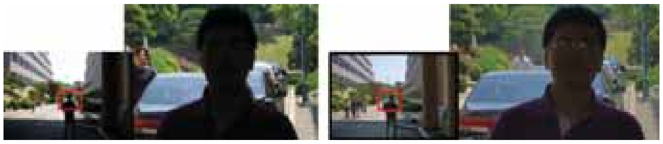


Figure 6. Examples of the auto-exposed image and the proposed HDR image

V. Conclusion

We have developed a new artifact-free HDRI based on a de-ghosting algorithm with the dual-brightness mapping. Evaluation results show that the proposed method can remove the ghost artifacts almost perfectly. Our method proved successfully on variety dynamic scenes. It will be applied to Samsung cameras (NX300) to create artifact-free HDR images on the real scenes.

References

[1] P. Debevec and J. Malik, “Recovering high dynamic range radiance maps form photographs,” SIGGRAPH ‘97, pp.369-378, 1997.
[2] S. Mann and R. Picard, “Being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed picture,” In Proc. IS&T 46th Annual Conf., pp.422-428, May 1995.
[3] S. B. Kang, M. Uyttendaele, S.Winder, and R. Szeliski, “High dynamic range video,” ACM Trans. On Graphics, vol.22, no.3, pp.319–325, 2003.
[4] O. Gallo, N. Gelfand, W. Chen, M. Tico and K. Pulli, “Artifact-free high dynamic range imaging,” IEEE Int’l Conf. on Computational Photo., 2009.
[5] A. Eden, M. Uyttendaele, and R. Szeliski, “Seamless Image Stitching of Scenes with Large Motions and Exposure Differences,” CVPR, pp. 2498-2505, 2006.
[6] N. Menzel and M. Guthe, “Freehand HDR photography with motion compensation,” Vision Modeling and Visual. 07, pp.127-134, 2007.
[7] JAI tech. note TH-1086, AccuPiXEL™ LUT Function, 2000.
[8] S. Park, J. Kwon, H. Oh, W. Choe, and S. Lee, “Motion artifact-free HDR imaging under dynamic environments,” 18th IEEE ICIP, pp.353-356, 2011.
[9] W. Choe, K Lee, J. Kwon, and S. Lee, “High Dynamic Range Imaging On Digital Still Camera,” IEEK Conf. Proceedings, pp. 2217-2220, June 2010.
[10] Y. Yoo, K Lee, W. Choe, S. Park, S. Lee and C. Kim, “A digital ISO expansion technique for digital cameras,” Proc. of EI, vol. 7537, 2010.
[11] K. Lee, W. Choe, J. Kwon, and S. Lee, “Locally Adaptive High Dynamic Range Image Reproduction Inspired by HVS,” Proc. of EI, vol. 7241, 2009.

Stereo Panoramic Image Stitching with a Single Camera

Junguk Cho, Joon Hyuk Cha, Yong Min Tai, Young-Su Moon, and Shihwa Lee
SAIT, Samsung Electronics, Korea

***Abstract*-- In this paper, we present stereo panoramic imaging method which generates stereo panoramas by stitching sequential images captured from a single camera under a hand-held sweeping condition. It is hard to take sequential images without a shake under a hand-held sweeping condition. The tilt of the camera makes the rotation of input images which causes the misaligned stereo panoramas. The misaligned stereo panoramas make the degradation of the stereo perception and the dizziness of the viewers. In order to make well-aligned stereo panoramas, we propose the vertical local alignment of the left and right strips using a reference panorama. The proposed method is measured against the existing method about in terms of disparity ranges.**

I. INTRODUCTION

A panorama is an image extending the field of view by stitching sequential images. A stereo panorama is an extension to add real depth to the panorama. A stereo pair is comprised of two panoramas which have two different viewpoints, corresponding to the position of the two eyes. Basically, stereo panoramas consist of two panoramas produced from several images which are taken from two separate cameras. Two separate cameras are improper for mobile devices like digital camera, smartphones, and tablets in terms of the cost and complexity of systems. Therefore, many techniques have been proposed for stereo panoramas with a single camera [1]-[4]. However, they have presented the photograph methods like the hand-held sweeping condition. They have not considered the misaligned stereo panoramas caused by the different position and the time delay of the strip extraction from the left and right images about the same viewpoint of the panorama. The misaligned stereo panoramas make the degradation of the stereo perception and the dizziness of the viewers.

In this paper, we propose a method that makes well-aligned stereo panoramas with impressive stereo perception. Especially we focus on the vertical local alignment of the left and right strips using the reference panorama. The proposed method is explained in detail in section II, and experimental results and conclusions are described in section III and IV, respectively.

II. THE PROPOSED ALGORITHM

Figure 1 shows the flowchart of the proposed stereo panoramic image stitching method. It consists of image selection and storing among sequential images, Spherical projection and alignment of stored images, reference panorama generation using the central region of images, left and right strip extraction used for stereo panoramas using the reference panorama, left and right panoramas generation, and Stereoscopic effect optimization.

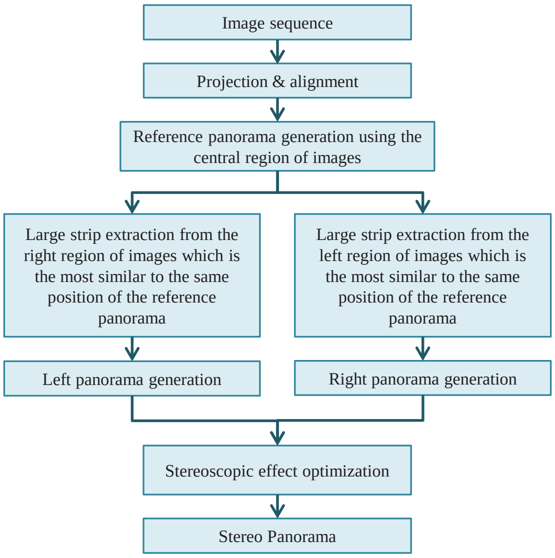


Image sequence
Projection & alignment
Reference panorama generation using the central region of images
Large strip extraction from the right region of images which is the most similar to the same position of the reference panorama
Large strip extraction from the left region of images which is the most similar to the same position of the reference panorama
Left panorama generation
Right panorama generation
Stereoscopic effect optimization
Stereo Panorama

Fig. 1. Flow of the proposed stereo panoramic image stitching method.

Since stereo panoramas consist of left and right panoramas having two different viewpoints, the misaligned stereo panoramas caused by the different position and the time delay of the strip extraction from the left and right images about the same viewpoint of the panorama. The misaligned stereo panoramas make the degradation of the stereo perception and the dizziness of the viewers [5].

Input images are projected using the spherical projection, and simply aligned by estimating translations between sequential images with matching corner points. Input images with uniform translation are necessary to make the well-aligned stereo panoramas. However it is hard to take sequential input images without a shake under a hand-held sweeping condition in practice. The tilt of the camera by the shakes makes the rotation of input images which causes the misaligned stereo panoramas as shown in Fig. 2. The misaligned stereo panoramas make the degradation of the stereo perception and the dizziness of the viewers. The rotation of input images makes the different translation of the left and right strips used for left and right panoramas as shown in Fig. 3. The top of Fig. 3 explains that the camera is slanted to the right during photographing; the bottom of Fig. 3 describes the rotation status of the captured image as input image. Since the global translation between images is the average of the local translation of an image, the variation of the global translation is small in case of the rotation of input images. However the vertical translation of the left and right strips is in the reverse direction. The important information is that the average of the horizontal translation in case of the rotation of input images is invariant. Therefore, we can make well-aligned stereo panoramas by the local alignment of the vertical direction about the left and right strips used for stereo panoramas after left and right panoramas are initially global aligned.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

In this paper, we propose the method of the vertical local alignment of the left and right strips. It makes another problem to align one strip to the other strip because both strips have incorrect translation caused by the rotation of input images. Since a reference is necessary to align both strips identically, the reference panorama is generated using the central region of images because the central region of images is the most robust to the rotation of images [6]. The left and right strips are extracted from the left and right region of input images which are the most similar to the same position of the reference panorama, respectively, as shown in Fig. 4. The results of the vertical local alignment of the left and right strips about reference panorama are shown in Fig. 5. The left of Fig. 5 shows the misaligned left and right strips at the same position before the vertical local alignment. The other shows the well-aligned left and right strips after the vertical local alignment.

III. EXPERIMENTAL RESULTS

A twenty image sequence sets are tested and the results of the proposed method are compared with the results of existing method which does not include the vertical alignment of the left and right strips. Input images were captured under a hand-held sweeping condition. Each test set has 700~800 still images that were captured as a 2304 * 1296 resolution using a compact digital camera with a manual setting in order to fix the exposure time and white balance parameters.

As shown in Table I, existing method has the large vertical disparity which causes the degradation of the stereo perception and the dizziness of stereo panoramas. The proposed method has the small vertical disparity that gives good stereo perception without any dizziness on 3DTV display. The disparity ranges in Table I are measured from histogram distribution of the feature disparity of left and right panoramas using SIFT feature matching. Fig. 6 shows the resulting stereo panorama using red-cyan anaglyph image which are overlapping two left and right panoramas.

IV. CONCLUSION

We present stereo panoramic imaging method which generates stereo panoramas by stitching sequential images captured from a single camera under a hand-held sweeping condition. Especially, we focus on the vertical local alignment of the left and right strips using the reference panorama.

Our method makes well-aligned stereo panoramas with impressive stereo perception without any dizziness on 3DTV display. The proposed stereo panoramic imaging method shows a sufficient capacity of the image quality as well as the processing time. As a result, it is possible to make a stereo panorama system for ordinary digital cameras, which have lower hardware specification, with only software implementation.

TABLE I
EXPERIMENTAL RESULTS

Metric	Proposed method	Existing method
Horizontal Disparity Range	0~40 (40)	0~50 (50)
Vertical Disparity Range	-4~+4 (8)	-7~+6 (13)

REFERENCES

[1] S. Peleg and M. Ben-Ezra, "*Stereo Panorama with a Single Camera*," In Proc. of CVPR, 1999.

[2] S. Peleg, M. Ben-Ezra, and Y. Pritch, "*Omnistereo: Panoramic Stereo Imaging*," IEEE Trans. on PAMI, 2001.

[3] T. Svoboda and T. Pajdla, "*Panoramic Cameras for 3D Computation*," In Proc. Of Czech Pattern Recognition Workshop, 2000.

[4] K. C. Zheng, S. B. Kang, M. F. Cohen, and R. Szeliski, "*Layered Depth Panoramas*," In Proc. of CVPR, 2007.

[5] R. Kosakai and S. Inaba, Image Processing Apparatus, Image Capturing Apparatus, Image Processing Method, and Program, Patent Number US20110157305A1, June 30, 2011.

[6] J. Cha, Y. Jeon, Y. Moon, and S. Lee "*Seamless and Fast Panoramic Image Stitching*," In Proc. of ICCE, 2012.

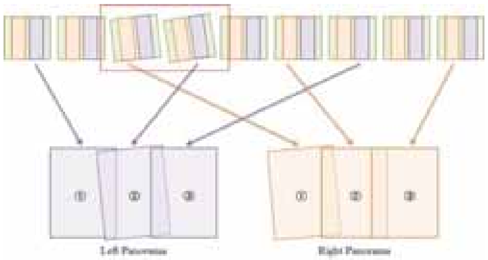


Fig. 2. Input image sequence captured from a single camera.

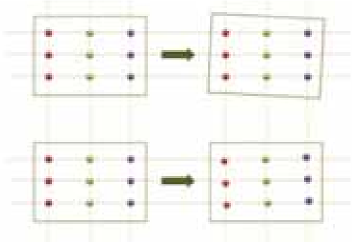


Fig. 3. Positions of subjects in the rotation of input images.

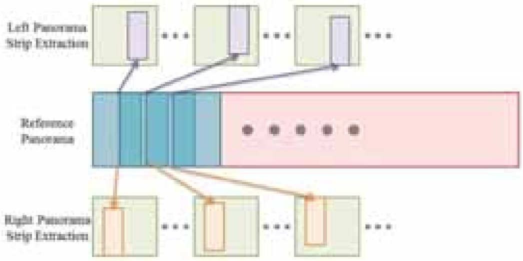

Left Panorama Strip Extraction
Reference Panorama
Right Panorama Strip Extraction

Fig. 4. Left and right strips extraction.

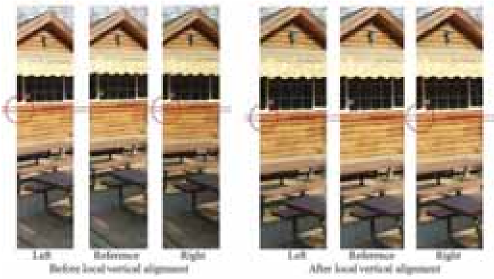

Left
Reference
Right
Before local vertical alignment
Left
Reference
Right
After local vertical alignment

Fig. 5. Results of vertical local alignment.

Fig. 6. An anaglyph image of proposed stereo panorama.

A Crash Course on Patents for Engineers

Peter M. Corcoran[1], *Fellow, IEEE*

***Abstract*-—— Most engineers are aware of the existence of patents. Indeed the preparation of patent disclosures is an important aspect of most research engineering job functions. Yet for many engineers the world of patents remains shrouded in a cloud of mystery. In this paper we try to dispel some of the mystery surround the quasi-technical, quasi-legal nature of patents and provide a set of practical, hand-on, guidelines to the patent literature from an engineer's perspective.**

I. INTRODUCTION

Patents are an important tool for technology-based businesses to protect rights and ring fence new technical and engineering techniques. Research and development is an expensive business and would not be viable commercially unless a company can protect new techniques from copying and duplication by competitors.

As most CE engineers work in a fast changing technological environment it is inevitable for most of us to eventually become involved in preparing technical disclosures and patent applications or being asked to analyze the patents of competitors. The patent process is, however, more complex that one might think and this article attempts to provide some additional insights to research engineers.

II. THE PATENTING PROCESS

On the face of it the patent process is quite simple – an inventor describes a new *invention* that is then formulated into a comprehensive legal document with well-specified legal claims. This *patent application* is evaluated by an examiner, who applies due process to determine the validity of the claims of invention. In practice, there are many shades of gray to this process. We will next consider some of these.

A. Disclosures and Applications

As mentioned above, the process begins with a *disclosure* from the engineer who has developed a new *invention*. However with the complexity of today's technologies it is not always clear what, if any, aspect of a new technique or technology is novel and non-obvious. At the same time it is not always desirable to reveal all the details of a new system or process to competitors, particularly when it may not be clear if a strong case for inventiveness can be shown.

There can often be conflicting views at this point between engineering management and the legal department. Attorney's prefer to err on the side of including more information, where as engineering may be reluctant to reveal every details of a process if only limited claims may issue.

At some point, however, a decision is taken and the material in the *invention disclosure* is converted to a *patent application*. The US patent office requires that this document provides a "best effort" description of the underlying invention – as the actual *invention* may not be quite clear at this point the legal team may request additional information from engineering to ensure support for the broadest scope of claims. Conversely, engineering may resist providing some details they are uncomfortable with revealing to competitors.

B. Provisional Vs Non-Provisional Applications

Due to commercial pressures engineers are often pushed to prepare a disclosure document before the final system or process is finalized. This can happen, for example, when some public demonstration of a technology is to be shown. In the US patent office it is acceptable to provide an initial "best effort" description and to later amend this to a more detailed description. However claims date back to the description where they first find sufficient technical support. This first application is known as a *provisional* patent application

While this mechanism appears to be a helpful one to allow the description of an *invention* to be gradually refined it tends not to work out well in practice. Typically a poor and incomplete description is filed and both engineering and legal teams feel the main work is done. Later, as the deadline for completing the application approaches both engineering and legal teams may have moved onto new projects and it becomes challenging to motivate the relevant personnel to focus and provide an improved specification.

A difficulty with partly complete specifications is that descriptive material cannot be added once a final application is filed. The description of the invention is cast in stone. And it is quite common that deficiencies in the descriptive material are only realized as the patent examiner begins to review and analyze the legal claims.

C. The Examination/Prosecution Process

The examination, or prosecution, stage is often the most interesting part of the patenting process. It is during this stage that an examiner from the patent office performs an analysis of the patent claims and then searches the technical and patent literature to determine if there is a basis for some of the claims applied for.

Typically the examiner will initially construct a range of arguments demonstrating that the claims were known, or could be easily deduced by an expert in the field from documents that were available in the technical or patent literature. The company attorney, supported by the inventor(s) and engineering experts within the company will counter argue, often limiting or modifying the scope of some of the original claim set. Claims that are too broad in scope allow the patent examiner to introduce documents that may be only marginally related to the field of the original invention.

The full details of this process are available publicly and can be accessed by obtaining a "file wrapper" via the "public PAIR" section of the US patent office website. PAIR is a

[1] *College of Engineering & Informatics, National University of Ireland Galway*

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

valuable tool to understand how a certain set of claims was eventually granted. Often a claim set that seems very broad in scope is in fact much more limited when considered in tandem with the relevant “file wrapper”.

This prosecution process can actually continue for several years. Even when an application receives a “final rejection” there are appeal mechanisms that allow new arguments to be presented. Naturally, this all takes time and money, and eventually either the patent examiner or the patent attorney will yield. This last aspect explains the lengthy and torturous claims that can be found in some patents. If a prosecuting attorney is tenacious enough, and has a large budget at his disposal, he will usually manage to get some grant of claims.

D. *Granted Patents*

Eventually the dance between patent examiner and attorney will end with one of two results. Either the patent will receive a final rejection or be abandoned in which case no patent will issue; or, more frequently than you might think, a set of patent claims will be agreed by the examiner and a patent will issue.

E. After Grant – Continuations, Continuation-in-Part (CIP) and Divisionals

We explained above that a provisional patent filing can be added to within a 12 month period. But in the US system there also exists a mechanism to continue with a patent even after it has been initially granted. This is known as a continuation and the ideas is that not everything that could be claimed may have been included in the original granted patent. Thus it is possible, for a small fee, to continue the patenting process, after grant, and pursue a different set of legal claims. These must be supported by the original specifications but may differ significantly from the granted claim set.

It is also possible to add material to a granted patent specification - for example, if you modified some aspect of the invention that adds additional value to the original invention then you may file a *continuation-in-part*. Claims priority is only from the date you added the new material, but it may be possible to have some claims which only rely on the original material in which case those claims would have an earlier priority date in line with the original application.

A third form of extension of the original application is known as a *divisional*. This occurs when the examiner considers that your original specification includes 2 distinct inventions. He may limit your original case to one of these distinct concepts, but you have the option to “divide” the case and pursue the 2nd invention at a later stage in what is known as a *divisional* action.

III. . The Value of Patents

Just because a patent issues does not imply that it is a very useful or valuable patent. Indeed many patents, on their own, would have very little value. Patents are generally more useful when combined into inter-related “families” with related or even overlapping claim sets. Such groups of patents are considerably more restrictive for competitors who might be able to “work around” a single patent.

It is also important to consider the commercial logic of the corporation or business that owns a patent. Very few larger companies want to engage in costly litigation and so most view their patent portfolios as tools to engage in business negotiations with competitors. Often patents become part of an overall business strategy and entire families of a portfolio may be sold off if a company decides to leave certain markets.

For smaller companies a patent portfolio is a badge of credibility and enables them to gain business from larger corporations.

IV. Searching the Literature

A word you'll hear a lot in the context of patents is *prior art*. This refers to knowledge and skills that were known and/or practiced prior to your invention. Naturally these should not teach or suggest the new contribution embodied in your patent concept. If they did, then it wouldn't be patentable, would it? As you can imagine, skillful searching of the research literature is a prerequisite before you get to filing, or ideally before you even start to talk to an attorney.

A. Getting original PDFs of a Patent

These are available from the USPTO but only one page at a time. A better source is Google patents or the free utility http:\\www.pat2pdf.org. The full PDF will contain patent drawings which are often very valuable to understand the underlying concepts of an invention and determine how it relates to your own.

B. Text Searching Techniques

Unfortunately the PDF documents that are available generally don't contain the original text. For this reason you'll also find the US patent office website a useful resource as it offers the original text. This can be very helpful when trying to search a long document for key terms or concepts.

C. Other Sources of Prior Art

There are other important places you need to search in order to validate if your concept or idea is patentable. *Google* is an important starting point. *Google Scholar* is an effective tool to cover most of the academic literature - a lot of ideas are initially presented at conferences. Even if your work is better developed and works better, a short paper presenting a similar idea is sufficient to destroy the novelty of your work. I once had an excellent idea for a secure keyboard ruined by a discussion among some techies on a public Internet forum.

In fact it doesn't matter if the idea originally appears in a movie or a work of fiction (think *Science Fiction!*) as long as there is enough detail to pre-empt your new innovation.

V. Concluding Remarks

This is a topic which deserves a more extended treatment than can be provided in a short digest paper. As this article has received a positive response from reviewers I will develop an extended version, most likely for publication in the IEEE Consumer Electronics Magazine in the near future.

AAM-based Face Reorientation Using a Single Camera

Dowan Kim[1], Sungjin Kim[1], Ying Huang [2], Jianfa Zou[2], JunJun Xiong[2] and Jongsul Min[1]
[1]Samsung Electronics, Suwon-si, Gyeonggi-do, Korea
[2]Samsung Electronics, Chaoyang, Beijing, China

***Abstract*-- In a video conference, eye-contact between conferees is an important issue because it makes them feel immersive and friendly. An eye-contact feeling does not mean only parallel gaze angle between conferees. For a real eye-contact feeling, it is necessary that the conferee's face is also perpendicular to the camera so that users feel a face-to-face conversation to the conferee. In this paper, we present a new method to reorient a face for the real eye-contact using a generic 3D face model based on a single camera.**

I. INTRODUCTION

Nowadays, telecommunication is an essential part in our life. Human network is linked complicatedly and enhanced via telecommunication. Development of high-speed internet and wireless communication makes it possible to transfer visual face, as well as voice. People can talk to their friends, family or loved one at the same time seeing their expressions or gestures. In addition, video conferencing has recently activated due to its convenience and cost reduction. However, there is an obstacle for visual telecommunication. The problem is a discomfort feeling causing that the face of a conversation partner in a display does not look natural during visual communication. The discomfort feeling disturbs an immersive conversation and makes a bad effect to their relationship. In the field of computer vision, we call it eye contact issue [1]. It is caused by differences in location of the video conferencing devices that are a camera and a display. In literature, a number of methods have been proposed in order to solve it. These are mainly classified into two categories: mechanical approach and image processing approach. The mechanical approach [2] provides a good performance but it needs extra installation cost, so that the device of video conferencing makes bigger and more expensive. In contrast, the image processing approach [3] doesn't require the extra cost. However, it hasn't shown a satisfactory result. Some researchers are just correcting one's gaze so that they tried to make an eye contact feeling. Recently, others have proposed a hybrid type method [4] using additional cameras. The hybrid method presents a moderate performance but it also causes an additional cost according to the number of the cameras.

For video conferencing, fundamental devices are a camera and a display. Conventionally, a commercial webcam is setup to the top of the display. In this paper, we propose an image processing approach to handle the single top camera issue using a generic 3D face model. Our method can be also expanded to the other camera position. We use a piece-wise linear warping method to generate an eye contacted virtual image.

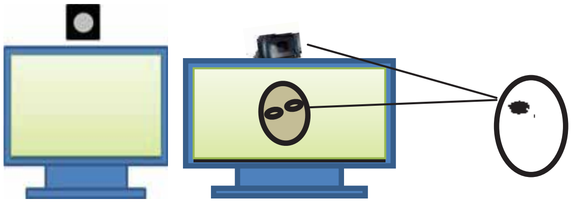


Fig.1. Top-attached video conferencing system and eye-contact failure reason

II. SYSTEM DESCRIPTION

The proposed system is composed of four parts. The first part is a face detection part to track a user's face and fit a generic 3D face model using the user's face features. The second part is the determination warping function to generate a reoriented face. The third part is the texture mapping part so that it procedures an inverse mapping method. The fourth part is the distortion protector part for eliminating artifacts, such as eyeglasses distortion and face distortion, of the image. Our system is implemented at desktop computer aided video conferencing.

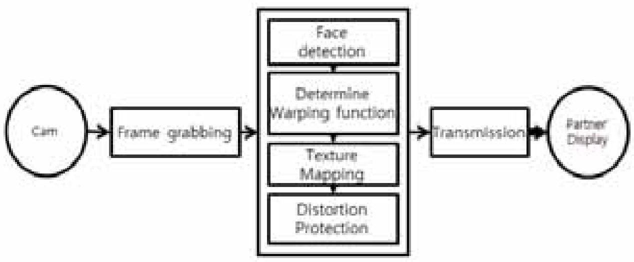


Fig.2. System block diagram

III. ALGORITHM

A frame captured from webcam goes though the face detection part, where the position and orientation of the face is calculated by the active appearance model (AAM) [5] and fitting 3D generic face model to one's face as shown in Fig.3. The desired position and orientation is calculated by the rotation angles with respect to 3D coordinate. Vertices of the fitted 3D generic model are very important for the proposed algorithm, and they are tracked every frame. But, 3D face mesh model is too coarse to describe the face in detail. So, we mix 2D face features and 3D mesh features and make a new mesh as shown in Fig.4. Then, they are handled by control

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

points of the face texture in the correction step. Using three vertices of every mesh triangle, we can extract the affine transformation matrix from the current frame to desired one. Next, texture of the current frame is mapped onto the generated frame at the virtual view based on the affine transformation.

The reoriented face at the virtual view is generated by accumulating all affine transformation, which is often called a piece-wise linear transformation [6]. It is widely used to apply the sophisticated transformation with a low computational burden. Thereafter, a gaze correction technique is applied to emphasize eye contact feeling. It is done by generating an eye-oriented mesh model. The gaze correction consists of two steps. First one is the size-up of middle area of the eye, and second one is the shift-down of tail area of the eye.

After warping process, artifacts may be occurred some part of virtually-generated face, such as chin fattening and eyeglasses distortion, as shown in Fig.4. In order to solve these distortions, we propose a distortion protection algorithm. Chin fattening distortion can be reduced by modification mesh points and the degree of modification is determined by face ratio that is invariant factor to the range. The proposed method maintains pixels near eyeglasses to reduce eyeglasses distortion as shown in Fig.4.

Fig.5. shows the main algorithm flow of the proposed face reorientation.

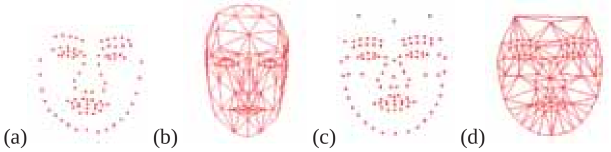


Fig.3. AAM Face Models (a) 2D features, (b) 3D Mesh Model, (c) Mixed features, (d) Delaunay triangulation

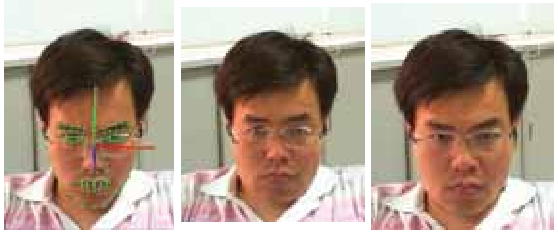


Fig.4. Distortion protection(Left: AAM image, Mid: Chin & eyeglasses distortion, Right: distortion adjusting)

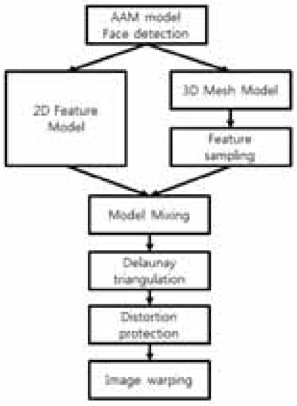


Fig.5. Main algorithm flow chart

IV. EXPERIMENT

We used Logitech Web cam C-910 for video conferencing and OpenCV library for image processing. The proposed method was tested at window 7 OS. In VGA Format, frame rate was 25 fps on Intel core I5 processor with CPU Clock 2.66 GHz. Fig 6. shows the result for face reorientation.

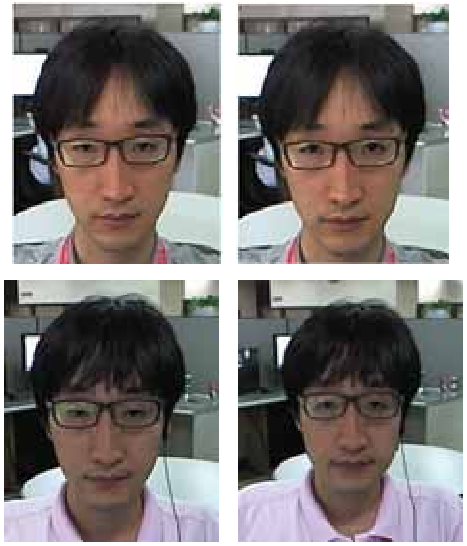


Fig.6. Up: Eye only reorientation, Down: Face reorientation result (Left: before, Right: after)

In order to verify our face reorientation performance, we measured an eye position ratio improvement indicator (EPRII) and head dimension ratio improvement indicator (HDRII)[7]. Our system shows that average EPRII is 0.91 and average HDRII is 0.90 by measurements using 100 frames.

V. CONCLUSION

In this paper, we have implemented a face reorientation algorithm for video conferencing. The proposed method shows good performance despite light computation and without extra cost. Even eyeglasses wearer person can enjoy our application. We expect that our research can be adapted to other devices or situation, for example, mobile, TV, side-mounted webcam, etc.

REFERENCES

[1] Steve McNelly. Immersive Group Telepresence and the Perception of Eye Contact. Nature 407, 477–483 (2000)

[2] CISCO. Cisco TelePresence 3000 . http://www.cisco.com/

[3] Ben Yip, Jesse S. Jin. An Effective eye gaze correction operation for video conference using anti-rotation formulas. Proceedings of the 2003 Joint Conference of the Fourth International Conference. (2003)

[4] Ruigang Yamg. Zjengyou Zhamg. Eye Gaze Correction with Stereovision for Video-Teleconferencing. Pattern Analysis and Machine Intelligence, IEEE Transactions on. Issue 7. 956 - 960 (2004)

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. IEEE TPAMI, 23(6):681–685. (2001)

[6] Seiichi Uchida, Hiroki Sakoe. Piecewise Linear Two-dimensional Warping. Pattern Recognition 2000 Proceedings. 15th International Conference on, Page(s): 534 - 537 vol.3. (2000)

[7] Ben Yip. Eye Contact Rectification In Video Conference With Monocular Camera. PhD Thesis at Sidney University(2007)

A Novel Stereoscopic Image Processing Pipeline

Ja-Won Seo*†, Hae-Sun Lee†, Jong-Hyub Lee†, Sungjun Yim† and Sangbae Park†
*Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea
†DMC R&D Center, Samsung Electronics, Suwon, Korea

***Abstract*—This paper presents a novel stereoscopic image processing pipeline which ensures the consistency of a stereoscopic image pair. Although a stereoscopic image calibration method should resolve both photometric and geometric binocular asymmetries, most previous approaches [1], [2] have mainly focused on the geometric calibration so far. In contrast to this biased research, we propose both photometric and geometric calibration methods which are readily applicable to an embedded pipeline for a stereoscopic camera system. The experimental results demonstrate the performance of the proposed pipeline for calibrating a stereoscopic image pair.**

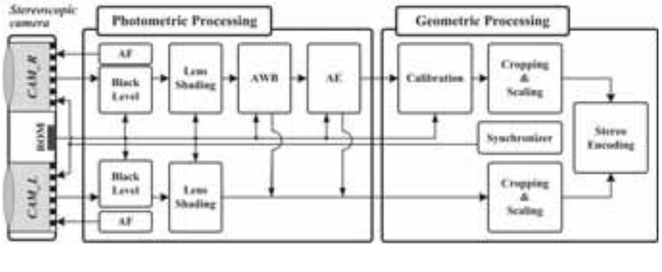


Fig. 1. The proposed embedded stereoscopic image processing pipeline.

I. Introduction

Recently, the deficiency of 3D contents slackens the pace of revolution from 2D to 3D, and some contents from inferior stereo cameras even hamper further progress. Thus, the abundance of user-friendly stereo cameras with acceptable image quality is an essential prerequisite for encouraging people to produce various contents. This paper proposes a practical implementation regarding both photometric and geometric image processing pipelines for stereoscopic camera systems.

II. Proposed Method

Figure 1 illustrates the proposed stereoscopic image processing pipeline which consists of "*Photometric Processing*" and "*Geometric Processing*" parts. In order to alleviate binocular rivalry, the former is in charge of maintaining consistency in photometric perspective, whereas the latter takes charge of reducing geometric mismatches of a stereoscopic image pair. Additionally, notice that our system is asymmetrically configured: several blocks, i.e., AWB, AE and Calibration, in the right camera (i.e., *CAM_R*) pipeline are not present in the left camera (i.e., *CAM_L*) pipeline as shown in Fig. 1. The implementation and benefits of this configuration will be discussed in the following subsections.

A. Photometric Processing

The *Black level* and *Lens shading* blocks compensate the dark current in image sensor and the vignetting effect of optical lens systems respectively. The compensation results for both cameras, i.e., average black level and anti-vignetting gain of each RGB channel, are stored in off-chip memory such as the ROM in Fig. 1 while producing each stereoscopic camera module. Then, they are referred during camera operation time.

The *AWB* (Automatic White Balance) block estimates the color temperature of an ambient illuminant, and the *AE* (Automatic Exposure) block maintains the image brightness to the target brightness automatically. Although both blocks are critical to retain the photometric consistency of a stereoscopic image pair, it is not recommended to integrate them for each camera's pipeline due to following reasons: 1) they require considerable amount of resources from an embedded system and 2) our concern is not the optimization for each camera but the compensation between left and right cameras. To these ends, we integrate the AWB and AE blocks for only the *CAM_R* as shown in Fig. 1, then each result is utilized respectively to approximate that of the *CAM_L* as follows. The proposed AWB block in the pipeline capitalizes on the relative sensitivity ratios between left and right cameras for red and blue channels (α and β), which are computed as (1).

$$\begin{aligned} \alpha &= (G^l_{max}/R^l_{max})/(G^r_{max}/R^r_{max}) \\ \beta &= (G^l_{max}/B^l_{max})/(G^r_{max}/B^r_{max}) \end{aligned} \quad , \tag{1}$$

where for $X \in \{R, G, B\}$, X^l_{max} and X^r_{max} indicate the maximum value of each RGB channel under a test illumination condition for the left and right camera respectively. Therefore, AWB gains for the left camera (i.e., R^l_{gain} and B^l_{gain}) can be approximated from those for the right camera as (2).

$$R^l_{gain} = R^r_{gain} \times \alpha, \ and \ B^l_{gain} = B^r_{gain} \times \beta \tag{2}$$

The proposed *AE* block demands average responsivity of each sensor ($\bar{\eta}_l$ and $\bar{\eta}_r$), which is a slope of the irradiance ($\mathfrak{L}$) against image sensor output (I_o). For two different irradiance levels (i.e., $\mathfrak{L}_1$ and $\mathfrak{L}_2$), we compute the responsivities by measuring corresponding sensor outputs ($I_{o,1}$ and $I_{o,2}$) as (3).

$$\begin{aligned} \bar{\eta}_r &= (I^r_{o,1} - I^r_{o,2})/(\mathfrak{L}_1 - \mathfrak{L}_2) \\ \bar{\eta}_l &= (I^l_{o,1} - I^l_{o,2})/(\mathfrak{L}_1 - \mathfrak{L}_2) \end{aligned} \tag{3}$$

Therefore, the AE gain for the left camera (i.e., A_l) is approximated from that for the right camera as (4).

$$A_l = A_r \times \bar{\eta}_l/\bar{\eta}_r \tag{4}$$

The *AF* (Auto Focus) block finds the focus position of the lens against target subjects. Since it is hardly possible to relate left and right optical systems, the AF is obliged to operate

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

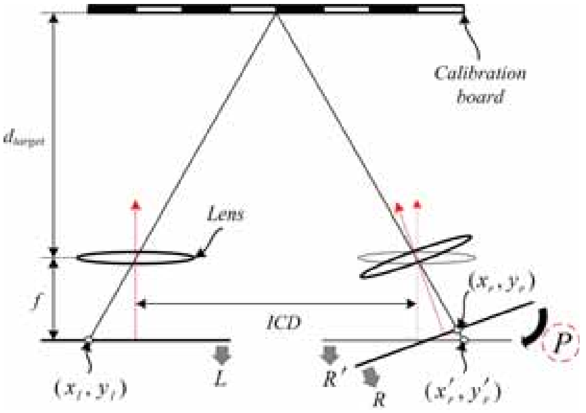


Fig. 2. Geometric calibration using the perspective transform.

individually in the pipeline. Note also that the AF should be performed after halting both AWB and AE algorithms to avoid an abrupt change of contrast in input image.

B. Geometric Processing

Unlike the previous approaches, we introduce a simple calibration method which is devised from a manufacturer's point of view. Figure 2 illustrates the stereo camera geometry, which transforms the right camera image (R) into the calibrated image (R') by applying the perspective transform matrix ($P \in \mathbb{R}^{3\times3}$). Therefore, the corresponding pixels in L and R' can be simply expressed as (5).

$$({x_r}',\ {y_r}') = (x_l + ICD \times f/d_{target},\ y_l)\,, \qquad (5)$$

where *ICD*, f and d_{target} denote respectively the inter-camera distance, focal length and target distance, which are traditionally obtained by estimating intrinsic and extrinsic parameters, but fortunately they are already known to manufacturers with allowable tolerances. Thus, for a given $m(\geq 4)$ corresponding points set, the P can be easily derived using a DLT (Direct Linear Transformation) algorithm [3] as follows, then saved to the ROM to be used during camera operation time.

$$Ah = 0 \quad \overset{SVD}{\Leftrightarrow} \quad U\Sigma V^T h = 0 \quad \Rightarrow \quad h = v_9\ , \qquad (6)$$

where $h = [p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23}, p_{31}, p_{32}, p_{33}]^T$ is the vectorized form of P, $A = [A_1, A_2, \cdots, A_m]^T$, v_9 is the ninth column of orthogonal matrix V after the SVD (Singular Value Decomposition) of A. The i_{th} component of A (i.e., A_i) is computed from the i_{th} corresponding point set as (7).

$$A_i = \begin{bmatrix} x_r^i & y_l^i & 1 & 0 & 0 & 0 & -x_r^i {x_r^i}' & -y_r^i {x_r^i}' & -{x_r^i}' \\ 0 & 0 & 0 & x_r^i & y_l^i & 1 & -x_r^i {y_r^i}' & -y_r^i {y_r^i}' & -{y_r^i}' \end{bmatrix} \qquad (7)$$

TABLE I
COLOR DIFFERENCES OF A STEREOSCOPIC IMAGE PAIR.

		Red ①	**Green ②**	**Blue ③**	**Gray ④**
Indoor	#1	1.7	1.5	0.4	1.8
	#2	2.1	2.2	1.3	1.5
Outdoor	#1	0.8	0.4	0.9	1.6
	#2	1.2	2.7	0.9	1.0

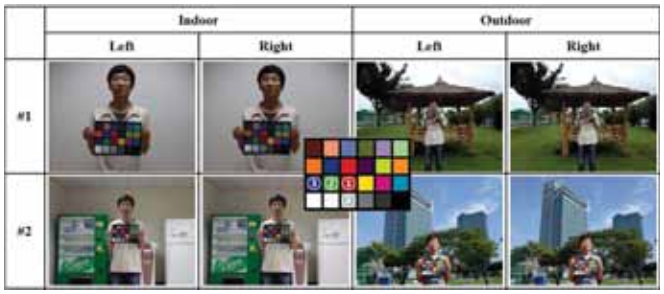
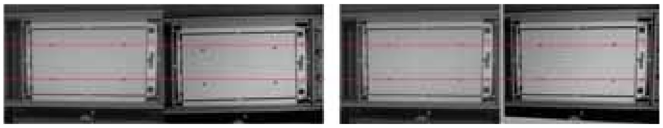


Fig. 3. Demonstration of the proposed photometric calibration.



(a) (b)

Fig. 4. Demonstration of the proposed geometric calibration. (a) before and (b) after the calibration.

The *Cropping & Scaling* block is deployed to adjust the disparity range of a scene and maintain the aspect ratio of an original image. In other words, the resulting crossed disparities from a parallel stereoscopic camera configuration can be expanded to the uncrossed disparity range by horizontal image shifting, and then the changed aspect ratio and resolution of stereoscopic images are compensated by the consecutive cropping and scaling.

III. EXPERIMENTAL RESULTS

Figure 3 verifies the proposed photometric calibration method. In each stereoscopic image pair, the CIE76 color differences of red, green, blue and gray patches in a Macbeth chart are summarized in Table I. In psychophysics, a JND (Just Noticeable Difference) [4] is customarily around 2.3, and most results in Table I are below this limit with enough margin.

Figure 4 also verifies the proposed geometric calibration method. By applying the perspective transform matrix to the right camera image, the calibrated image shows improved row-alignments with the left camera image.

IV. CONCLUSION

In this paper, we propose the overall image processing pipeline for an embedded stereoscopic camera system, which supports both photometric and geometric calibrations for a stereoscopic image pair effectively and efficiently.

REFERENCES

[1] H.-M. Wang, C.-W. Chang, and J.-F. Yang "An effective calibration procedure for correction of parallax unmatched image pairs," *IET Image Processing*, vol. 3, no. 2, pp. 63–74, 2009.

[2] Z.-W. Gao, W.-K. Lin, Y.-S. Shen, C.-Y. Lin, W.-C Kao "Design of signal processing pipeline for stereoscopic cameras," *IEEE Trans. on Consumer Electronics*, vol. 56, no. 2, pp. 324–331, 2010.

[3] R. Hartley, and A. Zisserman, "Multiple View Geometry in Computer Vision," *Cambridge University Press*, 2004.

[4] G. Sharma, "Digital color imaging handbook," *CRC Press*, 2003.

Depth Estimation Based on Blur Measurement for Three Dimensional Camera

Ikhyun Lee, *Student Member, IEEE*, Muhammad Tariq Mahmood, *Member, IEEE*, Seong-O Shim, *Member, IEEE*, Sung-An Lee, and Tae-Sun Choi, *Senior Member, IEEE*

***Abstract*—Depth from Focus (DFF) is the one of the optical methods to estimate depth information. In this paper, we propose a new approach based on blur estimation to obtain depth map. The optimal focused points are selected by maximizing absolute values of deferences in blur signal. The proposed method provides more robust depth information.**

I. Introduction

Depth estimation is an important research field in computer vision. It has numerous applications in consumer electronics and broadcasting. Depth from focus (DFF) [1] is a passive optical method for depth estimation using a sequence of images acquired by translating object in small steps. The objective of DFF is to determine the depth of every point of the object from the camera lens. The location of the best focused point or maximum sharpness provides information to calculate depth. The procedure to compute depth map through DFF technique can be divided into two main steps, (1) computing image focus volume through focus measure operator, (2) applying an approximation technique to enhance the initial results. A focus measure is applied in the small image regions of each image frame in the image sequence to measure image focus quality. The value of the focus measure increases as the image sharpness or contrast increases and it attains the maximum for the sharpest focused image. The maximum focus or sharpness value determines the depth. Thus, an initial depth map is computed by maximizing focus measure along the optical axis. In the second step, an approximation method [2], [3] is applied to enhance the results.

In this paper, we introduce an approach based on blur estimation for measuring the best focused points. The high frequency components contain image details where, low frequency components provides information about image smoothness. The smoothing image due to the blur results in loss of high frequency components. We have observed a significantly difference between gray level values of original image and its smooth version. It means that the gray level values in blurred image have high variation in sharp areas. Whereas variation is low in smooth areas of the blurred images. Figure 1 explains this phenomena and illustrates the main idea behind the proposed algorithm. Figure 1(a) is the original intensity profile and Fig. 1(b) is the blurred signal intensity profile. Figure 1(c) is obtained by taking the absolute value of difference between blurred and original signals. Edge component has high frequency, and it is considered that the best focused pixel has the highest frequency. On the other hand, the flat line is less blurred pixel than edge. Note that the high magnitude of difference between two signals Fig. 1(a) and Fig. 1(b) has significantly high variation than flat area.

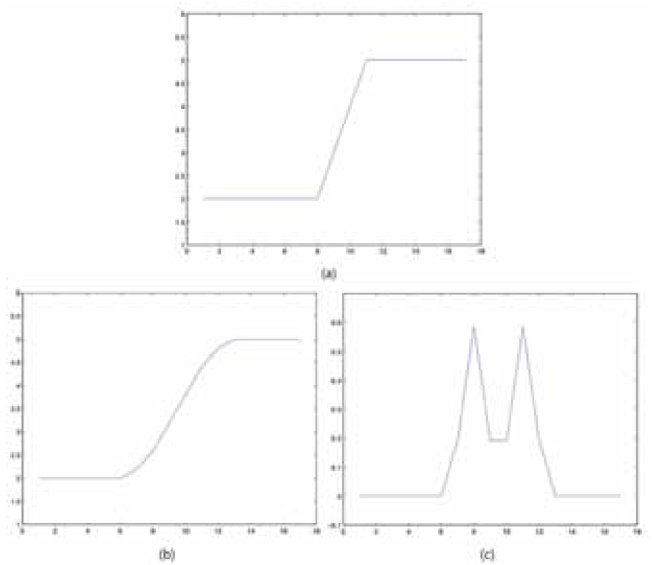


Fig. 1. 1-D illustration of the mechanism of proposed method: (a) Original signal (b) Blurred signal (c) Absolute difference of signal

II. Method

The main objective of DFF is to find the pixels that have maximum focus measure in image volume. The three dimensional image volume is as followed:

$$F_O^z(x,y) = I_z(x,y), \quad z = 1,2,\cdots Z. \tag{1}$$

A general linear filter can be written as:

$$B = w_1 f_1 + w_2 f_2 + \ldots + w_m f_n. \tag{2}$$

where w is the coefficient of an $m \times n$ filter and f is the corresponding image intensities encompassed by the filter. To blur the image volume, we apply the 3×3 averaging filter as

$$B = \frac{1}{9}\sum\nolimits_{i=1}^{9} w_i f_i. \tag{3}$$

where $w_i = 1, i = 1,2,3,\cdots 9$. A blurred image volume is obtained by convolving the original image volume with average filter.

$$F_B^z(x,y) = B * I_z(x,y), z = 1,2,\cdots Z. \tag{4}$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Now, the original image volume is compared with the blurred image volume to obtain the edges.

$$F_{AD}^{z}(x,y) = |F^{z}(x,y) - F_{B}^{z}(x,y)|, z = 1,2,\cdots Z. \qquad (5)$$

In order to improve the robustness for weak-texture images and noise, sum of values is calculated in a local window around (x, y).

$$F'^{z}_{AD}(x,y) = \sum_{x=i-1}^{i+1} \sum_{y=j-1}^{j+1} F^{z}_{AD}(x,y), \\ z = 1,2,\cdots Z. \qquad (6)$$

The best focused pixel has the maximum value in z-direction. The sharpest pixels in the focus volume provide the depth map $D(x, y)$ the object i.e.,

$$D(x,y) = \arg\max_{z} F'^{z}_{AD}(x,y). \qquad (7)$$

III. Experimental Results

In order to evaluate the performance of the proposed method, we use the images for simulated cone object with conventional methods such as SML [4], GLV [5], [6]. The simulated cone has been selected for the experiments because it is easy to verify the results for such an object with known data depth map. Besides, the planar images and real cone objects are used for experiments. Noise sensitivity and computational complexity are important factors for focus measure evaluation. Two statistical metrics mean square error (MSE) and correlation are used for quantitative analysis. Higher the correlation is closer to original image and smaller value of the MSE indicates a higher precision.

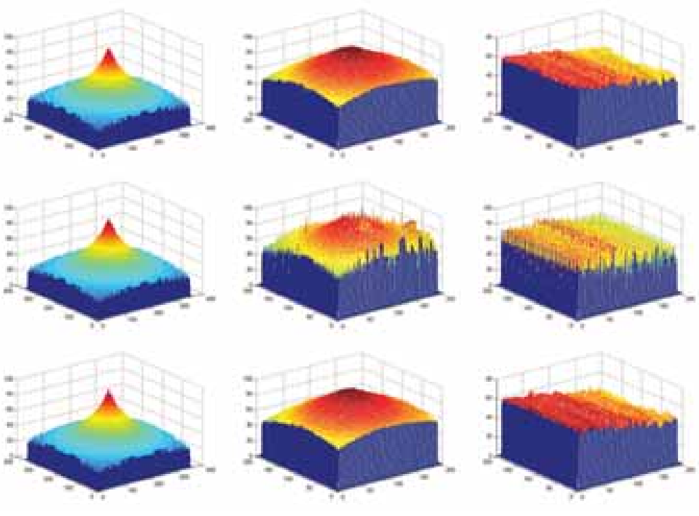


Fig. 2. Depth maps for simulated cone (left column), real cone (central column), planar (right cloumn) objects with SML (first row), GLV (second row), proposed method AD (bottom).

Figure 2 shows the comparison between conventional methods as SML, GLV and proposed method. Figure 3 shows the robustness of different methods against noise. The proposed method is more robust than traditional methods.

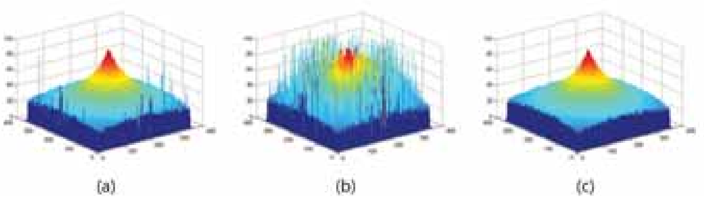


Fig. 3. Depth maps with speckle noise with zero mean and 0.04 variance: (a) SML, (b) GLV, (c) proposed method.

TABLE I
Performance comparison for DFF methods

Noise	SML	GLV	AD
(a) MSE			
Gaussian(0.01)	370.517	202.9424	174.9498
Salt&Pepper(0.05)	324.39	603.8847	250.788
Speckle(0.04)	54.2895	75.3873	53.6729
No noise	54.4975	55.8295	54.6544
(b) Correlation			
Gaussian(0.01)	0.3773	0.5747	0.615
Salt&Pepper(0.05)	0.419	0.1937	0.5003
Speckle(0.04)	0.9396	0.8619	0.942
No noise	0.9358	0.9350	0.9393

Table I(a) shows the MSE comparison of DFF methods with various noises. Proposed method (AD) has the lowest value for both no noise and with noise cases. Table I(b) shows the performance comparisons in terms of correlation. It can be observed that the proposed method has provided considerable robustness against noise than conventional measures.

IV. Conclusion

In this paper, we have proposed focus measure using blur estimation. The proposed algorithm is robust and fast for depth estimation. The experimental results have demonstrated the robustness of the proposed method.

Acknowledgment

This work was supported by the National Research Foundation (NRF) grant funded by the Korean government (MEST) (No. 2012-0001344). This work (2012-0005542) was supported by the Mid-career Researcher Program through a National Research Foundation (NRF) grant funded by the Ministry of Education, Science and Technology (MEST), Korea.

References

[1] A. Malik and T. Choi, "Application of passive techniques for three dimensional cameras," *Consumer Electronics, IEEE Transactions on*, vol. 53, no. 2, pp. 258–264, 2007.

[2] M. Ahmad and T. Choi, "A heuristic approach for finding best focused shape," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 4, pp. 566–574, 2005.

[3] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, no. 3, pp. 266–274, 1995.

[4] S. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 824–831, 1994.

[5] F. Groen, I. Young, and G. Ligthart, "A comparison of different focus functions for use in autofocus algorithms," *Cytometry*, vol. 6, no. 2, pp. 81–91, 1985.

[6] T. Yeo, S. Ong, R. Sinniah *et al.*, "Autofocusing for tissue microscopy," *Image and vision computing*, vol. 11, no. 10, pp. 629–639, 1993.

Visual Quality Improvement for Hybrid 3DTV with Mixed Resolution Using Conditional Replenishment Algorithm

Kyeong-Hoon Jung, Min-Suk Bang, Sung-Hoon Kim, Hyun-Gon Choo and Dong-Wook Kang

Abstract--**This paper proposes the conditional replenishment algorithm (CRA) to improve the visual quality of hybrid stereoscopic 3DTV where the spatial resolutions of left and right views are mismatched. The basic concept of CRA is to determine the better replacement from disparity compensated view and simply enlarged view in generating the enhanced right view. The simulation results show that the proposed CRA can successfully improve the objective quality of the poor view and give positive effect on the final 3D visual quality.**

I. INTRODUCTION

The 3DTV has been considered as one of major candidates for the next broadcasting services and many approaches are being suggested to initiate 3D broadcasting service. The most familiar examples are frame-compatible or service-compatible methods using single channel. Also, there have been several hybrid 3DTV methods which use two channels for transmission of left and right views. Kim et. al. have proposed a new hybrid 3DTV system which can be serviced via ATSC-M/H [1]. The block diagram of the hybrid 3DTV system is shown in Fig. 1. The left and right views are transmitted through DTV and mobile channels, respectively. If we have ATSC-M/H receiver with dual codec, both left and right views are decoded and combined to generate 3D content. The principal merit of this system is to guarantee the compatibility with existing services. That means it can provide 2D HDTV and mobile 2DTV without sacrifice in bandwidth.

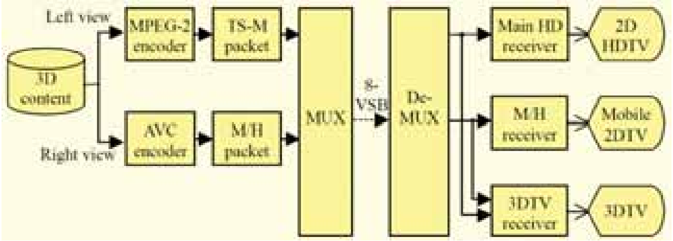


Fig. 1. The block diagram of ATSC M/H-based hybrid 3DTV system [1].

The quality mismatch may be an issue because the left view has resolution of 1080p but the right view has resolution of 240p or 480p. The binocular suppression theory describes the resolution mismatch issue and says it is the quality of better view what is critical for the overall 3D visual quality [2]. And the proposed system could provide quite satisfactory quality for many kinds of sequences even though there is a considerable mismatch in resolutions.

However, the quality of poor view should not be overlooked especially when the resolution gap is too high or there exist severe coding artifacts. Thus the smaller right view needs to be improved to get satisfactory 3D visual quality.

Scalable video coding (SVC) or multi-view video coding (MVC) can be thought as a method to improve the small view. But it is not desirable to directly use these familiar coding algorithms. Firstly, an additional complex SVC or MVC codec is required at both encoder and decoder. Also the amount of additional data due to SVC or MVC is not insignificant since they are kinds of residual coding. Moreover, these methods do not fully use the available data, that is, the high quality left view is not considered in SVC and the small right view is not considered in MVC. In this paper, we propose an effective algorithm to enhance the quality of small right view.

II. CONDITIONAL REPLENISHMENT ALGORITHM

In order to display the hybrid 3D content, the small-sized right view needs to be enlarged to the same size as the HD left view. The intra view approaches, such as bilinear interpolation, can be considered as simple solution, but it does not use any information from left view. It is obvious that there is much correlation between two views of 3D contents and a pixel in one view usually has its corresponding position in another view. If we use the information of disparity, we can expect the quality of enlarged right view is greatly enhanced. Meanwhile, there exist some areas where the proper disparity is not available like shadow or occlusion region. Thus the mode information is also required to indicate whether the disparity from left view is reliable or not.

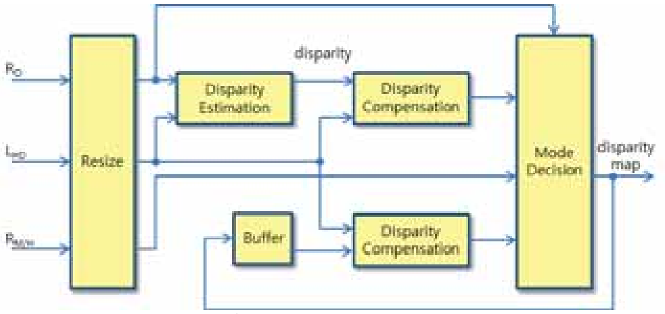


Fig. 2. The block diagram of conditional replenishment algorithm.

The block diagram of conditional replenishment algorithm (CRA) is shown in Fig.2. Firstly, the disparity vector between original right and the encoded left views is estimated. Next, the disparity compensated right view is generated and the small right view is interpolated to have the same size as left view. Also the previous disparity map is used for the generation of the third candidate to reduce the temporal

This work was supported by the IT R&D program of MKE/KCC/KEIT/ETRI [Development of ATSC 8-VSB based Single Channel Broadcasting System Technology Providing HD 3DTV, 2D HDTV, and Mobile 2DTV Services, 10041901].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

redundancy. The final step is to determine the mode which conveys information about the best candidate and to produce the disparity map.

There are two types in mode decision; one is intra and the other inter. In intra type, the mode is decided by comparing the disparity compensated right view and the simply interpolated right view. In inter type, the third mode of compensation with the previously buffered disparity map is also considered.

Meanwhile, the size of processing block for disparity estimation may be either fixed or variable. The variable-sized block is preferred by observing that there is much spatial correlation in disparity map. Thus we adopt the quad-tree structure to describe the mode distribution of disparity map. And the exponential Golomb code is used to encode the disparity.

III. SIMULATION RESULTS

We used several 3D video sequences with various characteristics for simulation. The color format of original sequence is 4:2:0 and its vertical resolution is 1080p. In the proposed hybrid 3DTV system, the left view is transmitted through DTV channel for the conventional 2D HDTV service and encoded by MPEG-2 MP@HL with 12 Mbps. Meanwhile the right view is transmitted through M/H channel for the mobile service. Thus the original right view is resized to the vertical resolution of 240p and encoded by H.264 with 480 kbps.

The Fig. 3 shows the typical pattern of mode distribution in disparity map. The colored blocks denote that the disparity compensated view is selected, and the empty blocks are replaced by the simply enlarged view. It can be noticed that the blocks are successfully partitioned.

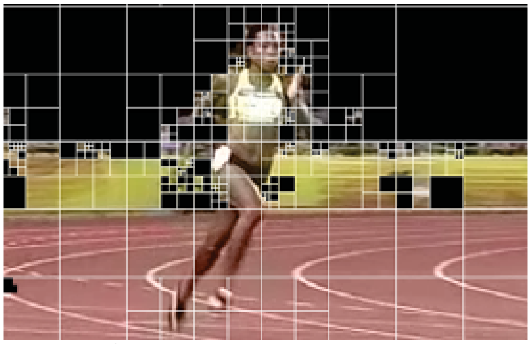


Fig. 3. An example of mode distribution in disparity map.

We analyzed the objective and subjective quality gain of CRA in many cases. But only Fig. 4 is given in this summary due to the page limit. We can notice that considerable PSNR gains are obtained by CRA. With a small amount of disparity map, that is, data of about 120 kbps, the PSNR was improved over 3dB. As expected, the PSNR gain depends on the amount of additional information. And the case of variable block outperformed that of fixed block. By using variable block, BD-PSNR [3] increased by 0.66dB and BD-rate reduced to 32.8%.

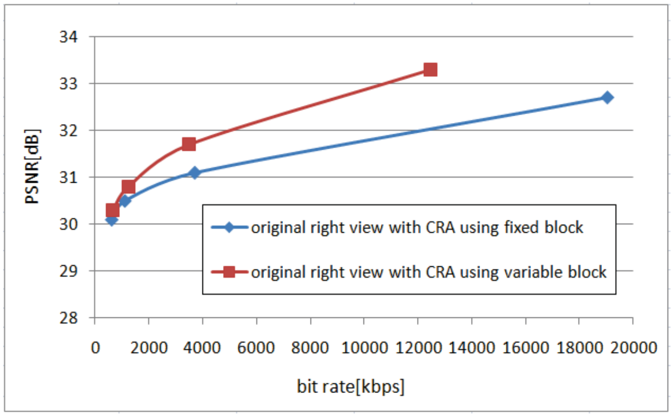


Fig. 4. The R-D curves of the enlarged right view by CRA (PSNR of enlarged right view: 26.81dB).

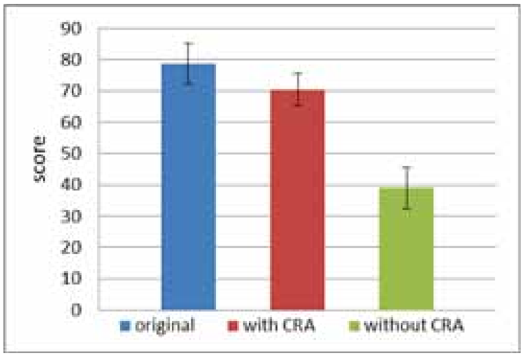


Fig. 5. The DSCQS [4] scores.

Meanwhile, we evaluated the subjective quality of final 3D sequence and a sample DSCQS scores are given in Fig. 5. This result also shows the effectiveness of CRA. On the other hand, in case of using fixed-sized block, there are many blocky errors and it directly has a bad effect on the 3D visual quality.

IV. CONCLUSIONS

The ATSC M/H-based hybrid 3DTV system has noticeable advantages over another competitors but the issue of quality mismatch need to be clearly answered. In this paper, an effective algorithm to improve the visual quality of the above hybrid 3DTV system is proposed. The basic idea is to determine the better substitute from the disparity compensated view and the simply interpolated view. This algorithm doesn't require any complex codec since we don't need to encode the residual data. And it produced considerable quality gain both objectively and subjectively with small amount of additional information.

SELECTED REFERENCES

[1] Byung-Yeon Kim et al., "A study on feasibility of dual-channel 3DTV service via ATSC-M/H," *ETRI Journal*. vol. 34, no. 1, pp. 17-23, Feb. 2012.

[2] L. Stelmach et al., "Stereo image quality: effects of mixed spatio-temporal resolution," *IEEE Trans.on Circuits Syst. Video Technol*, vol. 10, no. 2, pp. 188-193, March. 2000.

[3] G. Bjontegaard, "Improvements of the BD-PSNR Model," ITU-T SG16/Q6, VCEG-AI11, July. 2008.

[4] Recommendation ITU-R BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, 2002.

Analysis on the susceptible level of 3D crosstalk and Its effect on the 3D nausea

Hee-Jin Choi[1], Minyoung Park[1], Joohwan Kim[2], Jae-Hyeung Park[3], and Sung-Wook Min[4]
[1]Department of Physics, Sejong University, South Korea
[2]School of Optometry, University of California, Berkeley, USA
[3]Department of Electrical and Computer Engineering, Chungbuk National University, South Korea
[4]Department of Information Display, Kyung Hee University, South Korea

***Abstract*—The 3D crosstalk is one of the major problems of the current 3D displays. In this paper, the susceptible level of 3D crosstalk is researched through subjective experiments using a hafloscope.**

I. Introduction

Recently, the rapid progress in the flat panel display technologies opens a new era of three-dimensional (3D) displays that consumers can access easily. However, as the number of consumers who have an experience to use a 3 D device, complains about the 3D crosstalk are also growing. The 3D crosstalk is a phenomenon which the observer sees a mixture of the left-eye and the right-eye images. Since the 3D crosstalk can disturb the recognition of 3D image, the 3D display manufacturers are trying to find a way which measures and reduces the 3D crosstalk [1-5].

Regarding that most of the existing 3D displaying devices have some amount of 3D crosstalk, it is important to research about a susceptible level of 3D crosstalk. However, it is hard to control the level of 3D crosstalk since it is related with various hardware parameters of the system and mechanical modifications of the system structure is also required to change it. Therefore, it is needed to find a way to control the level of 3D crosstalk electrically. In this paper, a method using a hafloscope which composes of two display panel and a folded mirror is proposed to control the 3D crosstalk to find a susceptible level of it.

II. Principles

Since the existing 3D display requires an additional hardware such as 3D glasses or a lenticular array, the 3D crosstalk can be affected with the properties of the above hardware. Therefore, it is required to eliminate additional devices between the 3D image and the observer. The hafloscope is one of the good alternatives to realize the 3D image and control the level of the 3D crosstalk as well as and the cubic effect also. The principle of the hafloscope is to enforce the observer to see a different left-eye and right-eye images on different display devices through the folded mirror. The disparity can be controlled by changing the location of the images and the level of the 3D crosstalk also can be changed with the grayscale of them. The structure of the hafloscope used in the research is shown in Fig. 1.

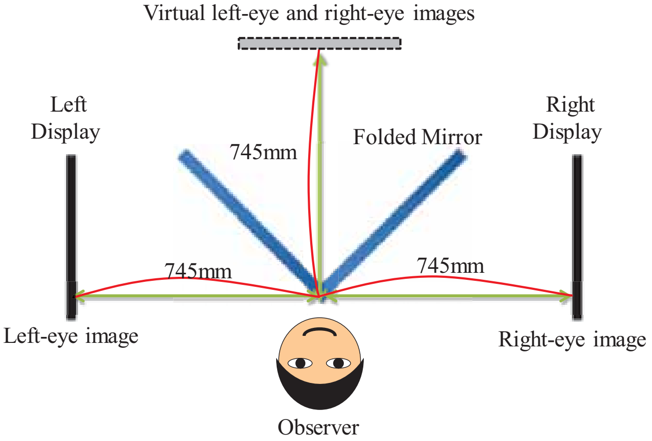


Fig. 1. The structure of the hafloscope

With the above experimental setup, it is possible to control the level of the 3D crosstalk and the cubic effect without any additional distortion from the additional devices such as 3D glasses or an array of optical elements. Therefore, the susceptible level of the 3D crosstalk and the effect of it to another human factor such as the 3D nausea can be researched.

In this paper, the relation between the level of the crosstalk and the 3D nausea is analyzed by measuring the maximum disparity which the observer can endure the accommodation-convergence mismatch in various condition of different 3D crosstalk levels.

III. Experimental Results

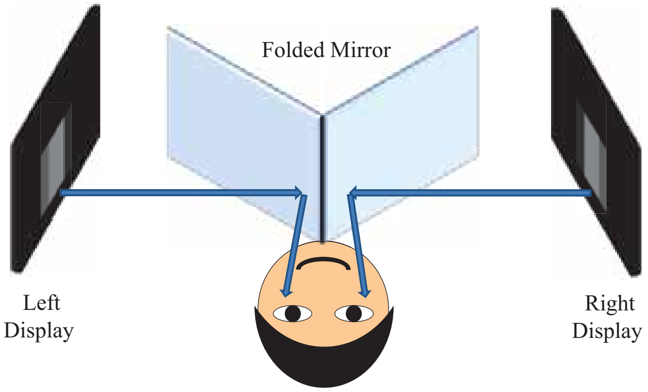


Fig. 2. The experimental setup

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

In the experiment, a 3D image with two kinds of different crosstalk is used. The first one is called bright or white crosstalk since it changes the luminance of the displayed image higher. The second one is called dark or black crosstalk due to the inverse reason of the above one(decreasing the luminance). The experiments were performed by comparing the condition without 3D crosstalk with the others. There were four observers who attend the experiment. While the 3D crosstalk varies, the observers were asked to notify the maximum disparity that they can combine the left-eye and the right-eye images to form the 3D image.

At the first stage, the observer sees a 3D image without any 3D crosstalk. Then, the disparity between the left-eye and the right-eye images is increased by moving them. If the observer appealed a severe 3D nausea due to the excessive disparity, the experiment was stopped and the operator recorded the maximum disparity that the observer could endure. The second stage had same procedure with the previous one except that the level of 3D crosstalk slightly increases to identify the susceptible level of it. Since the level of the 3D crosstalk is controlled by changing the grayscale of the displayed image, the level of 3D crosstalk realized in the experiment has discrete values. At the third stage, the observers were asked to watch a 3D image with large 3D crosstalk and the maximum disparities were also recorded. The experimental setup of the second and the third stage is shown in Fig. 2.

Table 1. The experimental results with the black crosstalk

Subject / Black Crosstalk	Point of Convergence (Diopter)			
	1	2	3	4
None	3.81	4.23	4.39	4.22
5.6%	3.83	4.13	4.32	4.20
8.4%	3.86	4.04	4.44	4.28
11.1%	3.98	4.20	4.37	4.32
12.7%	4.04	3.99	4.44	4.18
14.4%	4.03	4.15	4.49	4.20
20.8%	3.91	4.11	4.39	4.25
29.4%	3.95	4.08	4.39	4.10
40.2%	3.91	4.15	4.40	4.34

▨ : Observer's Notification of recognized black crosstalk

The experimental results are summarized in Table 1 and 2. The four observers notified their first recognition of the black crosstalk around a level of 10%. The maximal levels of the cubic effects which the observers can endure were similar with that of 3D image without black crosstalk. However, in the case of the white crosstalk, a level below 0.01% was recognized. Additionally, the observers also reported the 3D nausea with farther point of convergence if the white crosstalk is increased. Therefore, it can be derived that the observers are more sensitive with the white crosstalk than the black crosstalk.

Table 2. The experimental results with the white crosstalk

Subject / White Crosstalk	Point of Convergence (Diopter)			
	1	2	3	4
None	3.75	4.30	4.32	4.11
0.02%	3.79	4.32	4.30	4.09
0.04%	3.79	4.34	4.37	4.04
0.06%	3.59	4.28	4.35	3.90
0.08%	3.53	4.23	4.28	3.94
0.10%	3.34	4.25	4.20	4.01
0.12%	3.32	4.25	4.20	3.96
0.22%	3.13	4.06	4.20	3.89
0.35%	3.23	4.04	4.13	3.87
0.66%	2.99	3.87	3.96	3.75

▨ : Observer's Notification of recognized white crosstalk

IV. Discussion

The susceptible level of 3D crosstalk and its effect on the 3D nausea has been measured and analyzed through experiments using a hafloscope. The results of this research can be helpful to realize a 3D displaying device with optimized performance.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2012-0003578).

References

[1] S. Shestak, D. Kim, S. Hwang, "Measuring of gray-to-gray crosstalk in a LCD based time-sequential stereoscopic display," SID Int. Symp. Digest Tech. Papers, 132-135 (2010).

[2] S. S. Kim, B. H. You, H. Choi, B. H. Berkeley, and N. D. Kim, "World's first 240Hz TFT-LCD technology for Full-HD LCD-TV and its application to 3D display", SID Int. Symp. Digest Tech. Papers 424-427 (2009).

[3] D. –S. Kim, S. –M. Park, J. –H. Jung. And D. –C. Hwang, "New 240Hz driving method for Full HD & high quality 3D LCD TV," SID Int. Symp. Digest Tech. Papers, 762-765 (2010).

[4] D. Kim, J. Lee, T. Kim. S. Moon, "Method to reduce the cross-talk in 3D PDP TV," IMID 2009 Digest, 513-516 (2009).

[5] B. Lee, I. Ji, S. Han, S. Sung, K. Shin, J. D. Lee, B. H. Kim, B. H. Berkeley, and S. S. Kim, "Novel simultaneous emission driving scheme for crosstalk-free 3D AMOLED TV," SID Int. Symp. Digest Tech. Papers 758-761 (2010).

Scalable ECG Transmission to Improve the Diagnosability of Remote Patient

Yongwoo Cho[†], Junhee Ryu[†], Juyoung Park[*], Jaemyoun Lee[*], Heonshik Shin[†], Kyungtae Kang[*]
[*]Department of Computer Science and Engineering, Hanyang University, Korea
[*]School of Computer Science and Engineering, Seoul National University, Korea

***Abstract*—We present an adaptive framework for layered representation and transmission of electrocardiogram (ECG) data that can accommodate a time-varying wireless channel. The representation, combined with the layer-based earliest deadline first (LB-EDF) scheduler, ensures that the perceptual quality of the reconstructed ECG signal does not degrade abruptly under severe channel conditions and that the available bandwidth is utilized efficiently. Simulation shows that the proposed approach significantly improves the perceptual quality of the ECG signal reconstructed at the remote monitoring station.**

I. Introduction

In this paper, we present an adaptive framework to support high-quality remote electrocardiogram (ECG) monitoring over error-prone wireless networks. Our proposed adaptive framework consists of a layered representation of ECG data and an error control scheme based on automatic repeat request (ARQ) combined with a layer-based earliest deadline first (LB-EDF) scheduler.

The LB-EDF scheduling algorithm support the delivery of scalable ECG streaming over lossy channel in real-time. Scalable ECG streaming data have timing constraints because of their sensitivity to delay and jitter, and thus, the use of the EDF policy has the critical advantage of ensuring that higher (less important) enhancement layer(s) (EL(s)) can be discarded so that the base layer (BL) and the lower enhancement layers have a greater chance of arriving at the remote monitoring station (RMS) on time. Working in conjunction with the ARQ scheme, the proposed LB-EDF scheduler greatly improves the signal readability at the RMS by rescheduling the packets such that the more important lower-layer packets are transmitted first. This ensures *Graceful quality degradation* and efficient use of the bandwidth in a way that maximizes the perceptual quality and the resulting ECG readability, thus facilitating a correct diagnosis.

II. System architecture

A wearable wireless ECG sensor (also called electrode) continuously measures the heart activity of a mobile patient. The resulting digital stream is grouped into packets that then transmitted wirelessly to remote healthcare professionals in real time through a nearby access point [1].

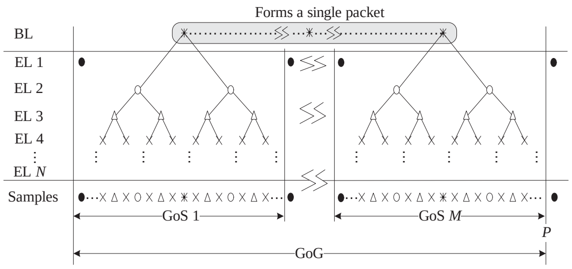


Fig. 1. Framework for layered temporal scalability, and packetization for transmission over the wireless channel.

III. Representation of ECG data

In general [2], an η-lead ECG is one in which η different electrical signals are recorded almost simultaneously, and it is often used as a one-off recording of an ECG. If η leads are recorded, and if the ECG output for each lead is digitized at a rate of R samples per second, each of which has a resolution of L_{smpl} bits, the resulting data-rate μ_{ecg} of the wireless ECG application is given as $\mu_{\text{ecg}} = \eta R L_{\text{smpl}}$. The digital stream is packetized and then sent to an RMS over a wireless channel.

It is clear from this definition that the quality of the obtained ECG signal improves with an increase in the sampling frequency. In a standard environment, scalability is achieved through a layered structure, where the ECG information is divided into two or more discrete bit streams corresponding to different layers, as shown in Fig. 1. The BL ECG stream contains fundamental ECG information that is periodically sampled at a low frequency. The EL contains ECG data sampled at higher frequencies in different time domains to produce the expected scalability.

Temporal scalability involves the partitioning of a group of samples (GoS) into a single BL and multiple ELs. Samples at the center of each GoS are packed into a single BL packet according to their sequence numbers; then this BL packet is transmitted with the highest priority. Assuming that the size of a packet and that of each ECG sample is L_{payload} and L_{smpl}, respectively, constructing a single BL packet requires $M = L_{\text{payload}}/L_{\text{smpl}}$ GoSs, and the interval of M GoSs is known as the period P. A layered structure for representing ECG data with N ELs is shown in Fig. 1.

Now, let S_0 and S_n $(1 \leq n \leq N)$ be a set that contains ECG samples corresponding to the BL and the nth EL, respectively; $\sigma_{i,j}$ is the ith ECG sample of the jth GoS. Then, S_1 includes

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012R1A1A2044653)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

the first ECG sample from each GoS; thus, it is defined as

$$S_1 = \{\sigma_{1,j} | j = 1, 2, 3 \ldots\}. \quad (1)$$

The BL sample in the jth GoS is located at $(\sigma_{1,j} + \sigma_{1,j+1})/2$. Next, the set S_2 contains two elements from each GoS, and it can be defined as follows:

$$S_2 = \left\{ \frac{\sigma_{1,j} + \frac{\sigma_{1,j}+\sigma_{1,j+1}}{2}}{2}, \frac{\sigma_{1,j+1} + \frac{\sigma_{1,j}+\sigma_{1,j+1}}{2}}{2} \right\}. \quad (2)$$

The first element of the set S_1 and S_n $(n \geq 2)$ corresponds to the $(2^{(N-n)}+1)$th ECG sample and the first ECG sample, respectively, whereas the interval between two consecutive elements of S_n is 2^{N-n+1}. Therefore, set S_n can generally be defined as follows when n is greater than one:

$$S_n = \{\sigma_{2^{(N-n)}+1,1} + 2^{N-n+1}k | k = 1, 2, 3 \ldots\}. \quad (3)$$

The number of samples in the BL $|S_0^g|$ and in the nth EL $|S_n^g|$ $(n \geq 1)$ in a period P is respectively defined as follows:

$$|S_0^g| = M, \; |S_n^g| = 2^{n-1}M \; (n \geq 1). \quad (4)$$

Now, the total number of samples R_g in a period P is

$$R_g = M \sum_{n=0}^{N} |S_n^g| = 2^N M. \quad (5)$$

As a result, the sampling frequency in the BL and in the nth EL is $R_B = M/P$ and $R_E^n = 2^{n-1}M/P$, respectively, where $R = R_B + \sum_{n=1}^{N} R_E^n$ and $P = R_g/R$.

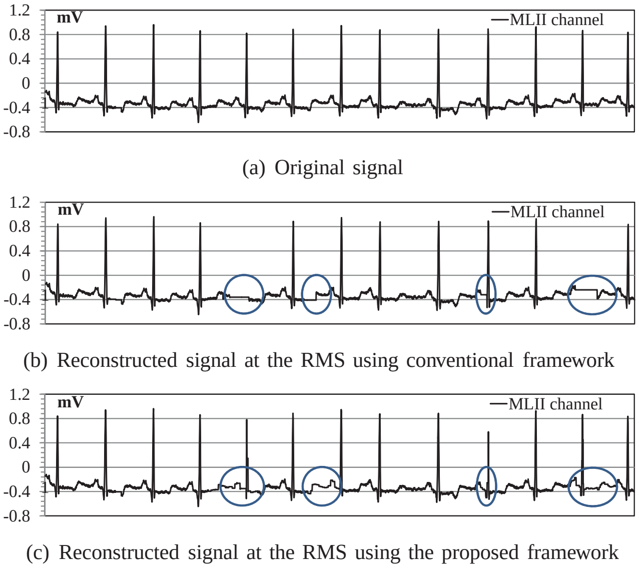
IV. ARQ-based error control using LB-EDF

Owing to the proposed layered representation of ECG data, it is intuitive to consider relative “importance” of the data in the scalable ECG stream in order to avoid an abrupt degradation in the quality of the ECG signal. The loss of consecutive ECG symbols has a greater effect on the ECG signal than the loss of a few random symbols. Therefore, it is desirable to prioritize the delivery of packets in the BL or lower ELs, even under severe channel conditions. For this purpose, we assign higher priority to packets in the lower layer, these can then be transmitted earlier, with a greater opportunity for retransmission in the case of loss. Packets in the same layer are served according to EDF policy. The scheme improves bandwidth utilization and the readability of the ECG signal in the case of some data loss via the prioritization of the low-frequency data in the BL or lower ELs.

V. Performance of wireless ECG transmission

In the simulation, we set the packet size to 512 bits; each packet contains a maximum payload of 490 bits and a packet header of 22 bits. The fundamental timing unit for packet transmission is set to 1.67 ms, as per the CDMA2000 1xEV-DO Revision A standard [3]. The transmission of a packet requires one time-slot, and the resulting reference channel data-rate is 307.2 kb/s.

The relative advantage of our framework can clearly be seen in Fig. 2, which depicts a snapshot of the original ECG signal obtained from patient and that of the corresponding signal reconstructed in the RMS when the channel error rate is 0.1 and the patient moves at 2 km/h (channel errors were modeled using the threshold model suggested by Zorzi et al. [4]). It is observed that compared to the original ECG signal in Fig. 2(a), the ECG signal reconstructed with conventional transmission framework (CTF), which serially packetizes consecutive symbols in order, frequently omits important ECG information; this might lead a physician to misinterpret a patient’s condition. However, for the same pattern of error in the wireless channel, the perceived quality of the reconstructed signal degrades very gracefully in our framework, as shown in Fig. 2(c), with the help of layered representation and by selectively recovering packets with higher priority. *This provides the physician with a better chance of arriving at an accurate diagnosis.*



(a) Original signal

(b) Reconstructed signal at the RMS using conventional framework

(c) Reconstructed signal at the RMS using the proposed framework

Fig. 2. Snapshot of ECG signal fluctuations for MLII channel when the channel error rate is 0.1 and the patient moves at 2 km/h.

VI. Conclusions

The proposed adaptive framework can effectively limit the effect of error bursts that are commonly occur in a wireless channel, hence ensures that the *perceptual quality degrades gracefully under severe channel conditions*.

References

[1] K. Kang, K.-J. Park, J.-J. Song, C.-H. Yoon, and L. Sha, “A Medical-Grade Wireless Architecture for Remote Electrocardiography,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 260–267, Mar. 2011.

[2] D. Cypher, N. Chevrollier, N. Montavont, and N. Golmie, “Prevailing over Wires in Healthcare Environments: Benefits and Challenges,” *IEEE Communications Magazine*, vol. 44, no. 4, pp. 56–63, Apr. 2006.

[3] N. Bhushan, C. Lott, P. Black, R. Attar, Y.-C. Jou, M Fan, D. Ghosh, and J. Au, “CDMA2000 1xEV-DO Revision A: A Physical Layer and MAC Layer Overview,” *IEEE Communications Magazine*, vol. 44, no. 2, pp. 37–49, Feb. 2006.

[4] M. Zorzi, R.R. Rao, and L.B. Milstein, “Error Statistics in Data Transmission over Fading Channels,” *IEEE Transactions on Communications*, vol. 46, no. 11, pp. 1468–1477, Nov. 1998.

Automatic Waist Airbag Drowning Prevention System Based on Underwater Time-lapse and Motion Information Measured by Smartphone's Pressure Sensor and Accelerometer

Mohamed Kharrat, Yuki Wakuda, *Member, IEEE*, Noboru Koshizuka, *Member, IEEE,*
and Ken Sakamura, *Fellow, IEEE,*

***Abstract*—We propose an automatic airbag system helping evacuating swimmer who is motionless or spent abnormally long time underwater. The system is composed of a customized waist airbag and Smartphone equipped with pressure sensor and accelerometer. The Smartphone application will try to define accurately the position of the swimmer's head the whether underwater or not by comparing the real-time measured pressure to an estimated value of over the water surface pressure. If the time-lapse spent by the swimmer exceeds a predefined maximum threshold or the measured accelerometer information shows that he has been motionless for long time period underwater, an on device alarm is triggered and a signal is sent to the servomotor connected to the deflation system of the airbag to trigger it and evacuate the swimmer.**

I. Introduction

Drowning is the third cause of unintentionally injury death in the world [1]. In United States it is considered the second cause of death among children under 12 [2]. Worldwide, children under 5 have the highest drowning mortality rate [1]. Drowning accidents occurs even in swimming pool staffed with professional lifeguards. It is very hard for normal people to identify people drowning. Parents are required to watch permanently their children while they are swimming in the pool. However, this is practically difficult to ensure, as humans lose easily their attention. Drowning is considered silent and rapid death . In less than one minute a person can drown silently without being able to call for help. There are few commercialized wearable drowning prevention system. SenTAG [9] is a wrist band based system which triggers an alarm if the swimmer is motionless for twenty seconds under a certain depth. WAHOO [10] is a head band based system which sends alarm if the swimmer spends a long period under water. These system require the installation of equipments in the area where they are used. Which can make from the systems costly for private swimming pool as well as not suitable for large swimming area such sea. These systems also consider the presence of lifeguard nearby to respond to the alarm.

In this research we aim to create an affordable drowning prevention system which can be flexibly used in various locations. For this we make use of the recent advancement in Smartphones equipped with pressure sensor and accelerometer which we attached at the swimmer's head level. We developed and application which can measure the motion and time spent by the swimmer underwater and compare it to a predefined maximum time-lapses threshold. So if an abnormal behavior is observed on device alarm is triggered and waist airbag is deflated to evacuate the victim.

II. Experimentation Setting

A. The Waist Airbag

We use a manual waist inflatable PFD belt. Which is composed of nylon bag, a compact 24 gram CO_2 gas bottle connected to a deflating system. Manually Pulling the cable connected to the deflation system cause the airbag to deflate. PFD belts are mainly dedicated for boating or fishing activity to save the user who fall incidentally in water. However users without swimming ability are not able to pull the cable manually once they fall in water as they enter in a panic situation and keep fighting for breathing.

So another variety of the system exist called automatic which include a water sensitive component. So when the victim fall in water the system deflate automatically.

In this research we use the manual edition of PFD system as the automatic one is not suitable to be used during swimming. We connect to the deflation cable to a high torque servomotor which is controlled by a microcontroller (IOIO board), as described in Fig. 6.

The servo motor deflate the system by pulling out the cable similarly to the human pulling action. The IOIO board can communicate with Smartphone using USB cable or wirelessly using Bluetooth communication. As we are using the system in water we need to ensure the protection of the electronic components in the servomotor and IOIO board. For this, we use epoxy potting and gum dipping to protect water sensitive components (Fig.7).

B. Processing unit

We are using Smatphone (Galaxy Nexus) which is equipped with built in pressure sensor and accelerometer.

We choose a Smartphone based platform due to several reasons.

- Smartphones are getting very popular
- It is possible to create an intuitive user interface
- A wide variety of water proof cases for Smartphone exist some makers start releasing waterproof models such Panasonic Eluga which is 1 meter waterproof .

The authors are with the Interfaculty Initiative in Information Studies, Graduate School of Interdisciplinary Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan ({mohamed, wakuda, koshizuka, ken}@sakamura-lab.org)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

- Smartphone system reusability can reduce the cost of the end product for the final customer.
- The possibility to use the network layer to communication an alert message.
- In addition to drowning prevention purpose. It might make sense to the user to carry out his Smartphone with him in water after protecting it with waterproof case to responds to phone calls or listen to sound tracks.

.

III. System Description

The user first installs a Smarthone application and set up the maximum time that he can spend underwater (Fig.8). He inserts then the device in a waterproof case and wear the waist airbag and turn it on . He inserts then the Smartphone in the swimming cap at the back level of his head as described in Fig.1.

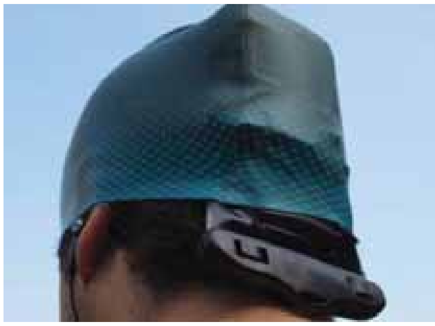


Figure1. Smartphone with waterproof case inserted in the swimming cap

To increase the reliability of the proposed system we use a second alarm scenario based on the swimmer motions. In the case the person is underwater and motionless for T2 period of time that might mean that the person is drowning so an alarm is triggered. To estimate the swimmer motion we calculate acceleration square root accSQR and we compare it with a threshold accSQRth.
The decision unit will trigger an alarm in the two following situation (Fig.2).

If (T>T1)
If ((accRoot<accTh) for duration (T>T2))
With T = time spend under water

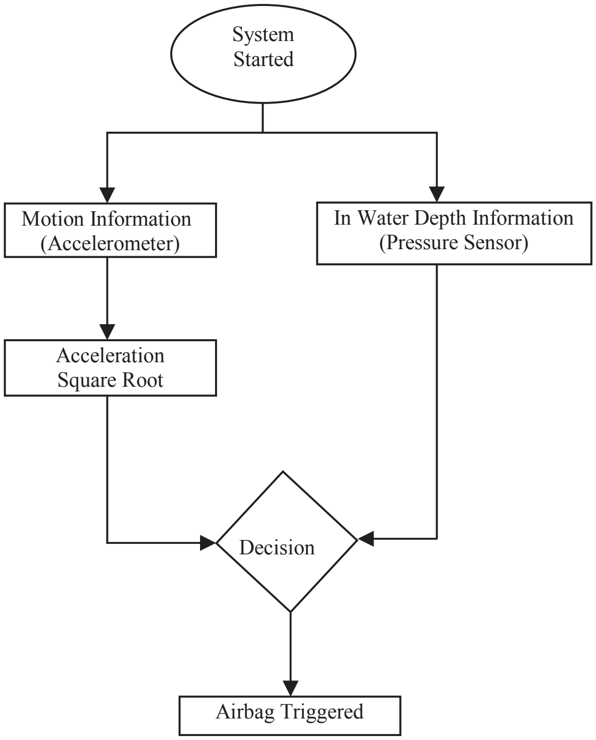

System Started
Motion Information (Accelerometer)
In Water Depth Information (Pressure Sensor)
Acceleration Square Root
Decision
Airbag Triggered

Figure2. System Block Diagram

A. *Accelerometer Data*

Accelerometer is a sensor that measures the proper acceleration of an object. In this research we use it to measure the motion of the swimmer. For this we calculate acceleration square root accSQR:

$$accSQR = \frac{x^2 + y^2 + z^2}{g^2} \tag{1}$$

We calculate then maximum accSQR (max_accSQR) during a predefined time window W.
We compare then accSQR with a predefined accSQR threshold (accSQRth) to check the swimmer motion.

B. *Pressure Data*

Pressure is the force per unit area applied in a direction perpendicular to the surface of an object. The measured pressure is sensitive to the environment where the sensor is located. We have conducted several experimentations in a water tank on the pressure sensor to identify the effect of waterproof case and swimming cap on the measured values (Fig.3).

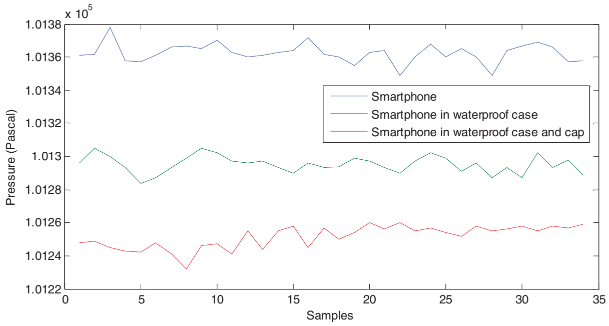


Figure 3. Pressure variation according to the pressure sensor location

The results show that the measured pressure decrease by about 65 Pascal when the Smartphone is enclosed in the waterproof case. The pressure decrease again by about 42 Pascal when the Smartphone is inserted in the swimming cap. The current air atmospheric pressure value is very important as it is used to identify the whether the head position of the swimmer is inside or outside the water. As the pressure fluctuation is relatively important from one environment to another, it is important to ensure that the system defines properly the correct atmospheric pressure reference value at the beginning.

IV. Results

We were able to the measure a significant pressure fluctuation between the outside and inside water regions (Fig. 4). After submerging the system in a water level of just 1 inch, the pressure increased by 133 Pascal compared to the average on air pressure avP (Table. I). At 5 and 10 inches depth the pressure level increased respectively by 1193 and 2683 Pascal. This information helps the system identifying the head level of the swimmer to trigger an alarm if he spends long time under water. We were also able to detect the motion and motionless situation from the accelerometer information (Table.II) and this after fixing the threshold accSQRth to 1.25 and time window W to 5 seconds. The system measures the maximum value of accSQR (max_accSQR) in each time window W and compares it with accSQRth. If max accSQR is higher than accSQRth is means the person is in motion. In the opposite case, when max_accSQR<accSQRth it means that the person is motionless for the window time duration W=5 second (Fig.5). Then the system checks the current pressure if it is higher than the on air pressure avP plus 133 Pascal (corresponding to the pressure underwater at one inch level), an on device alarm is triggered and the airbag is deflated

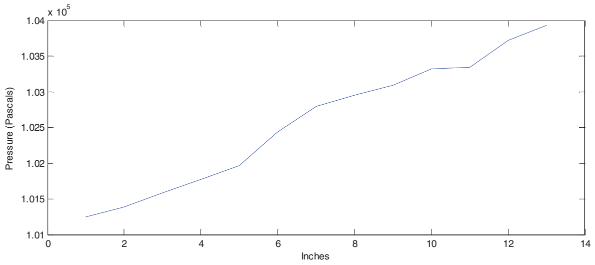


Figure 4. Pressure measurement in water at different depths

TABLE I
MEASURED PRESSURE FROM THE SYSTEM AT DIFFERENT DEPTH

Depth (Inches)	*Pressure (Pascal)*	*Pressure Differences In-Out Water*
0(avP)	101247	-
1	101380	133
2	101580	333
3	101770	523
4	101960	713
5	102440	1193
6	102790	1543
7	102950	1703
8	103090	1843
9	103320	2073
10	103345	2098
11	103720	2473
12	103930	2683

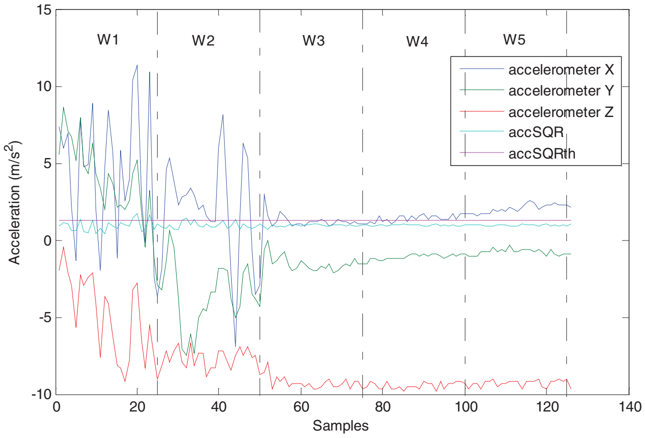


Figure 5. Accelerometer measured information

TABLE II
MOTION CONDITION IN WATER IDENTIFIED FROM ACCELEROMETER PROCESSED INFORMATIONS

Time Window	*W1*	*W2*	*W3*	*W4*	*W5*
Max accSQR	1.699	1.385	1.026	1.041	1.036
Situation	Motion	Motion	Motionless	Motionless	Motionless

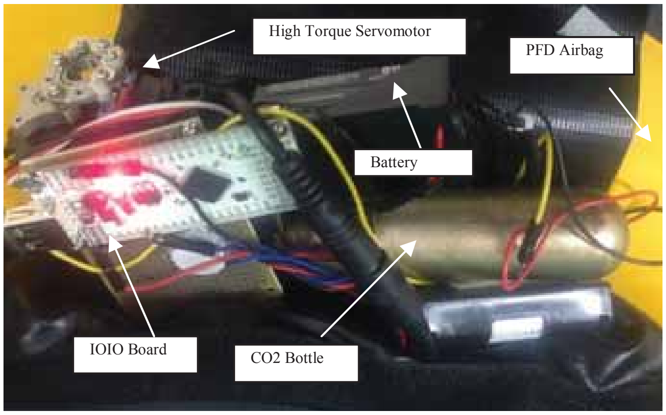


Figure.6 Picture of the system before applying epoxy potting and gum dipping

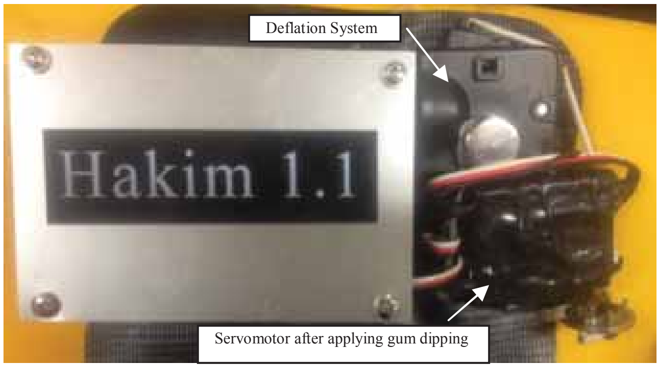
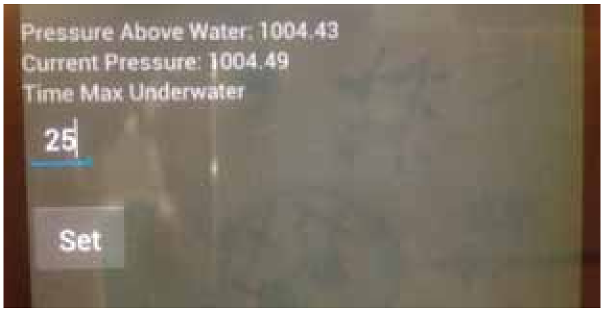


Figure 7. System after waterproofing



Pressure Above Water: 1004.43
Current Pressure: 1004.49
Time Max Underwater
25

Figure 8. Screen shot of the developed application

V. CONCLUSION AND FUTURE WORK

In this research we show how we can use Smartphone as a platform for drowning alarm system, using the embedded device accelerometer and pressure sensors. We show also the efficiency of pressure sensor in detecting swimmer's head position: whether it is outside or inside water. When the swimmer submerge his head in water, we were able to measure a sensitive pressure changes at just one inch depth. Information from accelerometer was used to identify the case when the swimmer is motionless. So if an abnormal behavior is measured in either of these two cases an airbag is deflated. In the future we consider conducting further experimentation in order to define the appropriate parameters of the system which are necessary to reduce false positive alarms. We are also currently working on developing further more this system to detect the victim at early drowning stage by analyzing his physiological body changes [4,5,7,8].

ACKNOWLEDGMENT

I would like to thank Mr. Mahdi Kharrat for his support in conducting the experimentation and Dr. M. Fahim Ferdous Khan for helping proofreading this paper.

REFERENCES

[1] World Health Organization, "Drowning Fact Sheet N°347", November 2010.

[2] Centers for Disease Control and Prevention USA, "Unintentional drowning factsheet", May 2011.

[3] F. Pia, 'Observations on the Drowning of Nonswimmers', Journal of Physical Education, July, 1974.

[4] M. Kharrat, Y. Wakuda, S. Kobayashi, N. Koshizuka, K. Sakamura, "Near Drowning Detection System Based on Swimmer's Physiological Information Analysis", World Conference on Drowning Prevention (WCDP), May 2011

[5] M. Kharrat, Y. Wakuda, S. Kobayashi, N. Koshizuka, K. Sakamura, "Adaptive Radial Artery Pulse Rate Measurement using Piezo Film Sensor Based on Ensemble Empirical Mode Decomposition," the 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB 2010),

[6] C. Edmonds, "Drowning Syndromes-The Mechanism", Forty-seventh Workshop of the Undersea and Hyperbaric Medical Society, 1997

[7] Kharrat, M.; Wakuda, Y.; Koshizuka, N.; Sakamura, K.; , "Near Drowning Pattern Detection Using Neural Network and Pressure Information Measured At Swimmer's Head Level' " 7th International ACM Conference on Underwater Network and Systems WUWNet'12, November 2012, to be appeared

[8] Kharrat, M.; Wakuda, Y.; Koshizuka, N.; Sakamura, K.; , " Near Drowning Pattern Recognition Using Neural Network and Wearable Pressure and Inertial Sensors Attached at Swimmer's Chest Level" 19th International Conference on Mechatronics and Machine Vision in Practice, November 2012, to be appeared

[9] SenTAG , www.sentag.com , Nov 2010.

[10] WAHOOO , www.wahooosms.com, Nov 2010.

Leveraging Smart Grid Technology for Home Health Care

Thomas THOMAS, *Member, IEEE,* Cade CASHEN, *Student Member, IEEE*, and Samuel RUSS, *Member, IEEE*

***Abstract*—Smart grid technology is emerging as a powerful method for managing home energy consumption and improving the efficiency of power delivery. Due to declining birth rates and advances in health care, the world population is aging and new technology can assist in care of elderly populations. Smart grid technology can be used to provide useful information on the activities of daily living and can be used to monitor both the short-term and long-term health of elderly individuals. This report outlines some experiments that demonstrate the concept.**

I. INTRODUCTION

After a review of current technology, this paper proposes network architecture for home-health monitoring and describes the results of some preliminary experiments in implementing the network. By "burying" the sensing and monitoring functions of the network into a pervasive home network, it is hoped that elderly patients will not find the new technology disturbing. By leveraging the information from the network, a detailed portrait of a patient's activities of daily living can be drawn unobtrusively.

II. REVIEW OF TECHNOLOGY

A. *Smart Grid Technology*

Smart grid technology is a very broad, very active area of research. The research outlined here is focused on measuring the power consumption of individual appliances. Previous work in this area include a patent for a network of wall plates to track RFID tags [1], a patent for a wall plate that measures over-current or electrical-fault conditions and that uses an RFID tag to determine the current limit of attached appliances [2], and commercially standard for measuring and reporting power consumption [3].

The systems described in [2] and [3] are designed primarily for electrical-fault isolation and energy management. The focus in this work is, instead, on the correlation of energy measurements to the activities of daily living.

B. *Monitoring Activities of Daily Living*

In order to monitor elderly patients, one important capability is to sense the activities of daily living (ADL). Many approaches have been proposed including wearable sensors [7],[8], machine vision [9], and infrared sensors mounted in the ceiling [10]. Because of concerns about technology acceptance and about "being spied on" by the elderly, the focus in this work is on completely passive sensing that does not involve imaging.

C. *Monitoring Power Consumption for Home Health Care*

As noted above, the ability to measure power consumption is an important smart-grid capability. The observation can be made that measuring the power consumption of individual appliances throughout the course of each day can provide a detailed portrait of the activities of daily living of an elderly patient living at home. For example, one might make coffee or use a toaster every morning and watch television every afternoon.

This information can then be aggregated and mined for useful patterns of information. Short-term anomalies might indicate an emergency condition. Long-term shifts might indicate decline in function.

The ability to monitor the power consumption of individual appliances is therefore useful for home health care.

Our proposed solution is a "smart wall plate" (along the lines of [3]) that can be inserted on top of existing wall outlets to add wireless connectivity, power-monitoring, and the ability to identify individual appliances. The use of this type of information for home health care appears to be novel.

D. *Other Sensor Capabilities*

The wireless network that connects the smart wall plates can be used for other functions, most specifically motion detection. See [4] for a more detailed discussion of this capability.

Other sensor nodes can be added to the wireless network, such as sensors to monitor pharmaceutical usage and to monitor water consumption. The former is of clear importance and is a very difficult issue in home health care. The latter is needed to round out a complete portrait of daily activities. For example, if no water is used (e.g. no toilet, bath, or kitchen sink) then it may flag an emergency medical condition.

There are commercially sold systems that monitor pharmaceutical usage. For example, one pharmaceutical company has begun placing RFID tags in medicine bottles [5] and there are commercially sold bottle caps that indicate when medicine needs to be taken [6]. The approach taken here is to connect a pharmaceutical sensor to the in-home network.

III. HOME SENSOR ARCHITECTURE

A. *Overall Architecture*

The overall architecture of the smart-grid network for home health monitoring is shown below in Figure 1.

The authors are affiliated with the University of South Alabama, Mobile, Alabama, USA.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

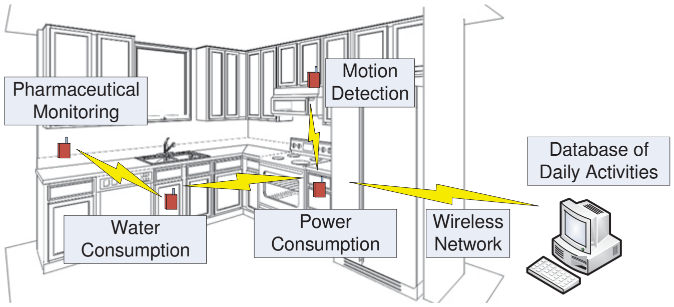


Fig. 1. Smart-Grid Home-Health network. This shows how data from a smart-grid network can be combined with data from other sensors to form a complete picture of the activities of daily living.

The network includes sensors to measure power consumption, pharmaceutical usage, and water consumption.

B. *Power-Consumption Node*

The power-consumption nodes form the "backbone" of the network. With the addition of an input device, like an RFID reader, the nodes can also determine which appliances are connected. (In such a case, RFID tags could be added to appliance cords to permit identification, as in [3].) Thus the nodes perform an important "smart grid" function of measuring the power consumption in the home down to the level of individual appliances. Additionally, the wireless traffic itself can also be used to perform important tasks such as motion detection [4].

C. *Other Nodes*

Measurement of the activities of daily living requires more than just power consumption. Two other important aspects are the patient's correct use of prescribed medications and the patient's consumption of water (e.g. bathroom and kitchen). Because the smart wall plates create an extensible wireless network, these types of sensors can easily be added.

IV. Experimental Node Design

A simple experimental node was constructed, as shown in Figure 2.

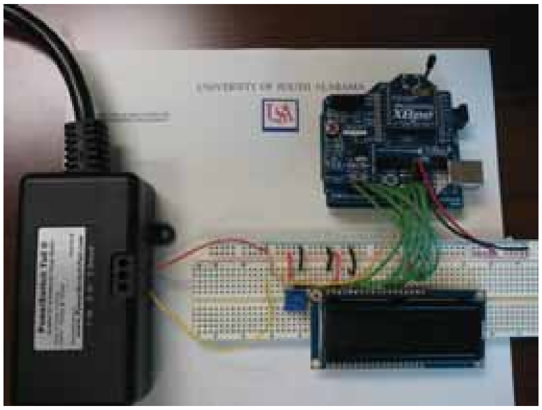


Fig. 2. Experimental node design. This prototype was built to implement the wireless-network, power-monitoring, and power-shutoff features as a proof of concept.

It included a microprocessor board and wireless add-on board, a transformer to convert AC current into an analog voltage, and a relay to provide power-disconnect capability. The experimental node was programmed to measure power consumption and send back current measurements, which it did successfully.

As described in [4], the node also made continuous received signal-strength indication (RSSI) measurements to perform motion detection.

In the experiment, motion-detection measurements were used to control the relay. In other use cases, the relay could be programmed, e.g., to trip during over-current conditions or under user command via the wireless network. The latter would add a remote-control capability to appliances in the home.

V. Results, Conclusions, and Future Work

The experimental node was successful, and now needs to be redesigned into a smaller form factor, and an RFID reader needs to be added. Accumulation of measurements over time needs to be carried out, and the correlation of the measurements to the activities of daily living can be demonstrated. Nodes for pharmaceutical monitoring and water-flow monitoring also need to be developed.

Acknowledgment

The authors acknowledge the support of the Department of Electrical and Computer Engineering and its senior project program.

References

[1] U.S. Patent 7,257,108, "Determining the physical location of resources on and proximate to a network", Issued August 14, 2007.

[2] U.S. Patent 8,000,074, " Electrical Power Distribution System", Issued August 16, 2011.

[3] N. Jones, "RightPlug Digital Plug Encoding: Synergies with the Smart Grid Initiative", Oct. 8, 2009, Available via http://www.rightplug.org/whitepapers/RightPlug%20&%20Smart%20Grid%20R001.pdf . [Accessed Sept. 25, 2012]

[4] C. Cashen, S. Russ, and T. Thomas, "Using a Wireless LAN to Perform Motion Detection", Submitted to International Conference on Consumer Electronics (ICCE).

[5] S. Gotensparre, "Rexam unveils RFID-tagged pill bottles," in-pharmatechnologist.com, Jan. 25, 2007, Available via http://www.in-pharmatechnologist.com/Drug-Delivery/Rexam-unveils-RFID-tagged-pill-bottles

[6] S. Race, " Now, Even Granny's Fuzzy Slippers Are Texting You," *Wall Street Journal*, July 15, 2011.

[7] Logan, B., and Healey, J., "Sensors to Detect the Activities of Daily Living", *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5362 - 5365.

[8] Bajcsy, R., "Distributed Wireless Sensors on the Human Body", *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering*, p. 1448.

[9] Sangho Park and Kautz, H., "Hierarchical recognition of activities of daily living using multi-scale, multi-perspective vision and RFID" , *2008 IET 4th International Conference on Intelligent Environments*, pp. 1 - 4.

[10] Shuai Tao, Kudo, M., Nonaka, H., and Toyama, J., "Recording the Activities of Daily Living based on person localization using an infrared ceiling sensor network", *2011 IEEE International Conference on Granular Computing (GrC)*, pp. 647 - 652 .

Active Monitoring for Lifestyle Disease Patient Using Data Mining of Home Sensors

Young-Sung Son[1], Topi Pulkkinen[2], *Member, IEEE* and Jun-Hee Park[1]
[1]ETRI, [2] VTT Technical Research Centre of Finland

***Abstract*-- This paper describes user activity recognition for lifestyle disease patients at home: ways to define data mining system for sensing, logging, analyzing, mining, measuring and recognizing user's daily activities. Lifestyle disease patients spend most of the time at home. There are lots of sensing data that can be based on home devices with home networking (sensors, gadgets, appliances, cameras, smart phones and some software applications running on computers). Main problem is interoperability, there is no standard framework for logging, analyzing and utilizing the available data sources.**

In this paper, we will introduce our layered architecture to do data mining for user's activity recognition. Understand user's life pattern can help medical services to cure and prevent diseases from developing.

I. INTRODUCTION

Recently, there exist a lot of sensors, home network devices, ehealth devices and computing devices at home environment. Most devices can create some logs such as status, sensing value, network traffic, user control and so on. This kind of sensing data can be used to recognize user's activity and to build user's lifestyle pattern at home [1, 2].

We are struggling with social problems for example emerging quantities of lifestyle disease patients such as diabetes, heart disease and hypertension. These lifestyle patients usually spend many hours a day at home, and only rarely visit a hospital. Regular treatment is usually lifestyle adjustments such as balanced meal, sufficient sleep and periodic exercise in daily life [3].

This paper will introduce the layered architecture for active monitoring of a lifestyle disease patient using data mining of home sensors

II. USER SCENARIOS

Mr. Park is a business man with unbalanced diabetes type 2. His home is a condo for two where he lives alone. He has installed some sensors in the kitchen and the bed room, a camera in the living room, and an activity detection app on his smart phone. He is afraid of diabetes complications and wants to get helpful messages from the system.

A. Scenario 1 - Data mining for life pattern

While Mr. Park is at home or carrying the mobile phone, the system can save all the measurement signals to the database. The measurements include door opening/closing pattern for refrigerator and microwave, noise, humidity, CO2, luminance, medical measurements (cholesterol, blood glucose and a blood pressure), appliances and basic questions from the mobile phone app such as "Did you sleep well?"

B. Scenario 2 - Event based decision

A camera, a mobile phone app or a noise sensor (or the combination of these) detects Mr. Park falling down and not getting up. The system interprets this information as an accident. An emergency unit receives the information about Park's disease and estimates the possible cause of accident (in this case the cause it is hypoglycemia – too low blood sugar, which leads to insulin shock). This was caused by Park injecting too much insulin after the meal accidently.

III. STRUCTURE AND COMPONENTS OF OUR SYSTEM

We defined our system as User Active Recognition Framework (UARF). The vertical communication between the layers will be handled by an inter-process communication interface. The network layer, device layer and application layer can provide abstract data that can be used in upper data analysis layer, which composes different services upon users' requests. The data forms a basis for enabling intelligent context aware application development, execution, composition and provisioning, which can be implemented by UARF. Features of the framework will support user-level context data creation and combination as well as context reasoning and context data mining.

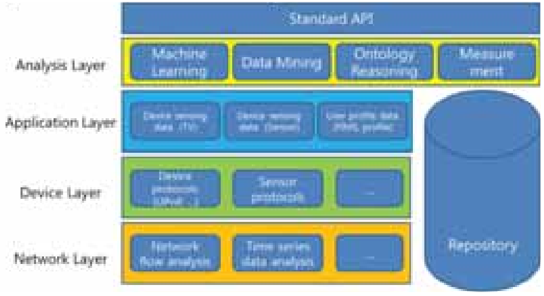


Fig. 1. Layered architecture of User Activity Recognition Framework.

Our lifestyle disease active monitoring system on UARF can connect sensor devices, medical devices, intelligent appliances, cameras and a smart phone together. In home, the system should have enough sensors for measuring user's activity in different situations. Medical sensor for measuring sugar blood rate is one of the most important devices in this system. The other sensors can detect the patient's activities indirectly. These sensing data would be a fundamental database for understanding user's life pattern. Outside of home, the system can gather data with a smart phone and web services. System can use a questionnaire to monitor the patient's current

This work is supported by the IT R&D program MKE/KIAT. [2010-TD-300404-001, Home Information Remote Aggregation and Context Inference Prediction Technology Development] and also VTT Technical Research Centre of Finland.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

situation. The conceptual system configuration is illustrated in Fig. 2. All gathered data should be stored at the Log Storage. The system can then analyze all the data within the Log Mining Server.

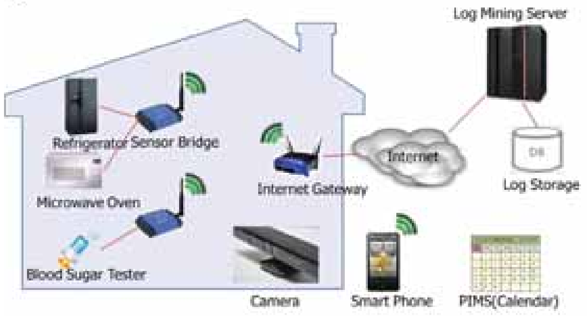


Fig. 2. Lifestyle disease active monitoring system configuration.

For monitoring the home environment, we developed a sensor board with six types of sensors: temperature, luminance, humidity, noise, co2, and magnetic sensors. For general purpose, this board is operated with embedded Linux and D-MAP (Device Management Architecture and Protocols) that is one of the major home network middleware proposals. The board presented in Fig. 3 has three packages: (1) an active monitoring package, which initiates periodic sensing probes to measure properties of sensors; (2) an event driven package, which records the camera's vision recognition and user's feedback & correction message procedure through a smart phone; (3) a management package, which facilitates remote management, such as software upgrades, as well as pushing sensing logs to a central server.

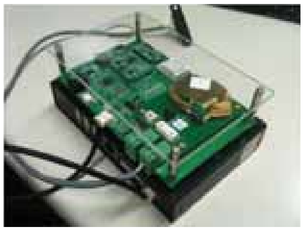


Fig. 3. Home sensor board.

IV. APPLICATIONS OF UARF

Fig.4 describes a life log auto-construction method to analyze sensor data and recognize patient's behaviors. In this graph, only the noise sensor data is handled and some time periods that show high level of noise signal. When we find user's patterns with supervised learning, these peak signals can match some user activities. If we combine several sensor data, we can get increase sensing correctness. In the learning phase, the system asks the patient to characterize an unknown event with a smart phone application. Through these feedbacks, the system can construct user life log automatically.

Medical doctors recommend that diabetes patient should keep regulated lifestyle that includes sufficient sleep, balanced meals and periodical exercise. We have implemented patient lifestyle monitoring service. Service checks their auto-constructed life log that contains several activities required by lifestyle regulation. The service calculates several items and shows up total score for every day. For example, a meal item has 15 points, sleep item has 10 points and exercise item has 10 points – total 35 points. Patients can know whether he/she had maintained balanced lifestyle that helps to prevent further complications of the disease.

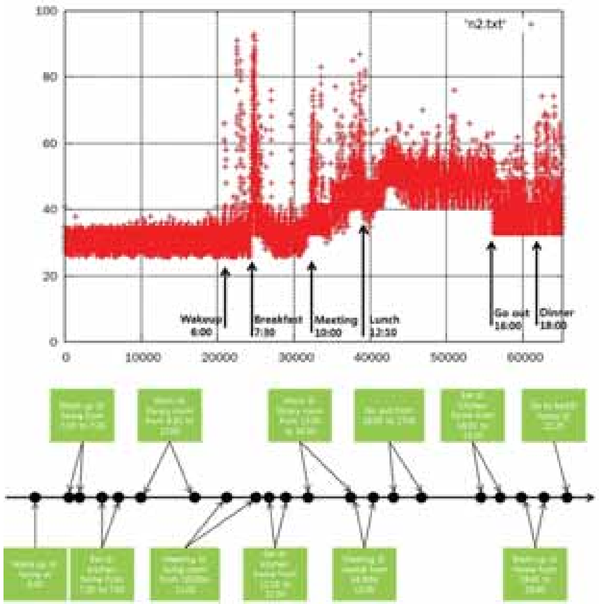


Fig. 4. Life log auto construction using noise signals.

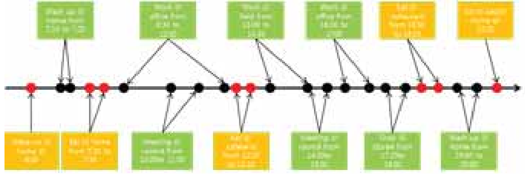


Fig. 5. Patient lifestyle monitoring and analyzing service.

V. CONCLUSION

This paper describes the active monitoring of a lifestyle disease patient using data mining of home sensors. With mining server, we can develop some applications such as life log auto-construction and balanced lifestyle service by utilizing data analyzing and medical knowledge. There are many challenges for supporting chronic disease patients at home. UARF can monitor user activities and make a plan to recommend patient to maintain or adjust their lifestyle to prevent disease status deterioration.

REFERENCES

[1] Y. Son, T. Pulkkinen, K. Moon, and C. Kim, "Home energy management system based on Power Line Communication, IEEE Transaction of Consumer Electronics., vol. 56, no. 3, pp. 1380-1386, Aug. 2010.

[2] Akyildiz I., Su W., Sankarasubramaniam Y., Cayirci E., "A Survey on Sensor Networks". IEEE Communications Magazine, Aug. 2002. pp. 102-114.

[3] Bisiani R., Merico D., Mileo A., Pinardi S. "A Logical Approach to Home Healthcare with Intelligent Sensor-Network Support". The Computer Journal Advance Access published May 14, 2009. 20p.

Bitstream Parsing Processor with Emulation Prevention Bytes Removal for H.264/AVC Decoder

Hyun-Ho Jo[1], Jung-Han Seo[1], Dong-Gyu Sim[1], *Member, IEEE*
Doo-Hyun Kim[2], Joon-Ho Song[2], Do-Hyung Kim[2], and Shihwa Lee[2]
Image Processing Systems Laboratory, Kwangwoon University, South Korea
Samsung Advanced Institute of Technology, South Korea

Abstract-- **In this paper, we present a bitstream parsing processor including emulation prevention bytes (EPB) removal for H.264/AVC decoder. The proposed bitstream parsing processor includes several specific instructions for bitstream parsing. Furthermore, it employs double bitstream buffers to remove EPBs for sequential bitstream parsing. Experimental results show that the proposed bitstream parsing processor achieves a cycle reduction of 18%, compared with conventional EPB removing methods. In addition, the proposed method reduces the buffer size by a large amount to preserve a network abstraction layer (NAL) unit for the removal of EPBs.**

I. INTRODUCTION

Since various video codecs have been widely utilized in the market, the demands of multi-format video decoders are dramatically increasing for mobile devices. Conventional full hardware development to support a wide range of video standards is pretty ineffective in several viewpoints of development cost, time, and its scalability. Under consideration of these issues, software-based development with reconfigurable processors (RP) and application-specific instruction-set processors (ASIP) is regarded as a competitive approach [1]. In addition, in consideration of low-power consumption in mobile devices, multi-core platforms with relatively lower-clock cores become the popular approach for high-complexity applications such as video processing. By utilizing data-level parallelism for video decoding, inverse quantization, inverse transform, intra prediction, inter prediction, and deblocking filter module can be parallelized with macroblock-level data partitioning. However, the entropy decoding cannot be decoded in parallel because of a bit-by-bit dependency. As the performance of parallel video decoders is highly affected by the decoding speed of the entropy decoder, the speed of entropy decoding is very important for the high-speed video decoder on multi-core systems.

Here, we propose a bitstream parsing processor that includes several specific instructions for multi-format bitstream parsing. In addition, the proposed bitstream parsing processor supports removal of emulation prevention bytes (EPBs) for H.264/AVC with a proposed double bitstream buffer structure. This paper is organized as follows. Section II presents parsing process of H.264/AVC with the EPB removing process. In Section III, we present the proposed bitstream parsing processor and its double bitstream buffer structure for removing EPBs. The experimental results and conclusion are followed in Section IV and Section V, respectively.

II. BITSTREAM PARSING AND EPBS REMOVAL

Bitstream decoding is a mapping process where a codeword is converted into a number or symbol. For bitstream decoding, bitstream parsing should be conducted in advance. Due to the large size of bitstreams, they are usually stored in an external memory. Furthermore, we cannot know a proper codeword length because its length is variable. Thus, the external memory should be frequently accessed for the bitstream parsing process. To decrease the numbers of external memory access, the operations for bitstream parsing listed in Table I were proposed with two 32-bit bitstream registers in hardware-based development [2]. All operations with conventional bitstream parsing processors employing dedicated parsing logics can be conducted in one or two cycles when a part of the bitstream is stored in the internal bitstream registers in advance.

TABLE I
BITSTREAM PARSING OPERATIONS

Operation	Description
Getbits(*n*)	The Getbits operation reads and skips *n* bits from bitstream buffer
Showbits(*n*)	The Showbits operation reads *n* bits from bitstream buffer
Skipbits(*n*)	The Skipbits operation skips *n* bits from the bitstream buffer

By utilizing the bitstream parsing operations and simple internal bitstream register structure, we can effectively parse multi-format bitstreams except for the EPBs removal in H.264/AVC. For the conventional removal of EPBs, we need to find two consecutive start code prefixes and extract the network abstraction layer (NAL) unit in between them [3]. In this step, all the consecutive bits from the first to the next start code prefixes should be sequentially loaded into an internal bitstream register. As the size of NAL units can be quite large, up to 3Mbytes (Level 5.1), we should store the extracted NAL unit in external memory. Then, we need to find EPBs and remove them from the extracted NAL unit by sequentially loading all the data into the internal bitstream register buffer. By removing the EPBs, we can obtain the raw byte sequence payload (RBSP) but it should be stored in the external

This research was partly supported by the Samsung Electronics, and partly supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2012-H0301-12-1011) supervised by the NIPA (National IT Industry Promotion Agency).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

memory again. This conventional method requires a huge number of bitstream loading processes from an external memory to internal registers, resulting in the removal delay of the EPB. Furthermore, it requires an additional buffer to save the extracted NAL unit.

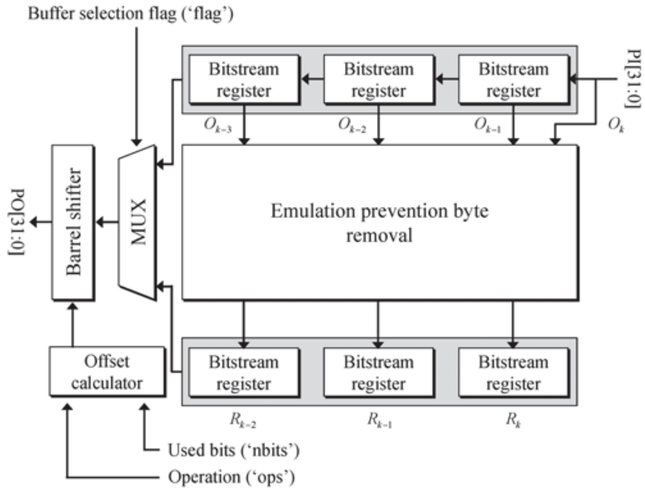


Fig. 1. Block diagram of proposed 'I_STREAMS' instruction

III. THE PROPOSED BITSTREAM PARSING PROCESSOR

To remove EPBs of H.264/AVC without one slice delay and a large buffer space to preserve a NAL unit, we propose a double bitstream buffer structure for bitstream parsing speed-up, as shown in Fig. 1. Conventional bitstream parsing methods only employ two 32-bit internal registers to load bitstreams from an external memory. In contrast, we use six 32-bit internal registers to parse bitstreams with EPBs removal for the H.264/AVC bitstream. The bitstream parsing operations in Table I were implemented with one consolidated instruction, 'I_STREAMS'. By using 'I_STREAMS' instruction in the parsing process, EPBs in a NAL unit are automatically removed. The six 32-bit internal registers are categorized into two sets. The register set without the EPBs (R_k, R_{k-1}, and R_{k-2} registers) contains part of a bitstream after removing the EPBs. The register set with the EPBs (O_{k-1}, O_{k-2}, and O_{k-3} registers) contains the original bitstream, which may include EPBs. When all the bits in O_{k-3} are consumed, the register set with EPBs is updated from its adjacent register. At the same time, the register set without EPBs is updated from the register set with EPBs (O_{k-2}, O_{k-1} and O_k) after removing the EPBs. For this, the EPB removal detects the EPBs from the first byte in O_{k-2} to last byte in O_k. Note that O_{k-3}[15:0] should be used to check whether the first byte in O_{k-2}, O_{k-2}[31:24], is an EPB or not. One parameter of 'I_STREMAS', 'flag', indicates which register set is used. For example, the register set with the EPBs is used to detect the 'end of slice'. Otherwise, three operations in Table I are executed on the register set without the EPBs.

IV. EXPERIMENTAL RESULTS

The performance of bitstream parsing and decoding can vary depending on the characteristics of bitstreams and their bitrates. In this paper, various standard videos inputs such as CIF, HD, and Full-HD were used. To generate the bitstreams, we employed the JM 18.2 with the quantization parameter (QP) 27 for I frames and 28 for P frames.

To evaluate the performance of the proposed EPB removal, we counted the total number of cycles for bitstream decoding in regards to the test bitstreams without EPB removal in 'I_STREAMS'. EPB removal was performed with the generic C code implemented in the reference software of JM 18.2. As only the EPB removal is disabled, 'Getbits', 'Skipbits', and 'Showbits' can be executed in two cycles with the 'I_STREAMS' instruction. Table II presents the total number of mega cycles per second (MCPS) during entropy decoding of each test case. We found that the proposed EPB removal achieves a cycle reduction of 18%. Furthermore, the proposed method does not need to have a memory to keep the NAL unit and one slice delay. As a result, we can save around 1.5 and 3.0 Mbytes memory for level 4.0 and level 5.1, respectively.

TABLE II
COMPARISON OF ENTROPY DECODING CYCLES ACCORDING TO THE EPB REMOVING METHOD

Resolution	Bitstream	Decoding cycle (MCPS)		Reduction Ratio (%)
		External EPB removal	EPB removal in 'I_STREAMS'	
CIF	Foreman	5.5	4.5	18
	Mobile	16.6	12.9	22
	Paris	7.2	5.7	21
	Tempte	12.7	10.0	21
HD	Bigship	28.4	23.6	17
	City_corr	35.2	28.8	18
	Crew	30.7	25.6	17
	Jets	14.7	12.9	12
Full-HD	Rush_hour	63.5	51.8	18
	Station2	42.4	36.0	15
	Sunflower	48.7	40.7	16
Average	-	-	-	18

V. CONCLUSION

In this paper, we proposed a bitstream parsing processor with EPB removal for H.264/AVC decoder. By consolidating an EPB removal operation into the bitstream parsing instructions, the proposed bitstream parsing processor can remove EPBs during the bitstream parsing process. Furthermore, we remove one slice delay and a large amount of memory in the process of EPBs removal. Experimental results demonstrate that the bitstream parsing processor achieves a cycle reduction of 18%, compared with the conventional EPB removal method.

REFERENCES

[1] J.H. Song, W.C. Lee, D.H. Kim, D.-H. Kim, and S.H. Lee, "Low-power video decoding system using a reconfigurable processor," *Proc. IEEE Int. Conf. Consumer Electronics*, pp. 532-533, January 2012.

[2] M. Berekovic, H.-J. Stolberg, M.B. Kulaczewski, P. Pirsch, H. Moller, H. Runge, J. Kneip, and B. Stbernack, "Instruction set extensions for MPEG-4 video," *Journal of VLSI Signal Processing Syst.*, vol. 23, no. 1, pp. 27-49, Oct. 1999.

[3] JM 18.2 software, http://iphome.hhi.de/suehring/tml/

Positive and Negative Max Pooling for Image Classification

Bin WANG, Yu Liu,WenHua XIAO, Zhihui XIONG, Maojun ZHANG

College of Information System and Management,
National University of Defense Technology, Changsha, China

Abstract—Max pooling has been regard as the best pooling method in image classification when image features are coded by sparse coding [2]. However, max pooling reduces the classification discrimination, since it doesn't distinguish the sign of coding coefficient but only selects the max absolute value. In order to increase the image representation discrimination, we preserve the sign of code coefficient and develop a feature pooling method named PN-Max pooling. Experimental results show that PN-Max pooling achieves higher image classification accuracy than Max pooling.

I. INTRODUCTION

Commonly, there are four steps in the popular Bag of Words based image classification: feature extraction, coding, pooling and classification. The feature pooling refers to the process that combines the responses of feature detectors (eg.SIFT) into some statistic which summarizes the joint distribution of the features. This idea of feature pooling originates in Hubel and Wiesel's seminal work on complex cells in the visual cortex [3], and is related to Koenderink's concept of locally orderless images [4]. These researches showed that the global information of a patch can be approximately represented as the max signal of the descriptors detected in this patch. Inspiring from this idea, Yang et al.[1] proposed Max pooling method which selects the code coefficient that has the max absolute value to pool features for image classification under ScSPM (Sparse coding Spatial Pyramid Matching) framework. Their experiments showed that Max pooling outperforms other alternative pooling methods (such as Avg pooling and Sqrt pooling). Max pooling method becomes a popular feature pooling method in visual recognition field such as image classification, scenes recognition and action recognition [5]. However, the hidden assumption of pooling method in complex cells [3] is that all the value of coding coefficient is non-negative, which is based on analyzing the biology signals (eg: the signals of complex cells in the visual cortex) [6].

While coding features with sparse coding [2,7], Max pooling is not suitable, since there is not only positive but also negative value in the code coefficient. Additionally, recent research showed that sparse coding is a new union sub-space model [8]; in the sub-space, the codes which have same absolute value but different sign lie in different locations; the different locations have different meaning. Therefore, Max pooling method neglects the sign of codes, it also loses their discriminative.

In this paper, we propose a simple but effective extension of Max pooling called positive and negative max pooling (PN-Max). It preserves the sign information which aims to increase the classification discrimination. Experimental results show that PN-Max pooling achieves higher image classification accuracy than Max pooling.

II. FRAMEWORK

Our image classification framework is based on ScSPM[1], which includes four steps: feature extraction, sparse coding, feature pooling and linear classification. In our framework (showed in Fig.1), we replace Max pooling by PN-Max pooling.

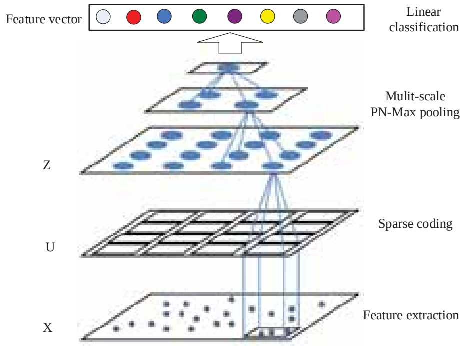


Fig.1. The illustration architecture of ScSPM with PN-Max pooling method.

III. FROM MAX POOLING TO PN-MAX POOLING

Let X be a set of D-dimensional local descriptors (e.g. SIFT) extracted from an image, i.e. $X = [x_1, x_2, ..., x_N] \in \mathrm{R}^{D\times N}$. Given a codebook with M entries $B = [b_1, b_2, ..., b_M] \in \mathrm{R}^{D\times M}$, sparse coding converts the descriptors X into a $M \times N$-dimensional code matrix C.

$$\arg\min_{C} \sum_{i=1}^{N} \left\| x_i - Bc_i \right\|^2 + \lambda \left\| c_i \right\|_{\ell^1} \qquad (1)$$

Where $C = [c_1, c_2, ..., c_N] \in \mathrm{R}^{M\times N}$ is the set of codes for X. The sparsity regularization term ℓ^1 plays very important roles: First, the codebook B is usually over-complete, i.e., M > D, and hence ℓ^1 regularization is necessary to ensure that the under-determined system has a unique solution; Second, the sparsity prior allows the learned representation to capture salient patterns of local descriptors; Third, the sparse coding can achieve much less quantization error than vector quantization such as K-means.

This work was supported by National Natural Science Foundation (NSFC) of China under project No.61175006, No.61175015, No. 60803101 and No.60872150; iCORE (Alberta Innovates) Alberta, and NSERC Canada.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Feature pooling methods convert the code matrix C into a feature vector $Z \in R^M$ to generate the final image representation. In ScSPM [1], they defined the pooling function F as a max pooling function on the absolute sparse codes:

$$Z_j = \max\{|C_{j1}|, |C_{ji}|..., |C_{jn}|\} \qquad (2)$$

where Z_j is the j-th element of z, C_{ji} is the matrix element at j-th row and i-th column of C, and n is the number of local descriptors in the relevant image region. Their image classification experiments showed that Max pooling outperforms other alternative pooling methods (such as Avg pooling and Sqrt pooling) [1].

Recent research has shown that sparse coding is a new union sub-space model [8]. In sub-space, the same absolute value codes with different sign lie in different locations. For example, in Fig.2, we select two bases b1 and b2 from the codebook B to construct a sub-space (b1 and b2 are not orthogonal, here, we draw them like orthogonal to look expediently). In the sub-space, there are two codes A and B which have the same absolute value but different sign. In Fig.2, the code A with positive sign lies in 1st quadrant while the code B with negative sign lies in 3rd quadrant. Therefore, code A and code B should be regard as different feature codes in the feature pooling step. Nevertheless, they cannot be distinguished when using Max pooling method, since both A and B have the same absolute value.

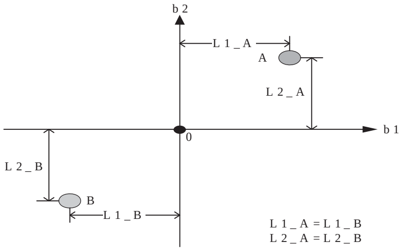


Fig.2. The illustration of that the codes which have same absolute value but different sign lie in different locations in sub-space.

Since the different locations have different meaning in sub-space, taking account of the sign of sparse coding coefficients can improve the distinguish ability for different signal coded by sparse coding. We preserve the sign information to increase discrimination of feature pool for image classification. Max pooling is improved to positive and negative max pooling (PN-Max) as Eq. (3).

$$Z_{j,pn-\max} = (Z_{j,\max+}, Z_{j,\max-}) \qquad (3)$$

$$Z_{j,\max+} = \max\{C_j \geq 0\} \qquad (4)$$

$$Z_{j,\max-} = \max\{C_j < 0\} \qquad (5)$$

In order to retain the same dimension as Max pooling, we also propose SPN-Max pooling method (sum of positive and negative max pooling) to avoid heavy cost of the classification time. The SPN-Max summates the two parts of PN-Max pooling result and can be written as Eq. (6):

$$Z_{j,spn-\max} = sum(Z_{j,\max+}, Z_{j,\max-}) \qquad (6)$$

IV. Experiment and Results

In order to verify the efficiency of our proposed PN-Max pooling method, we conduct image classification experiment on benchmark dataset Caltech-101 and 15 Scenes based on the code provided by Yang [8]. In our experiment, we used three pooling methods: Max pooling, PN-Max pooling and SPN-Max pooling. In each experiment, we randomly selected parts of the image from each class as training data, the remaining images were used to test. Then, we repeated these experiment 10 rounds to conduct the result. The detailed comparison result shows in table 1. All the result reported in this paper is obtained by our experiment. We note that it cannot reproduce the result reported in [1] when using max pooling, that proximately due to the parameter setting. However, in our experiment, the parameter is on the baseline. Experimental results show that the performance of this pooling method improved about 3% than Max pooling which is commonly regard as the best pooling method.

Table 1: Classification accuracy for different pooling operators on Caltech-101 and 15-Scenes

Dataset	Train times	Codebook size	512	1024	2048
Caltech-101	15	Max[1] PN-Max SPN-Max	63.65 ± 0.65 ***65.84 ± 0.49*** ***66.21 ± 0.45***	65.30 ± 0.62 ***67.04 ± 0.4***0 ***66.55 ± 0.77***	66.58 ± 0.45 ***67.21 ± 0.41*** ***67.01 ± 0.32***
Caltech-101	30	Max[1] PN-Max SPN-Max	69.60 ± 1.13 ***72.49 ± 1.69*** ***72.65 ± 0.94***	71.11 ± 0.85 ***73.25 ± 0.42*** ***73.72 ± 0.67***	72.68 ± 1.07 ***74.19 ± 1.10*** ***74.40 ± 0.94***
15-Scenes	100	Max[1] PN-Max SPN-Max	78.11 ± 0.78 ***80.98 ± 0.28*** ***81.12 ± 0.39***	79.69 ± 0.53 ***82.57 ± 0.18*** ***82.15 ± 0.57***	81.53 ± 1.03 ***83.34 ± 0.43*** ***82.74 ± 0.72***

V. Discussions

In this paper, we proposed a simple but effective pooling method called PN-Max pooling .It is a extension of Max pooling method and take the sparse coding efficient in two parts: positive part and negative part. Our method performs the max pooling over each part to get the final representation. Future work will aim at theoretical analysis why the PN-Max pooling method performance better than Max pooling method.

Reference

[1] Yang, J,et.al, "Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification". CVPR, 2009.

[2] Donoho, D.L."Compressed sensing", IEEE Transactions on Information Theory, 2006,52(4),pp: 1289 - 1306

[3] Hubel, D. H and Wiesel, T. N. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". J Physiol, 160:106–154, Jan 1962.

[4] Koenderink, J and Van Doorn, A. "The structure of locally orderless images". IJCV, 31(2/3):159–168, 1999.

[5] Yan Zhu.et.al."Sparse Coding on Local Spatial-Temporal Volumes for Human Action Recognition".ACCV,2010

[6] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, no. 6755, pp. 788–791, 1999.

[7] Lu, Y. M. and M. N. Do, "A theory for sampling signals from a union of subspaces". IEEE Transactions on Signal Processing, 2008, 56(6): 2334-2345.

[8] http://www.ifp.illinois.edu/~jyang29/ScSPM.htm

Visual Duplicate based Topic Linking using a Robust Video Signature

Kota IWAMOTO, Takami SATO, Ryoma OAMI, and Toshiyuki NOMURA
Information and Media Processing Laboratories, NEC Corporation, Japan

Abstract—**This paper proposes a topic linking using a robust video signature to detect visual duplicates for grouping coherent topics in a video archive. The proposed video signature, which was accepted as part of a new ISO/IEC standard "MPEG-7 Video Signature Tools", is designed for robust and high-speed detection of visual duplicates in a large database. It represents intensity differences between various sub-regions in a frame, which provides robustness to various modifications to videos, including caption overlay and compression. We show that the proposed video signature significantly improves the detection rate of visual duplicate segments, by more than 40% under caption overlay, compared with conventional visual features. We also present our topic linking system with its visual presentation of topic groups for efficient browsing and viewing of videos.**

I. INTRODUCTION

With the expansion of broadcast channels and IPTV services, and with the spread of digital video recorders, storing large volumes of video contents has become easy. An efficient method of consuming such vast amounts of videos is essential. Topic tracking or topic linking [1][2], a method of organizing contents based on topics, provides such efficient way of browsing and viewing videos, such as news programs. These methods use visual features of videos to detect visual duplicates between different contents, which are grouped together to form coherent topics groups. It utilizes the fact that the scenes depicting the same topic recurrently use the same video segments from a common source, even across different broadcast channels.

The key to topic linking is the robust detection of visual duplicates from the same source, even with differences caused by modifications such as caption overlay and compression, which are imposed during the editing or storage of contents. For example, different captions are overlaid at different regions on a same video source for different programs or broadcast channels, as shown in Fig. 1. Compression (transcoding) at different bitrates may be applied when recording videos. The conventional features for visual duplicate detection [6]-[8] cannot robustly detect duplicate segments under these modifications, especially at low false alarm rates, and thus are not able to provide an accurate topic linking.

In this paper, we propose a topic linking for video archives using a robust video signature (video fingerprint) developed by the authors. The proposed video signature, which was accepted as part of a new ISO/IEC standard "ISO/IEC 15938-3/Amd.4 MPEG-7 Video Signature Tools" [3]-[5], is designed to be robust to various modifications, and is suited for duplicate detection of short segments in a large database. We also present our topic linking system which visually presents the generated topic groups to the users for efficient browsing and viewing of contents based on topics.

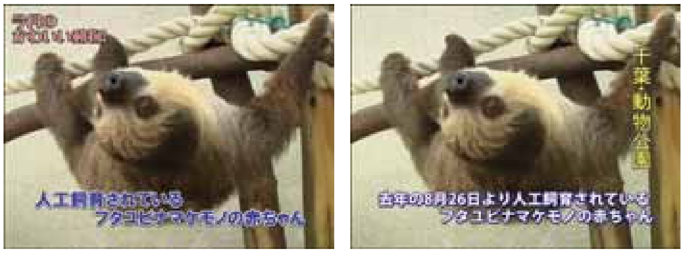

人工飼育されている
フタユビナマケモノの赤ちゃん
去年の8月26日より人工飼育されている
フタユビナマケモノの赤ちゃん
千葉市動物公園

Fig. 1: Example of different caption overlay on a same video source, from news programs of different broadcast channels.

II. PROPOSED ROBUST VIDEO SIGNATURE

The requirements of visual duplicate detection for topic linking are as follows.

(A) Robust detection of short duplicate segments under various modifications, such as caption overlay and compression.
(B) Fast feature extraction and matching for feasible implementation.

Visual features extracted at frame-level have been used for accurate detection of duplicate segments. However, the conventional color-based features such as color histogram [6], and spatial features such as difference block luminance [7] and ordinal measure [8], do not satisfy (A). Local-descriptor based features [9] do not satisfy (B). The proposed video signature has been developed to satisfy both (A) and (B).

A. Video Signature Extraction

The proposed video signature is a frame-level feature, and is composed of a frame signature representing the content of the frame and its confidence value. Fig. 2 illustrates the extraction procedure of these components.

The frame signature represents quantized intensity differences of various sub-regions in a frame. There are total of 380 sub-region pairs, which are configured at various scales, shapes and locations, as shown in Fig. 3, to provide uniqueness and robustness to the feature. Furthermore, they are sampled more densely at the center of the frame, where region of

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

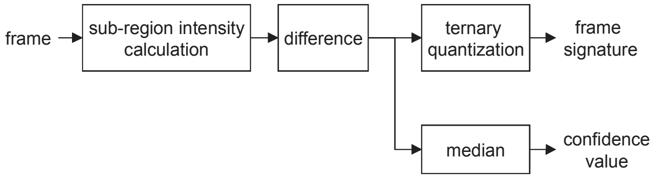


Fig. 2: Extraction procedure of a video signature.

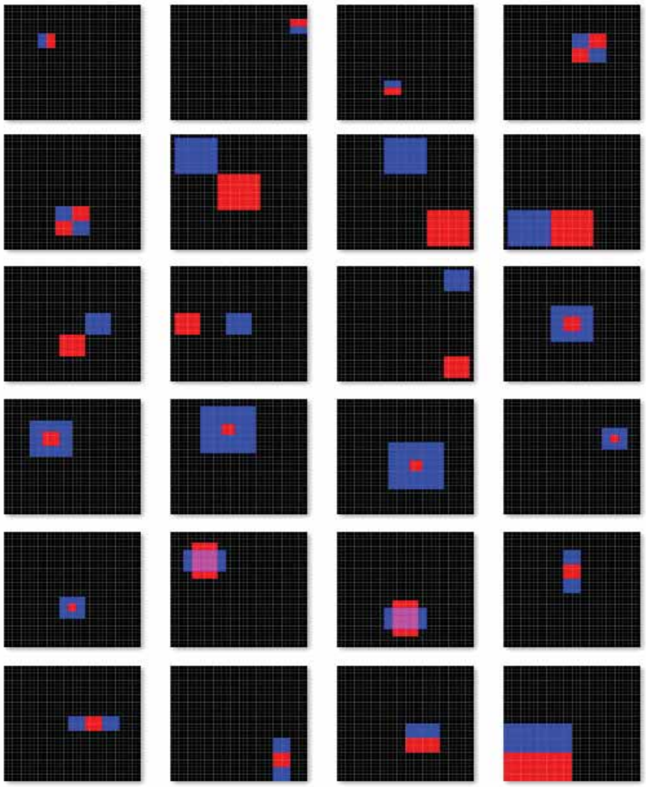


Fig. 3: Sample of sub-region pairs. The black image represents the whole frame region, and the sub-regions within the frame are shown in blue and red. Full specification of the sub-region pairs is described in [5].

interest is more likely to lie. The intensity differences of the sub-regions are quantized into ternary values {0, 1, 2}, which results in a 380 dimensional ternary vector, here denoted as $\mathbf{x}=\{x_1,x_2,\cdots,x_{380}\}$. Let $v1_i$ and $v2_i$ denote the average intensities of the two sub-regions[1] and d_i=$v1_i$-$v2_i$ denote the intensity difference of the sub-region pair for dimension i. The ternary value x_i is calculated by,

$$x_i=\begin{cases}2 & (\text{if} \quad d_i>th) \\ 1 & (\text{if} \quad |d_i|\le th) \\ 0 & (\text{if} \quad d_i<-th)\end{cases}, \tag{1}$$

where *th* is a threshold. The threshold *th* is not fixed, but is adaptively determined for each frame, so that distribution of the quantized ternary value across the vector becomes uniform, i.e. 1/3 each. This quantization strategy provides robustness to changes in intensity ranges, while at the same time maximizing discriminability of the feature.

The extracted ternary vector is compactly encoded into 76 bytes representation. Each group of five consecutive elements in the vector is encoded into one byte value. The encoded value b_j (j=1,…,76) is calculated by the following equation.

$$b_j=81\times x_{5j-4}+27\times x_{5j-3}+9\times x_{5j-2}+3\times x_{5j-1}+x_{5j} \tag{2}$$

This encoding scheme reduces the size of the feature by 20% compared with that of encoding each element with 2 bits representation.

The confidence value represents the complexity of the image and is a measure of reliability of the frame signature. It is calculated by taking the median value of the absolute differences $|d_i|$ of the vector elements, and representing it in one byte value (0-255). Low confidence value means that the intensity differences between sub-regions are small, representing a flat image with little content information. Thus, a frame signature with low confidence value can be considered as less reliable. This is used during the matching to filter out unreliable false matches caused by flat images.

B. Video Signature Matching

Frame-by-frame matching using is carried out between two frame signature sequences to detect duplicate segments. The frame signatures between two frames $\mathbf{x}^1$ and $\mathbf{x}^2$ is matched by calculating the L1 distance between them, given as,

$$dist(\mathbf{x}^1,\mathbf{x}^2)=\sum_{i=1}^{380}\left|x_i^1-x_i^2\right| \tag{3}$$

For fast calculation, the distance can be computed in the encoded domain by using a look-up table. The distance of 1 byte encoded representation corresponding to 5 ternary elements can be pre-calculated in a look-up table. The distance can then be calculated by 76 look-ups and 76 additions, significantly improving the computation speed. A consecutive sequence where the distances are below a certain threshold is detected as a duplicate segment using a matching method described in [10]. Finally, the detected segments which the overall confidence value is low are discarded to filter out false matches caused by flat images. The detected duplicate segments are clustered together to form groups of visually linked segments, which are regarded as topic groups.

III. Topic Linking System For Video Archive

Fig. 4 illustrates the architecture of our topic linking system [11] implemented on a PC. This system records incoming broadcast programs and stores them in a hard disk. At the same time, the video signatures of the video are extracted simultaneously with the recording, and are also stored. The

[1] Thirty-two out of 380 dimensions have only one associated sub-region, and $v2_i$ is defined as a fixed value of $v2_i$ =128 for these dimensions.

video signature size is 77bytes/frame (76 bytes for frame signature, and one byte for confidence value), which is more than 2-3 orders of magnitude smaller than the video contents. The total video signature size for 100 hours of contents is less than 800Mbytes at 30fps. After extraction, vide signature matching is carried out to detect duplicate segments between the newly recorded video and the videos already stored in the database. The duplicate segments are grouped together into topics, and the topic group information is stored in a hard disk. The topic group information is used to present topic based browsing and viewing to the users, which are explained in section IV *B*.

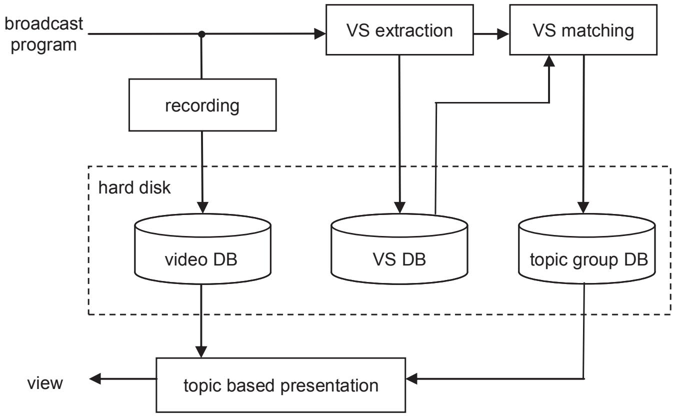


Fig. 4: Architecture of the topic linking system.

IV. Evaluation

A. Performance Evaluation

We evaluated the duplicate detection performance under modifications of the conventional color histogram (CH) [6], difference block luminance (DBL) [7], ordinal measure (OM) [8], and the proposed video signature (VS). Specifications of the conventional visual features are as follows.

- CH: YUV histogram with 64 bins; 32 bins for Y, and 16 bins each for U and V.
- DBL: Binarized block intensity differences using 24×24 block partitioning.
- OM: Ordered ranking of block intensities using 10×10 block partitioning.

Duplicate detection between query videos of 30 minutes and DB videos of 100 hours were carried out, where the duration of each duplicate segment was 5 seconds, assuming a practical use case. This is to simulate a practical use case of a topic linking system, where a newly recorded 30 minutes news program is matched against 100 hours of news programs already recorded, where the duplicate segments between different news reports could be as short 5 seconds. The duplicate segments in the query videos were edited with the following modifications.

(A) caption overlay of 10%-30% region
(B) compression to bitrates of 64Kbps-512Kbps

Total number of duplicate segments was 1,635 for each modification. Detection rate (recall) of the duplicate segments was evaluated, under a threshold which achieved a low false alarm rate (FAR) of 1 false match per 30 minutes of query video. We selected this low FAR for evaluation, as a practical level of false alarm that a user can tolerate.

Table 1 shows the results of the detection rates for each modification. Note that the low FAR used for evaluation was too severe for color histogram (CH) to achieve any meaningful detection. The results show that the proposed video signature (VS) achieves a high detection rate under both modifications. The video signature significantly improves the detection rate under modifications compared with the conventional features. In particular, more than 40% improvement was achieved for caption overlay.

Next, we evaluated the processing time on a PC with a CPU of Core i7-2600 3.40GHz. The extraction time of the VS was 1.47 ms/frame for a 720×480 size video. This enables simultaneous recording and extraction of real-time broadcasts. The matching time between a 30 minute query video and 100 hours DB videos was on average 1.33 minutes, which is much shorter than the query video length. This also means that matching can be done alongside the recording and extraction, and the topic links can be created with just little delay after the broadcast.

The evaluation results on detection performance and processing time show that the proposed video signature is a practical and feasible technology for achieving topic linking of video archives.

Table 1: Detection rate of duplicate segments under modification (at 1 false alarm per 30 minute query video).

modification	CH[6]	DBL[7]	OM[8]	**VS**
caption overlay	0.0% (*)	37.92%	43.79%	**86.06%**
compression	0.0% (*)	4.65%	96.88%	**99.20%**

(*) No detection was achieved due to extremely low FAR used for the evaluation.

B. Topic based Browsing and Viewing

Our topic linking system provides three ways of presenting topic links to the users, a map view, a list view, and a playback view. Fig. 5-7 shows examples of each view on a news program archive.

The map view (Fig. 5) shows the overview of topic groups, with thumbnails of duplicate scenes displayed on a 2D map of time and broadcast channels. The coherent topic groups are marked with the same color in a network. This view helps users to understand the overall structure of the topics, and to decide the topic of their choice. It also allows users to understand the hot topics.

The list view (Fig. 6) presents the thumbnails of duplicate scenes of the same group in a list format. The thumbnails of the duplicate segments are shown in the middle, and the shots preceding and following the duplicate segments are shown to the left and right. This view helps users to compare the news reports from different programs in the same topic group.

Once a video is selected for viewing, users can take advantage of the playback view (Fig. 7), which presents links to scenes in the same topic group by thumbnail popups while watching the video. Users can easily jump to the related contents by clicking on the popups. This enables non-linear viewing of video contents by hopping from one scene to another.

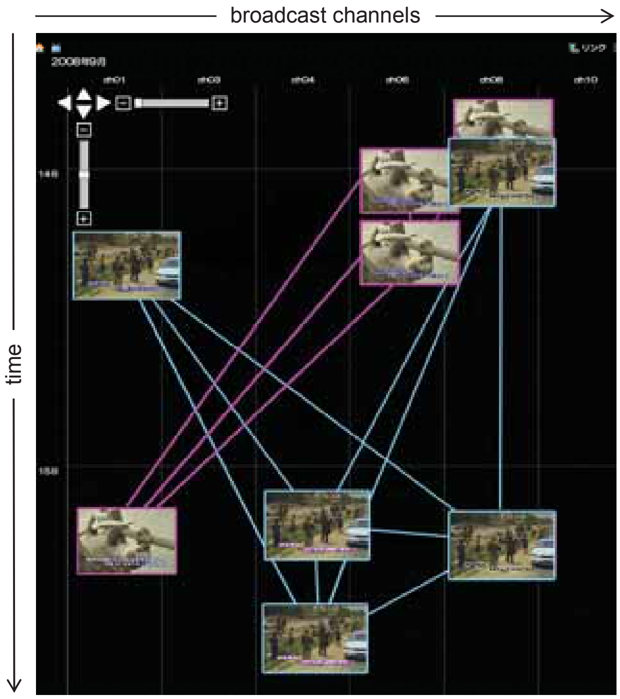


Fig. 5: Map view. Two topics groups are shown here, marked with red and blue respectively.

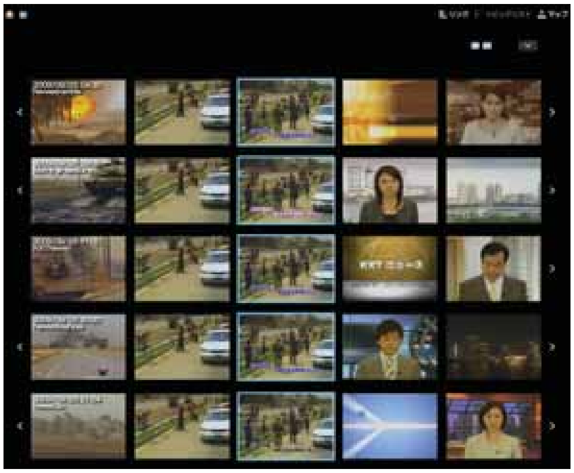


Fig. 6: List view. Scenes from five news reports of the same topic group are listed, with the duplicate segments shown in the middle.

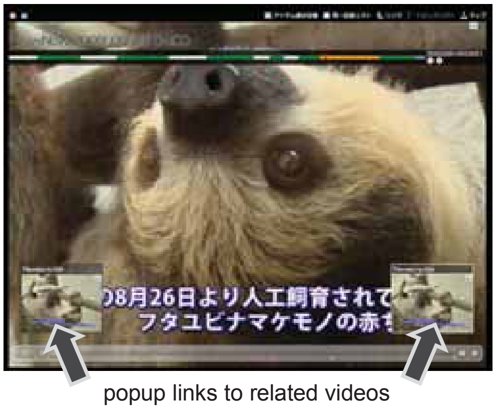


Fig. 7: Playback view. Popup links to the related videos in the same topic group are shown in the left/right bottom of the viewer.

V. Conclusion

We have proposed a topic linking using a robust video signature to detect visual duplicates for grouping coherent topics in a video archive. The proposed video signature is designed for high-speed duplicate detection with robustness to various modifications to videos, including caption overlay and compression. Evaluation results show that the video signature significantly improves the detection rate of duplicate segments, by more than 40% under caption overlay, compared with conventional visual features. We have also presented our topic linking system for efficient browsing and viewing of videos using the detected topic links.

References

[1] W. Hsu and S.-F. Chang, “Topic tracking across broadcast news videos with visual duplicates and semantic concepts”, Proc. of ICIP2006, 2006.

[2] X. Wu, I. Ide, S. Satoh, “Large-scale news topic tracking and key-scene ranking with video near-duplicate constraints”, Proc. of the First ACM workshop on Large-scale multimedia retrieval and mining (LS-MMRM’09), pp.129-136, 2009.

[3] K. Iwamoto, R. Oami, and T. Nomura, “MPEG-7 Video Signature for robust video identification”, Forum on Information Technology, 2011.

[4] S. Paschalakis, K. Iwamoto, P. Brasnett, N. Sprljan, R. Oami, T. Nomura, A. Yamada, and M. Bober, “The MPEG-7 Video Signature Tools for Content Identification”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 22, issue 7, pp.1050-1063, 2012.

[5] ISO/IEC 15938-3:2002/AMD 4:2010, Information Technology - Multimedia content description interface - Part 3: Visual, Amendment 4: Video signature tools.

[6] M. R. Naphade, M. M. Yeung, and B.-L. Yeo, “A novel scheme for fast and efficient video sequence matching using compact signatures”, Proc. of SPIE, Storage and Retrieval for Media Databases, vol.3972, pp.564-572, 2000.

[7] J. Oostveen, T. Kalker, and J. Haitsma, “Feature extraction and a database strategy for video fingerprinting”, Proc. of 5th Int’l Conf. on Recent Advances in Visual Information Systems, pp.117-128, 2002.

[8] X.-S. Hua, X. Chen, and H.-J. Zhang, “Robust video signature based on ordinal measure”, Proc. of ICIP2004, 2004.

[9] C.-Y. Chiu, C.-C. Yang, and C.-S. Chen, “Efficient and effective video copy detection based on spatiotemporal analysis”, Proc. of Ninth IEEE Int’l Symposium on Multimedia (ISM2007), pp.202-209, 2007.

[10] E. Kasutani, R. Oami, A. Yamada, T. Sato, and K. Hirata, "Video material archive system for efficient video editing based on media identification", Proc. of ICME2004, vol.1, pp.727-730, 2004.

[11] H. Kaneko, T. Ozawa, T. Nomura, and K. Iwamoto, “Video identification solution using a video signature”, NEC Technical Journal, vol.6, no.3, 2011.

Vision-based Absolute Indoor Point Positioning in the Hallway without Image Database

Hyunho Lee, Jaehun Kim, Seok Lee, Sanghoon Lee, *Member, IEEE,* Taikjin Lee

***Abstract*--This paper proposes a method for the vision-based absolute point positioning in the hallway using a single camera without image database. Our vision-based absolute point positioning system has less than 1 m error with respect to the root mean square and 0.5 standard deviation.**

I. INTRODUCTION

In recent years, studies about the vision-based indoor navigation is actively researched [3], [4]. Also, accurate initial point positioning is important in indoor navigation. However, the conventional vision-based navigation is impractical because it needs the image database that is hard to make [1]. In this paper, vision-based indoor point positioning without image database is conducted using the depth weighting function and the ratio of two perspective lines in the hallway. We believe that the vision-based indoor point positioning without image database is very powerful solution to build practical navigation systems.

II. THE ALGORITHMS

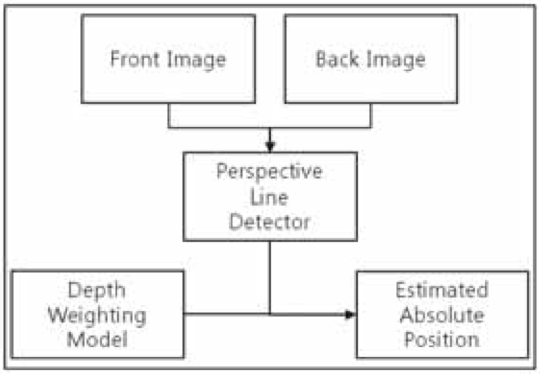


Fig.1 Entire system framework of the absolute indoor point positioning in the hallway

Fig.1 illustrates the entire system framework of the absolute indoor point positioning. From the front and back image captured by any handheld device in the hallway, we can get the length of the perspective line using perspective line detector. And the depth compensation is applied to the measured perspective line because the length of the perspective line from a single camera does not have the depth information. Finally, the absolute point positioning is performed by the ratio between two perspective lines from front and back images.

A. Robust perspective line detection

In the general hallway, there are two perspective lines between wall and ground because it has straight and long structure. The left perspective line appears 45 to 90 degrees and the right perspective line appears 90 to 135 degrees from x-axis because two perspective lines are converged to the vanishing point that is one of possibly several points in a 2D image where lines that are parallel in the 3D source converge. After the image transform using the steerable pyramid decomposition, the energy of the orientation of two perspective lines is gotten [2]. And then the length of the perspective lines is calculated through the Hough transform [5].

$$PL_L = H\left(\zeta\left(I,\phi\right)\right) \tag{1}$$

Where, PL_L is the length of the perspective lines, H is the Hough transform, and $\zeta\left(I,\phi\right)$ is the steerable pyramid decomposition of image I about orientation ϕ. Fig.2 (a) shows the hallway image and Fig.2 (b) shows the result of the perspective line detection.

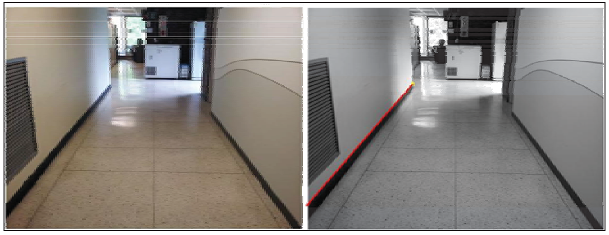


Fig.2. (a) Hallway image and (b) perspective line detection.

B. Depth weighting function

The absolute position can be estimated by the ratio of the front and back perspective lines.

$$P_{est} = \left(\frac{PL_f}{\left(PL_f + PL_b\right)}\right)L \tag{2}$$

Where, P_{est} is the estimated position, PL_f and PL_b are each length of the front and back perspective lines, and L is the total length of the hallway. However, the perspective line cannot reflect the accurate length because it that is estimated by counting the pixel does not have the depth information. Therefore, the length of the perspective line based on counting the pixel should be compensated the depth information. The depth weighting function can be calculated using accurate length and estimated perspective line.

$$\omega\left(PL\right) = \frac{PL_G}{PL} \tag{3}$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Where, $\omega\left(PL\right)$ is the depth weighting function about the length of the perspective line, PL_G is the accurate length of the perspective line, and PL is the estimated length of the perspective line. After the Gaussian smoothing for minimizing the error of estimated length of the perspective line, PL can be modeled as following (4).

$$PL = \alpha_1 e^{\beta_1 \cdot PL_G} + \alpha_2 e^{\beta_2 \cdot PL_G} \tag{4}$$

Where,

$\alpha_1 = 347.5$, $\alpha_2 = -344.4$, $\beta_1 = 0.0028$, $\beta_2 = -0.2535$.

Using (4), (3) is represented as follows:

$$\omega(PL) = \frac{PL_G}{\alpha_1 e^{\beta_1 \cdot PL_G} + \alpha_2 e^{\beta_2 \cdot PL_G}} \tag{5}$$

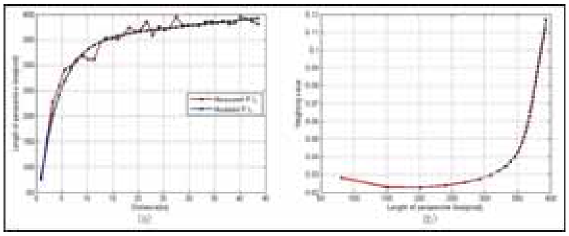


Fig.3 (a) Modeled perspective line and (b) depth weighting function.

Fig.3 (a) shows the measured perspective line and modeled perspective line and fig.3 (b) shows the depth weighting function. And (2) is represented as follows:

$$P_{est} = \left(\frac{PL_f \cdot \omega\left(PL_f\right)}{\left(PL_f \cdot \omega\left(PL_f\right) + PL_b \cdot \omega\left(PL_b\right)\right)} \right) L \tag{6}$$

III. Experimental results

The front and back images were captured at intervals of $1.15m$ with a full resolution of 480×360 pixels using the handheld device. And total length of the hallway was $46m$.

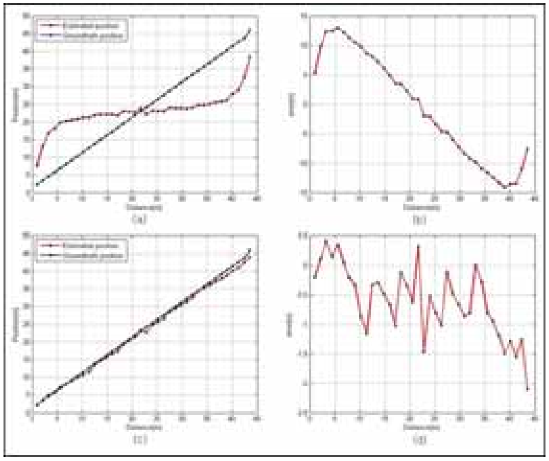


Fig.4 (a) Estimated position without depth weighting function, (b) error without depth weighting function, (c) estimated position with depth weighting function, and (d) error with depth weighting function.

Fig.4 shows the results of the absolute point positioning. In fig.4 (b), Maximum error is about 13 m without the depth weighting function. The root mean square error is 8.11 m and the standard deviation is 4.04. After adapting the depth weighting function, Fig.4 (d) shows that maximum error is about 2 m . The root mean square error is 0.68 m and the standard deviation is 0.5.

IV. Conclusions and further works

We presented a method for the vision-based absolute point positioning of a handheld device in indoor environments using a camera as the sole sensor without image database. Our system does not need to use camera calibration. Our key idea is that the length of the perspective line that does not have depth information is compensated using depth weighting function. Experimental results show that our vision-based absolute point positioning system has less than 1 m error with respect to the root mean square.

For further works, we will test the different method to detect perspective line efficiently. Our next goal is that other indoor structure such as hall, stair, etc. should be considered although the hallway and to develop the vision-based absolute point positioning system in entire indoor environments.

Acknowledgment

This research is supported by the robot R&D program (1415121080) funded by the MKE of the Korean government and also partly supported by NGTN R&D program (1575000034) funded by the KCC of the Korean government and the Pioneer Research Center Program (1415121978) funded by the MEST of the Korean government.

References

[1] J. Courbon, Y. Mezouar, N. Guenard, and P. Martinet, "Vision-based navigation of unmanned aerial vehicles," Control Engineering Practice, Vol. 18, Issue. 7, July 2010, pp. 789-799.

[2] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable Multiscale Transforms," IEEE Transaction on Information Theory, vol. 38, no. 2, March 1992

[3] G. C. Gini, A. Marchi, "Indoor robot navigation with single camera vision," Pattern Recognition in Information Systems, 2002, pp. 67-76

[4] J. Kim, H. Jun, "Vision-based location positioning using augmented reality for indoor navigation," IEEE Transactions on Consumer Electronics, Vol.54, no.3, August 2008, pp. 954-962

[5] D. H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," Pattern Recognition, Vol. 13, Issue. 2, September 1981, pp. 111-122

New Multi-Step Sampling with Adaptive Sampling Patterns in Particle Filtering for Tracking in Surveillance Systems

Zhang Chen, and Wan-Chi Siu, *Fellow, IEEE*
Dept. of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong

***Abstract*—Particle filtering is one of the most efficient approaches for object tracking in video application systems. In this paper, we propose a new multi-step recursive sampling method to replace the conventional direct importance sampling. An online-adaptive sampling pattern for proposal distributions is established. New particles are then sampled recursively from the existing particles with high weights. A 2D predictive transition vector is used to update the pattern of the multivariate Gaussian sampling. Experimental results illustrate that the proposed method reduces computation substantially and it also preserves good tracking results comparable to other algorithms in the literature.**

I. INTRODUCTION

Video Tracking is a significant issue in object monitoring and surveillance systems. It tracks objects and locates their trajectories. Recently, there is much interest in developing efficient algorithms which are robust handling varying situations. Particle filter (PF) [1] is the sequential Monte-Carlo method in setting up a discrete expression of posterior pdf (probability density distribution) with known observations. For the sake of simplicity, sequential importance resampling (SIR)[1], equal state transition probability, and random Walk Transition Model with multivariate Gaussians to generate new particles [2,3] are normally assumed. Conventional approaches, such as multistage sampling [4], MCMC (Markov Chain Monte Carlo) PF [5] and Annealed PF [6] do not use any filtering process to restrict particle sampling, which makes them less efficient.

We propose in this paper a new multi-step recursive sampling in resampling algorithm. Particles are drawn based on the residual particles with the top weights in the previous steps. Thus fewer low-weight particles are used. Only the ones with high weights are kept to form the center of the next sampling pattern and details have been described in [9]. In this paper we propose, rather than using a fixed sampling model, a new adaptive pattern which can react to possible state transitions and keep the efficiency of sampling. Besides we have also used the orientation for rotating to a better sampling region.

II. MULTI-STEP RECURSIVE SAMPLING

A. *Multi-step sampling in particle filtering*

Let us define x$_k$ as the target state at time k and z$_{1:k}$ denotes a set of past k observed appearances of target. x$_0$ is the prior knowledge, z$_k$ is the observation at time k. In particle filtering, the posterior probability density function (pdf) p(x$_k$|z$_{1:k}$) is characterized by a set of N weighted samples or particles, namely,$\{x^i_k, \omega^i_k\}^N_{i=1}$, where *i* identifies the particle, ω^i_k is the weight of particle x^i_k and is usually normalized as $\Sigma_i\omega^i_k=1$. In our tracking algorithm, the objective is to find the closest estimate to the target. Normally, particles with high weight are considered as close candidates to the referenced target. It is assumed that the area nearby higher-weight particles must be more probable to cover the high-likelihood region as well. Hence, new samples are randomly sampled from the existing top-weight ones. The general structure of our algorithm is drawn in a flow chart as Fig. 1. For instance, at step s, the top-weight particle set $\left\{\tilde{x}^j_{k,s-1}\right\}$ can be searched from known all as

$$\left\{\tilde{x}^j_{k,s-1}\right\} = \arg\max\left(\left\{\omega^i_{k,t}\right\}^{i=1:N_s}_{t=0:s-1}\right), \tag{1}$$

where *j* is sorted ascendingly by the weights, N_s is the allocated particle number (N_s) at each step . The s^{th} step particle tran-sition is formulated as $x^i_{k,s} = \tilde{x}^j_{k,s-1} + \varepsilon_s$, (2)

where $\varepsilon_s \sim N(0,Q)$, which is sampled from the multivariate Gaussian distribution with covariance Q and zero mean.
Finally, the posterior density is modified as

$$p(x_k \mid z_{1:k}) \approx \sum_{s=0}^{S}\sum_{i=1}^{N_s}\tilde{\omega}^i_{k,s}\delta(x_k - \tilde{x}^i_{k,s}) \text{ , where S is the size.} \tag{3}$$

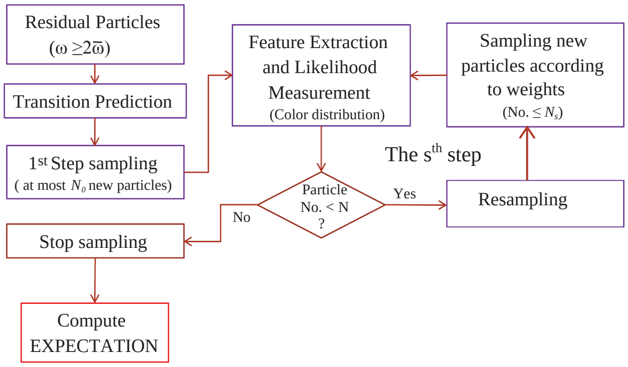


Fig .1 the flow chart of our proposed sampling method

B. *Adaptive sampling pattern*

Let us refer to eqn (2) for particle transition. Usually, its strategy follows the Gaussian distribution.

1) Adaptive Averaging Transition Model [2]

If the state x$_k$ contains the 2D location (x, y) in the frame, the state transition can be predicted by $v^-_k = \lambda\hat{v}_{k-1} + (1-\lambda)v^-_{k-1}$, (4)

where $v^-_k = (v^-_{k,x}, v^-_{k,y})$ stands for the prediction of an adaptive velocity from frame k-1 to k; and $\hat{v}_{k-1}$ is the estimated velocity from k-2 to k-1. The major idea is that the average of Transition relies partially on the prediction, but the most recent transition still has strong influence on the results. Coefficient λ, with value 0.2 obtained from experiments, is a tradeoff of the inertia of previous motions and sudden variation.

2) Adaptive Covariance Matrix

The covariance matrix Q can be Eigen-decomposed [7] into

$$Q = R^TPR,\ \text{where } R = \begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix},\ P = \begin{bmatrix}\sigma^2_{w,k} & 0 \\ 0 & \sigma^2_{h,k}\end{bmatrix} \tag{5}$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

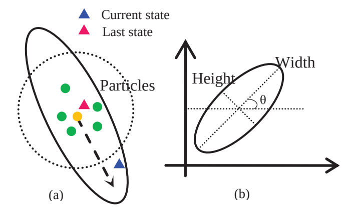


Fig. 2. (a) Simulation of comparison between the new sampling pattern and the standard one. (b) The description of sampling pattern in

where R is the rotation and P is the diagonal covariance matrices. In Fig. 2 (a), the color blobs are particles. The dash-line arrow shows the predicted state transition. Inner part of the dot-line circle stands for the sampling region by the pattern using the diagonal covariance matrix. The solid-line ellipse declares our new pattern in the sampling. It is observed that the covariance matrix can own its orientation (θ) as in Fig. 2(b). It can be computed by following the orientation of the predicted transition vector after eqn (4).

$$\theta = \arctan\left(v_{k,y}^{-} / v_{k,x}^{-}\right),\ \theta \in \left[-0.5\pi, 0.5\pi\right] \tag{6}$$

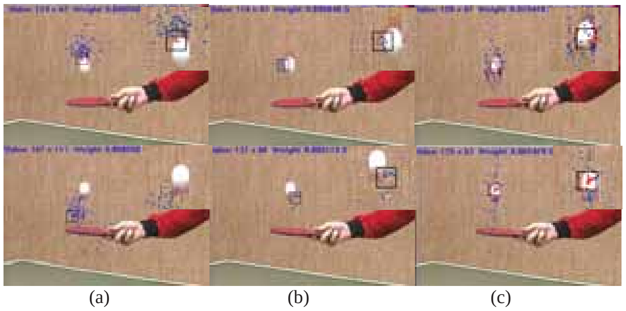
The area with the direction perpendicular to the motion vector stands for a smaller chance of state occurring. Thus we can improve the pattern to cover the possible sampling region. Diagonal matrix P is the constraint which controls the shape of the sampling pattern. Accordingly, we can lengthen the axis (width) following orientation, and shorten the one (height) perpendicularly.

$$\sigma_{w,k} = \begin{cases} \beta D, & \beta D > \sigma_{w,0} \\ \sigma_{w,0} & \end{cases}, \sigma_{h,k} = \gamma \frac{\sigma_{w,0}\sigma_{h,0}}{\sigma_{w,k}} + (1-\gamma)\sigma_{h,0}, \tag{7}$$

where $D = \sqrt{\left(v_{k,x}^{-}\right)^2 + \left(v_{k,y}^{-}\right)^2}$ is the length of the vector v_k^{-} , β (β=1.2) is the coefficient which expands the pattern to assign more probability to the predictively transited area. γ (γ=0.2) is the multiplier that compensate some extreme situation by ignoring the robustness in sampling. Thus, through combining eqns (5), (6) and (7), the new sampling pattern can be updated in the frames with acceptable tracking results.

III. EXPERIMENTAL RESULTS

Much experimental work has been done. An evaluation procedure is carried out in each step. For example, the algorithm is implemented to track the 2-dimensional positions (x, y) of a soccer ball, a ping pang ball or a car. As we only focus on locating objects, we need a clear target (easy to classify) as the subject to be tracked.



(a) (b) (c)

Fig. 3. Comparisons of subjective results by SIR PF (a), Annealed PF (b), and our proposed method (c) in the ball video sequence

The target [8] is represented by its color histogram as the reference model. We take uniform parameters in sampling patterns for different video sequences, i.e. Gaussian distribution with zero mean and the diagonal matrix Q_0 = diag(10^2, 10^2). Initially, SIR (sequential importance resampling) PF [1] uses 100 particles while Annealed PF [6] samples 70 particles in 5 layers. Our proposed algorithm only requires N_0=22 samples initially, and follows by several N_s= 12 samples in steps. The tracking results are illustrated in the Fig. 3. The black box represents the current estimates. The dots in the frame are the current distribution of particles. Using SIR PF, most particles (blue dots) are spreading out in a large area with little contributions. The rest particles with weights higher than the average are drawn in red dots. It can be observed that our proposed method obtained the best tracking results compared with other two methods. In the table tennis sequence, the ball dropped down and was pounced up subsequently. As the velocity dramatically changed, those two methods lost the track of ball, while ours kept sustainable. Also regarding to the particle consumption in tracking, from Fig. 4, the proposed one needs the fewest number of particles (64 in average), while having better subjective results. To sum up, our proposed algorithm is the most stable algorithm in addressing sudden acceleration and occlusion problems and it also allows efficient computation.

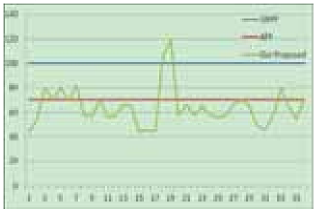


Fig. 4. Comparison of used particle number in frames

IV. CONCLUSION

This paper gives a new multi-step recursive sampling method in particle filtering. It is designed to look for high-weight particles as many as possible in steps. Meanwhile, relying on the averaged transition model, we have proposed a new method which is able to update sampling patterns adaptively. Hence, the sampled particles are more efficient in the estimation process, which is closer to the highly probable area. Experimental results have verified the effectiveness of the proposed method in reducing computation and improving the robustness of our algorithm.

V. REFERENCE

[1] Arulampalam M.S., Maskell S., & Gordon N., Clapp T., “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking”, *IEEE Trans. on Signal Processing*, vol. 50, no.2, pp. 174–88, 2002.

[2] S. K. Zhou, Chellappa R., & Moghaddam B., "Visual tracking and recognition using appearance-adaptive models in particle filters”, *IEEE Trans. on Image Processing*, vol.13, no.11, pp.1491-15, Nov. 2004

[3] Pan P., & Schonfeld D., , "Dynamic Proposal Variance and Optimal Particle Allocation in Particle Filtering for Video Tracking," *IEEE Trans. on CSVT*, vol.18, no.9, pp.1268-1279, Sept. 2008

[4] Bohyung Han, Ying Zhu, Comaniciu, D. & Davis, L.S.,"Visual Tracking by Continuous Density Propagation in Seq. Bayesian Filtering Framework", *IEEE Trans. on PAMI*, vol.31, no.5, pp.919-30,May 2009.

[5] Khan Z., Balch T., & Dellaert F., “MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets,” *IEEE Trans. on PAMI*, Vol. 27, No. 11, pp.1805–19, Nov. 2005.

[6] J. Deutscher, A. Blake, & I. Reid, “Articulated body motion capture by annealed particle filtering”, Proceedings, *CVPR’00*, Hilton Head, SC, USA, pp. 126–133, 2000.

[7] Khan Z.H., Gu I.Y., & Backhouse A.G., "Robust Visual Object Tracking Using Multi-Mode Anisotropic Mean Shift and Particle Filters", *IEEE Trans. on CSVT*, vol.21, no.1, pp.74-87, Jan. 2011

[8] Kin-Yi Yam, Wan-Chi Siu, Ngai-Fong Law, & Chok-Ki Chan, "Effective bi-directional people flow counting for real time surveillance system", Proceedings, *ICCE’11*, Las Vegas, pp. 863-864, Jan. 2011.

[9] Zhang Chen, & Wan-Chi Siu, “Novel Multi-Step Recursive Sampling Strategy for Particle Filtering in Object Tracking”, *Proceedings, Int. Conf. on Signal Processing (ICSP’12),* Beijing, China, Oct. 21-5, 2012

An Extensible Framework for Facial Motion Tracking

Xiaolu SHEN, Xuetao FENG, Jungbae KIM, Hui ZHANG, Youngkyoo HWANG, Ji-yeun KIM

***Abstract*--Facial motion tracking is a challenging task because of highly flexible head pose and facial expression. An extensible tracking framework is proposed in this paper. Within the framework, proper models are selected according to requirements and restrictions of the application, and different trackers can be constructed to handle different tasks. Experimental result shows that our tracker outperforms the existing commercial software.**

I. INTRODUCTION

Computer vision based human-machine interaction technology has been introduced to consumer electronics products recently, such as smart TV controlled by viewer's gesture. In order to achieve natural and smooth user experience, a motion capture system with high stability and precision is required. This paper discusses the facial motion tracking problem, which aims to obtain head pose and facial expressions in video stream. Such tracking results can be utilized in many applications, for example, high quality glassless 3D display according to the location of viewer's eyes, head pose and gaze correction in video communication, virtual character animation driven by user's facial expression, etc.

Tracking facial motion with high robustness and accuracy is difficult because a head may rotate in a large pose range and facial expression is highly non-rigid. This problem has been widely studied in recent decades. Many successful algorithms [1][2][3] have been proposed and several commercial solutions[4][5][6] are available. However, there may be various requirements and restrictions in practice. For example, the required tracking result may be head pose, positions of facial features or parameters that depict facial expressions. The product could be built on hardware platforms with significantly different computing power. Video source could be full HD camera or ordinary webcam. No perfect solution can handle all possible conditions. To solve this problem, an extensible tracking framework is proposed in this paper.

In our framework, multiple models are used to depict the pose, shape and appearance of a face. They have different capability and limitation, but can work together in a unified framework. We can select a subset of them to build a tracking engine according to the requirement and restriction. A system designer can also add new models into the framework, to fulfill his own condition.

In Table I, 14 different models are listed and labeled with their spatial and temporal attributes. Some of them are generated during online tracking process, and others are constructed by offline learning from huge amount of data. Some of them carry global information of a face, and others describe local features. These models are further addressed in Section 2.

TABLE I
14 MODELS FOR TRACKING FACIAL MOTION

Building time	Scale	Model Content
Offline	Global	2D shape/appearance
Offline	Global	3D shape
Offline	Global	2D/3D deform energy
Offline	Global	Prior pose
Offline	Local	2D shape/appearance
Offline	Local	2D/3D feature point
Key frame	Global	Skin
Key frame	Local	2D/3D feature point
Online	Global	Texture
Online	Local	2D/3D feature point

All selected models can be utilized in a unified optimization procedure to estimate the motion of a face, which is called model fitting. By solving the energy minimization problem in a stage-wise manner, model fitting is implemented with high efficiency and good convergence. The flow chart of our tracking system is shown in Figure 1.

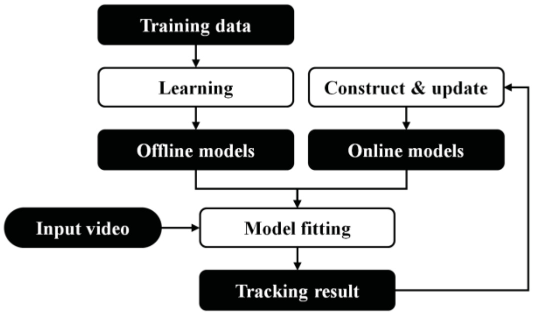


Fig. 1. Structure of facial motion tracking

The rest of this paper is organized as follows. Section 2 presents the 14 models with their construction method, usage and limitation. Section 3 addresses the stage-wise model fitting algorithm. Section 4 presents two implementations according to our framework: a facial key point tracker and an eye tracker. Finally, Section 5 gives the conclusion.

II. MODEL REPRESENTATION AND CONSTRUCTION

Most of facial tracking systems aim to find the 2D or 3D position of facial key points, which can be expressed by a 2D shape $\mathbf{s}$ and a 3D shape $\bar{\mathbf{s}}$. Following [1], target of facial motion tracking turns to finding the flexible deformation parameter $\mathbf{p} = (\mathbf{p}^{2D}, \mathbf{p}^{3D})$ and the rigid transformation parameter $\mathbf{q} = (\mathbf{q}^{2D}, \mathbf{q}^{3D})$, which is realized by minimizing the cost functions constructed from the models listed in Table 1. They are explained as follows:

A. *Offline Global 2D Shape/Appearance*

The face appearance is represented by a linear model as in [2]:

$$A = A_0 + \sum_i \lambda_i A_i$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

The appearance model is used to get 2D shape parameters by minimizing the error between synthesized appearance and input face:

$$E_{\mathrm{global2D}} = \left\| \mathrm{A} - \mathrm{I}\left(\mathrm{N}(\mathrm{W}(\mathbf{x}, \mathbf{p}^{2D}), \mathbf{q}^{2D})\right)\right\|^2 \tag{1}$$

where $\mathrm{N}(\mathrm{W}(\mathbf{x}, \mathrm{p}^{2D}), \mathrm{q}^{2D})$ is the warped coordinate of a pixel **x** in the mean shape mesh.

In order to cover the pose range from -90^o to $+90^o$, 5 groups of shape and appearance model corresponding to different poses are defined, including frontal models, left and right half profile models, and left and right full profile models.

2D global shape/appearance model is the most effective one, and acts as the kernel of tracking algorithm. To get the best performance, only face images with the poses, expressions and illuminations required by usage scenario should be selected to form the training database of shape and appearance model.

B. Offline Global 3D shape

We use a 3D model (Figure 2) modified from the widely used Candide-3 [7] to evaluate the 2D-3D shape difference as in [8]:

$$E_{3D} = \left\| \mathbf{s}(\mathbf{p}^{2D}, \mathbf{q}^{2D}) - \mathrm{P}\left(\bar{\mathbf{s}}(\mathbf{p}^{3D}, \mathbf{q}^{3D})\right)\right\|^2 \tag{2}$$

where P is the perspective projection that is applied to 3D shape $\bar{\mathbf{s}}$.

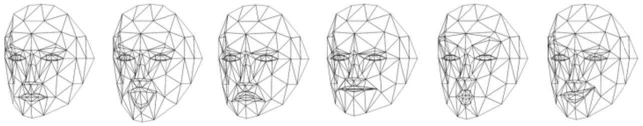


Fig. 2. Shape samples generated by the 3D shape model

Offline global 3D shape model is the first choice when 3D pose information is demanded. It also prevents unnatural 2D shape.

C. Global 2D/3D Deform Energy

In order to avoid twisted shape, we employ the following two energy models:

$$E_{\mathrm{Energy2D}} = \|\mathbf{e}^{-1}\mathbf{p}^{2D}\|^2,\ E_{\mathrm{Energy3D}} = \|\mathbf{p}^{3D}\|^2$$

where **e** is a diagonal matrix comprised of the eigen values of 2D shape covariance matrix. These items are very fast to compute. Adding weight to them can enhance stability if users do not tend to perform strong expression.

D. Offline Global Prior Pose

If head pose $\mathbf{q}_0$ can be learned independently from modules such as face detector and model choosing, the following model can be used to adjust $\mathbf{q}^{3D}$:

$$E_{\mathrm{Pose}} = \|\mathbf{q}^{3D} - \mathbf{q_0}\|^2$$

E. Offline Local 2D Shape/Appearance

To achieve high precision as well as stability, local shape and appearance models corresponding to isolated facial components are built. Since changes in different facial component can be approximated as independent, we need much less sample to train separate local models comparing with a global model with the same flexibility.

Local models are similar to that defined in paragraph II.A, but have smaller scale and higher resolution, covering different facial regions as shown in Figure 3.

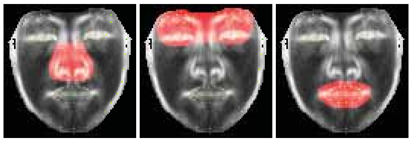


Fig. 3. Local 2D shape/appearance models: nose, eyes and mouth.

Besides, to prevent the local models from stuck into local minimum, it is required that key points in local models should not drift too far away from the global shape as:

$$\begin{aligned} E_{\mathrm{Local2D}} = &\left\| \mathrm{A} - \mathrm{I}\left(\mathrm{N}(\mathrm{W}(\mathbf{x}, \mathbf{p}^{2D}), \mathbf{q}^{2D})\right)\right\|^2 \\ &+ \mathrm{w_s}\|\mathbf{s}_{\mathrm{local}}(\mathbf{p}^{2D}, \mathbf{q}^{2D}) - \mathbf{s}\|^2 \end{aligned} \tag{3}$$

where $\mathrm{w_s}$ is the weight of global-local shape difference item.

This model is useful when particular facial component is of special concern, such as in lip reading and eye status detection. It is a bit time consuming, since pixel warp and appearance calculation is repeated in each iteration step.

F. Offline Local 2D/3D Feature Point

The purpose of offline feature point model is to find the position of each feature point detected from input image. Here each feature point has a fixed position with respect to the mean shape mesh, described by barycentric coordinates in triangles, and is modeled by a pre-constructed classifier.

The first step is to decide which feature points to model. A FAST corner detector [9] is used to find all local feature points of multiple scales in a database containing 4000 face images of various pose, illumination and expression. The training DB is divided into 15 subsets according to head pose. By backwarping feature points to the 15 mean shape meshes, the ones with high repeatability can be found, as illustrated in Figure 4.

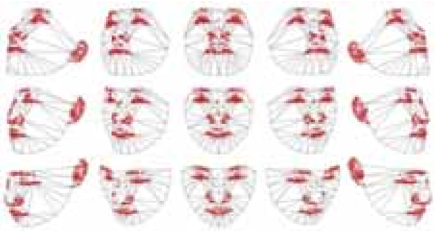


Fig. 4. Local feature points that have high repeatability

Each offline feature point defines a class. We use local random binary features to represent them, and train random ferns classifiers with original images plus artificially deformed samples.

During tracking process, all classifiers are used to recognize detected feature points. The current pose type is determined by the classifier which gives the most positive results. Since their positions with respect to the mean shape mesh are known, we can find the required parameter **p** and **q** by deforming the shape model mesh.

Let $\mathrm{v_i}$ be the coordinates of the detected feature point, $\mathrm{V_i}$ be the corresponding offline feature point in 2D mean shape mesh and $\mathrm{V_i}(\mathbf{p}^{2D}, \mathbf{q}^{2D})$ be its position warped by a certain parameter. Objective of tracking is to minimize their distance:

$$\sum_i \|V_i(\mathbf{p}^{2D},\mathbf{q}^{2D}) - v_i\|^2 \quad (6)$$

To handle incorrect result in classification, we use a robust error function [10] to suppress large error and exclude outlier:

$$E_{Offline2Dpoint} = \sum_i \rho(\|V_i(\mathbf{p}^{2D},\mathbf{q}^{2D}) - v_i\|,\gamma)^2 \quad (7)$$

Similarly, such constraint can be applied to 3D model:

$$E_{Offline3Dpoint} = \sum_i \rho(\|P(U_i(\mathbf{p}^{3D},\mathbf{q}^{3D})) - u_i\|,\gamma)^2 \quad (8)$$

where u_i is the detected feature point, U_i is class point in 3D mean shape mesh and P is a projection transform.

When the faces in input image and training image have obvious and stable details (e.g. under fine illumination and high-resolution camera), the offline feature point models are particular effective in spite of time consumption in feature extraction and classification.

G. Key frame Global Skin

Color delivers strong prior information for facial area. We represent skin / background color with a K-cluster mixture Gaussian model. For a pixel with RGB color v, its probability of belonging to either of the two categories has the following form:

$$G(v) = \sum_{k=1}^{K} w_k \frac{1}{|\Sigma_k|^{0.5}} \exp\left\{-\frac{1}{2}(v-\mu_k)^T \Sigma_k^{-1}(v-\mu_k)\right\}$$

Skin and background models are built from well-tracked key frame after a tracking failure or model switching, in case of user or illumination change.

For each video frame during tracking, a skin map $C(x)$is computed from distance map of skin/back binary result as in Figure 5.

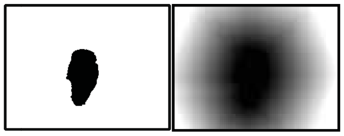


Fig. 5. Skin binary result (left) and skin map (right)

This model sum up skin map value on all 2D shape vertexes, and has the following form:

$$E_{skin} = \left\| C\left(N(\mathbf{s}(\mathbf{p^{2D}});\mathbf{q^{2D}})\right)\right\|^2$$

Mixture Gaussian model and distance map both take some computation time. This model contributes additional stability for fast motion and complex background situation as long as illumination is soft and skin color is stable. Aside from skin color, any kind of foreground/background binary information can be fit into the framework in this manner.

H. Key frame local 2D/3D Feature Point

This model is similar to offline local feature point described in paragraph II.F only that it doesn't depend on learnt model. Instead, we create a key frame space spanned by x, y and z rotation angle, and divide it evenly by 5 degrees into $L_1 \times L_2 \times L_3$ bins. Each time the head rotates to an empty bin and get a successful result with mild expression, we put the frame into this bin as key frame along with its feature point information, i.e., coordinates with respect to shape mesh and feature descriptors. Once tracking fails, the key frame space is empted.

For an input frame, we match its feature points with those in all key frames, and refine 2D/3D parameter as in (7) and (8). This model relies more on intra-user consistency than offline point feature.

I. Online Global Texture

The role of online global texture is very similar to the appearance model used in II.A. The difference is, instead of synthesized appearance as in (1), this item tries to fit the image to the appearance of current user, denoted by T. The difference between T and input face is minimized as:

$$E_{Texture} = \left\|T - I\left(N(W(\mathbf{x},\mathbf{p}^{2D}),\mathbf{q}^{2D})\right)\right\|^2 \quad (9)$$

The texture template T comes from the first successfully tracked frame, and is updated when the tracking error is small enough or after a tracking failure.

When input face cannot be expressed by linear combination of appearance model, such as under large pose and unfamiliar look, the online global texture will play a complementary role.

J. Online Local 2D/3D Feature Point

Online local point feature resembles key frame feature point model very much except for that it uses the previous frame's feature points instead of key frame.

Among the three point feature models in paragraph II.F, II.H and II.J, online model has the shortest time span, and is the most accurate one under stable environment in terms of feature matching. In contrast, offline model does not depend on previous tracking, and thus has the ability to rectify drifting. The demerit of offline model is that offline training process is time consuming and the classifiers require large memory. And key frame feature point model is the compromise of the other two. These models should be chosen according to circumstance.

III. Model Fitting

Target of model fitting is to find the optimal parameter **p** and **q**, which can be achieved by minimizing the following cost function:

$$E(\mathbf{p},\mathbf{q}) = \sum_i w_i E_i \quad (10)$$

where E_i is the cost items defined by each model that have been described in section II, and w_i is the weight of each item.

Cost items in (10) have different effects. Some have high precision but is easy to fall into local minimum while others have good monotonicity but lack accuracy. Therefore, a stage-wise optimization strategy is designed as follows to guarantee the precision, stability and efficiency.

1) Predict stage

Minimize feature point model related items to estimate **p** and **q**. As introduced in section II, they are based on recognition or matching result of key points of limited number, results in low precision. But the cost item is defined by distance between points, which provides good convergence.

2) Rough fitting stage

Minimize all items except the one defined by local 2D shape / appearance models to estimate rigid transformation parameter **q**. In this stage, parameters are still far from their optimal positions. If **p** and **q** are updated simultaneously, high dimension will make the problem unstable.

3) Fine fitting stage

Minimize all items except the one defined by local 2D shape / appearance to estimate **p** and **q** to adjust rigid transformation and non-rigid deformation simultaneously.

4) Refinement stage

Minimize the cost item defined by local 2D shape /appearance model to update **p** and **q** in local models. It can be considered as a further refinement to the location of facial key points, to improve the precision for strong expressions.

With the stage-wise model fitting strategy, high stability, precision and efficiency can be achieved. Besides, all the items in (10) can be optimized in an invert compositional manner [11], which reduces the computation cost remarkably and therefore increase the speed.

IV. EXPERIMENT

We build two trackers with different combination of models for two different tasks to show the flexibility of our framework.

A. Facial Expression Tracker

We build this tracker with the models that are introduced in paragraph II.A, II.B, II.C, II.E, II.F, II.I and II.J for both high accuracy and stability. Training samples includes all kinds of variations such as pose, expressions, illumination and races.

We compare this tracker with a representative commercial face tracking software [4] designed for a similar purpose. On a testing database including 20 videos (10 of large poses and 10 of strong expressions), we measure the average localization errors of eye corners and mouth corners. Our tracker exhibits higher accuracy as in Table II.

TABLE II
AVERAGE ERROR COMPARISON [a]

	Our tracker	Software [4]
Pose videos	0.062	0.129
Expression videos	0.051	0.075
All videos	0.055	0.095

[a] Error is calculated by comparing tracking result with manually labeled result and normalized by pupil distance.

Besides, our tracker is more stable during high speed motion, large pose and unfamiliar appearance such as thick frame glasses, as shown in Figure 6.

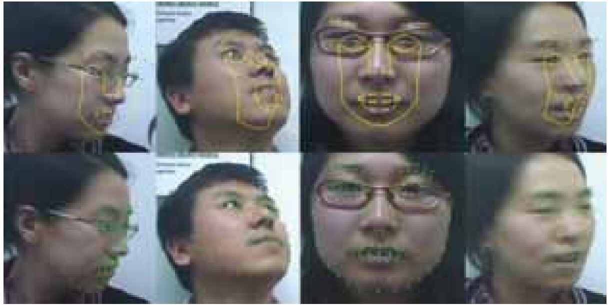


Fig. 6. Tracking result of a commercial software [4] (1^{st} row) and our tracker (2^{nd} row)

B. Eye Tracker

We build this tracker with fewer models, including that introduced in paragraph II.A, II.B, II.C, II.I and II.J. Besides, we only construct models covering front and half profile poses, and with fewer shape components. This tracker aims to locate eye position stably under uncontrolled light condition with very high speed.

Fig. 7. Examples of eye tracker testing database

The testing database consists of 270 videos (Figure 7). The frame size is 1280×720. The videos are captured under light conditions of large range intensity and strong directional contrast, and at distances from 0.5m to 3.0m. This tracker achieves average eye center localization error of 0.067 pupil distance, and speed at about 50 fps with a 900MHz Intel Celeron processor.

V. CONCLUSION

This paper proposes an extensible facial motion tracking framework. Multiple models can be selected and combined together to construct a cost function. For different application scenarios, we can choose different models to adapt the tracker to various requirement and restrictions. By minimizing the cost function, robust and accuracy tracking result can be achieved. Such facial motion tracker can be widely used as an attractive and convenient human-computer interaction medium in smart TV, handheld device, intelligent house, security system, media search engine, etc.

REFERENCE

[1] T. Cootes, C. Taylor, D. Cooper, et al., Active shape models - their training and application, Computer Vision and Image Understanding, 1995, 61(1): 38-59

[2] T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, European Conference on Computer Vision, 1998

[3] D. Cristinacce and T. Cootes, Feature detection and tracking with constrained local models. British Machine Vision Conference, 2006

[4] http://www.seeingmachines.com/product/faceapi

[5] http://www.mobinex.com/technology.htm

[6] http://www.oki.com/jp/fse/development

[7] http://www.icg.isy.liu.se/candide/

[8] J. Xiao, S. Baker, I. Matthews, et al., Real-time combined 2D+3D Active Appearance Models, IEEE Conference on Computer Vision and Pattern Recognition, 2004

[9] E. Rosten and T. Drummond, Machine learning for high-speed corner detection, European Conference on Computer Vision, 2006

[10] Julien Pilet, Vincent Lepetit, Pascal Fua, Real-time Non-Rigid Surface Detection, IEEE Conference on Computer Vision and Pattern Recognition, 2005

[11] Iain Matthews, Simon Baker, Active appearance models revisited, International Journal on Computer Vision, 2004, 60(2):135-164

Flash Image Quality Enhancement by Compensating the Quantity of Flash Light

Sung-Kwang Cho, Won-Ho Cho, and Tae-Chan Kim, *Member, IEEE*
Samsung Electronics Co., Ltd.

***Abstract*--The most of mobile phone camera are equipped with LED flash to acquire image under dark illuminant. However, the power of LED flash is not enough to illuminate the background area of scene as it becomes diminished in a distance. Moreover, mobile phone camera still suffers from low dynamic range. The facts bring unbalanced image in terms of brightness. To solve this problem, we propose method that estimates the intensity of flash light and compensate the difference of flash using flash and no-flash image pair. We confirmed the promising performance in an experiment using mobile phone camera.**

I. INTRODUCTION

Camera flash is used to brighten a dark environment. However, using flash often results in unnatural image with defects. To solve the problem, there have been many researches in computational photography field. Red eye removal has been very famous research in flash imaging [1]. In [2], they proposed method to get high dynamic range using several images with the different level of flash. In [3, 4, 5], they developed the way to acquire natural flash image combining flash and no-flash image pair. Besides, [5] also suggested the way to remove shadow which is caused by flash. [6, 7] show that several images with different flash angles are used to get the detail of object image or silhouette information. However, none of the methods provides way to modify flash intensity to enhance flash image quality.

In this paper, we propose method to improve the artifacts of flash image with darkened and saturated brightness using flash and no-flash image pair. The problem is mainly caused by the weak power of LED flash and the failure of auto exposure caused by limited dynamic range of mobile image sensor. The most similar works are [3, 4, 5] as they use flash and no-flash image pair. The goal of those methods is to recover the mood of no-flash image. It combines large scale of no-flash image with detail and color information of flash image. Meanwhile, in our method, we focus on controlling flash effect of flash image layers which are segmented into background and foreground based on derived flash intensity.

II. PROPOSED METHOD

In the experiment, input images are acquired by commercial mobile phone (640 × 480, 'jpg' format). The first image is taken with long exposure time without flash, and the second image is taken with short exposure time turning on flash under low illuminant condition (Fig.1). Fig.2 shows the block diagram of our method to enhance flash image quality. We will explain the processing blocks in the following sections.

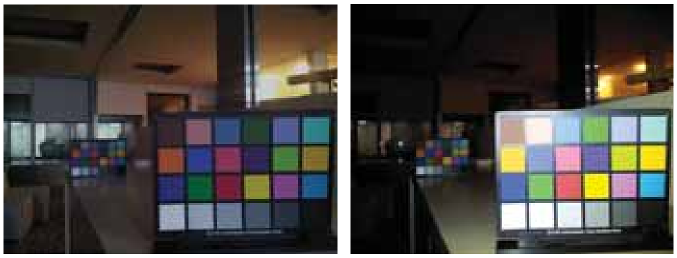


Fig. 1. Input image pair. (Left) Long exposure time without flash. (Right) Short exposure time with flash. The center is overexposed and the background becomes too dark. There are more noise in the left image.

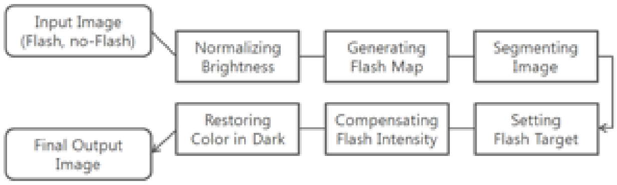


Fig. 2. The block diagram of proposed method.

A. Normalizing the brightness of image

The first step is to normalize the brightness of input images to put input images in a linear space. Mobile phone camera automatically calculates exposure time according to the average brightness of input scene. So that flash and no-flash images are in the different linear space with different exposure time. We find the relative exposure time by fitting the level of dark area.

$$I_{o\_norm} = I_0 / e_0 , \quad I_{f\_norm} = I_f / e_f \qquad (1)$$

Where, $\boldsymbol{I_o}$ and $\boldsymbol{I_f}$ are the intensity values of flash image and no-flash image, and $\boldsymbol{I_{o\_norm}}$ and $\boldsymbol{I_{f\_norm}}$ are the normalized value of $\boldsymbol{I_o}$ and $\boldsymbol{I_f}$. $\boldsymbol{e_o}$ and $\boldsymbol{e_f}$ are the relative exposure time of image pair. We set $\boldsymbol{e_f}$ as 1, $\boldsymbol{e_o}$ is close to 5 in this experiment.

B. Generating flash intensity map

The next step is to generate flash intensity map which is needed to segment image and set flash target in the latter step. With flash and no-flash images, we can estimate both flash intensity and flash radiance of object surface. Flash radiance is proportional to flash intensity and the reflectance rate of object. Flash radiance corresponds to the different intensity values between normalized inputs. Here, we assume that the intensity of no-flash image represents the reflection rate of object surface.

$$D_f = I_{f\_norm} - I_{o\_norm}, \quad Q_f = D_f \,/\, I_o \qquad (2)$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

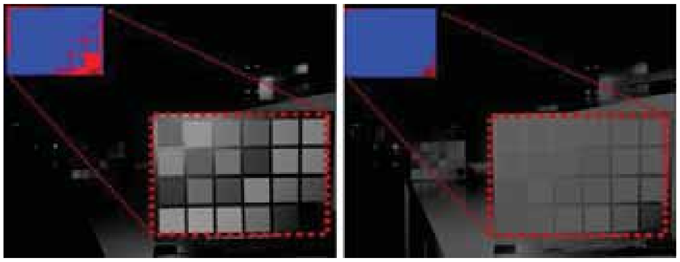


Fig. 3. (Left) The difference of image intensity ($\boldsymbol{D_f}$). (Right) Visualized flash intensity map near object surface ($\boldsymbol{Q_f}$).

Where, $\boldsymbol{D_f}$ is the intensity difference between input image pair and it refers to flash radiance. $\boldsymbol{Q_f}$ is an estimated flash intensity.

C. Segmenting image

With $\boldsymbol{Q_f}$, we segment image into foreground and background layers based on the mean of $\boldsymbol{Q_f}$. Only with $\boldsymbol{D_f}$, it is hard to segment image because it cannot determine whether it is in front or back side when they have similar intensities. $\boldsymbol{Q_f}$ varies with the distance of object regardless of the reflectance rate so that it is useful to segment image into layers with different distance. In Fig.3, $\boldsymbol{D_f}$ has different level according to the brightness near Macbeth chart. However, $\boldsymbol{Q_f}$ has flatter level than $\boldsymbol{D_f}$. It shows that $\boldsymbol{Q_f}$ can represent the distance of object from flash point.

D. Compensating flash intensity

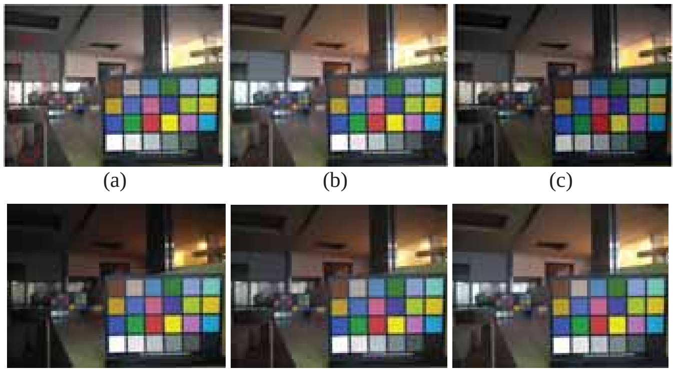
In this step, we compensate flash intensity to manipulate the unbalanced brightness of flash image. As shown in Fig.1 (right), the brightness of the center of flash image is over saturated and the background is relatively dark. First, we need to set the target value of flash intensity ($\boldsymbol{T_f}$). The segmentation information is used to get $\boldsymbol{T_f}$. We set the mean value of flash intensity in foreground layer as $\boldsymbol{T_f}$. Compensated flash image can be derived as below.

$$C_{off} = (Q_f - T_f) \times I_o, \quad I'_f = I_f + F\{C_{off}\} \times K \qquad (3)$$

Where, $\boldsymbol{C_{off}}$ is compensation offsets which is added into $\boldsymbol{I_f}$. $\boldsymbol{I'_f}$ is the compensated flash image. $\boldsymbol{K}$ is parameter to control compensation level. ***F{ }*** refers to bilateral filter which removes noise and keep large scale [8]. It is required to filter $\boldsymbol{C_{off}}$ because there is noise from no-flash image ($\boldsymbol{I_o}$). Fig.4 (a) shows processed image. The background becomes brighter and oversaturated area in the foreground is removed.

E. Restoring color information

In the above step, flash intensity is compensated in the foreground and background. However, Fig.4 (a) shows that the color of background is still faded even after casting the color information from flash image into $\boldsymbol{I'_f}$. To solve this problem, we cast color information from no-flash image into the background layer of $\boldsymbol{I'_f}$. As we can see in Fig.4 (b), the color of background layer is recovered.



(a) (b) (c)

(d)

Fig. 4. (a) Result image without color casting in background (*K*=1). (b) Result image (*K*=1). (c) Result image using previous method (d) Result images with different *K* values (*K*=0.25, 0.5, 0.75) in background layer.

F. Results

We compare our result with previous method [3, 4, 5], which merges the large scale of no-flash image and the edge and color information of flash image (Fig.4 (c)). However, it does not selectively control each intensity value of segmented layers. Moreover, the color of background area is faded because the background of flash image is too dark to carry color information. Meanwhile, with our method, the color of background layer is recovered by casting the color of no-flash image into final output image (Fig.4 (b)). Besides, user can control the flash intensity of each layer in the several levels according to the parameter (***K***) (Fig.4 (d)).

III. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed method to enhance flash image quality compensating flash effect based on flash information of flash and no-flash image pair. However, we did not consider the other important processing units in flash imaging, such as shadow correction, motion compensation and white balancing. In the future work, we are planning to work on those methods to get better result.

REFERENCES

[1] Gaubatz and Ulichney, “Automatic red-eye detection and correction,” in IEEE Int. Conf. on Image Processing, 2002.

[2] H. Hoppe and K. Toyama, "Continuous flash," Technical Report MSR-TR-2003-63, Microsoft Corporation, October. 2003.

[3] D. Krishnan_ R. Fergus, "Dark flash photography,” in ACM Trans. Graphics, pages 1–11, 2009.

[4] Petschnigg, Agrawala, Hoppe, Szeliski, Cohen, and Toyama, “Digital photography with flash and no-flash image pairs,” in ACM Trans. on Graphics, 2004.

[5] E. Eisemann, F. Durand, "Flash photography enhancement via intrinsic relighting," in ACM Trans. on Graphics, Vol. 23, Issue 3, Aug 2004

[6] Akers, Losasso, Klingner, Agrawala, Rick, Hanrahan, “Conveying shape and features with image-based relighting,”14th IEEE Visualization 2003.

[7] R. Raskar, K. Tan, R. Feris, J. Yu, M. Turk, "Non-photorealistic Camera:Depth Edge Detection and Stylized Rendering using Multi-Flash Imaging", in ACM (TOG), Vol.23, Issue 3, Aug 2004.

[8] Durand and Dorsey, “Fast bilateral filtering for the display of highdynamic-range images,” ACM Trans. on Graphics, vol. 21, 2002.

Real-Time Digital Zooming for Mobile Consumer Cameras Using Directionally Adaptive Image Interpolation and Restoration

Wonseok Kang[1], Jaehwan Jeon[1], Eunjung Chae[1], Minkyu Park[2], and Joonki Paik[1]
[1]Image Processing and Intelligent Systems Laboratory, Graduate School of Advanced Imaging Science, Multimedia, and Film, Chung-Ang University, Seoul, Korea
[2]Camera module Laboratory, Advanced R&D Team, Samsung Electronics, Suwon, Kyunggi-Do, Korea

Abstract— **In this digest, we present a novel real-time digital zooming method based on directionally adaptive image interpolation and restoration. The proposed method first estimates an edge direction using steerable filters and performs weighted smoothing along the estimated edge direction. Bi-cubic and bi-linear interpolations are selectively used according to the estimated edge direction. Degradation and artifacts caused by interpolation are removed by employing a directionally adaptive truncated constrained least squares (TCLS) filter. The proposed digital zooming method followed by image restoration provides high-quality magnified images which are similar to the result of computationally intensive super-resolution algorithms. The proposed method can be applied to real-time image processing, and embedded in the form of the finite impulse response (FIR) filtering structure. It is suitable for digital zooming system of mobile phone cameras, tablet PCs, and digital camcorders.**

I. INTRODUCTION

Mobile phones generally have limited–resolution, low-cost camera modules. Therefore, most mobile phones use digital zooming function instead of bulky, power-consuming optical zooming. Digital zooming, however, results in various interpolation artifacts that make the image unacceptably degraded especially with a high zooming ratio. To solve this problem, many image interpolation methods have been proposed, but are not suitable for the digital zooming function of mobile devices which have limited computational power and memory space [1][2].

In this digest, we present a novel digital zooming system which is able to minimize image degradation for various mobile applications. The proposed method is precisely estimates edge direction using steerable filters [3] with an edge smoothing method. The input image is then adaptively interpolated along the estimated edge direction. As a result the proposed method can provide digitally zoomed images without interpolation artifacts such as aliasing and jagged edges. Image details lost in the interpolation process are also recovered using the directionally adaptive truncated constrained least-squares (TCLS) filter [4].

II. DIRECTIONALLY ADAPTIVE IMAGE INTERPOLATION AND RESTORATION

The proposed directionally adaptive image interpolation and restoration method is shown in Fig. 1.

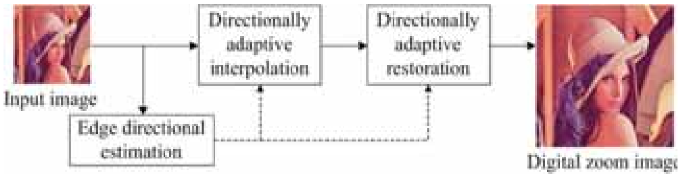

Input image
Directionally adaptive interpolation
Directionally adaptive restoration
Edge directional estimation
Digital zoom image

Fig. 1. The proposed directionally adaptive image interpolation and restoration framework.

A. *Estimation of Edge Direction Using Steerable Filters and Edge Smoothing*

In order to determine the edge direction of the input image, we use steerable filters [3] and edge smoothing. The resulting image of the steerable filter is obtained by the convolving the input image and a steerable filter as

$$R^{\theta} = G^{\theta} * I, \tag{1}$$

where G^{θ} represents a 5×5 steerable filter, I the input image, and $\theta \in \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$. The optimum kernel orientation for directional filtering can be estimated as

$$R^{*} = \arg\min_{\theta} \left\{R^{\theta}\right\}. \tag{2}$$

The direction of an edge is classified into one of four ranges of angle, such as $\theta \in \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$. The estimated edge direction is further smoothed using a weighting factor which is computed from two adjacent edges. The proposed edge direction estimation and smoothing method is shown in Fig. 2.

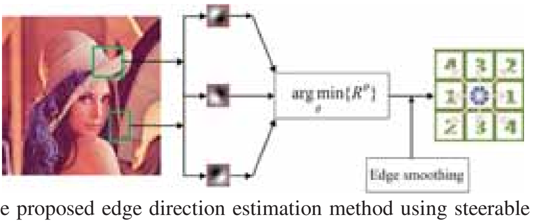


Fig. 2. The proposed edge direction estimation method using steerable filters and weighted smoothing

This research was supported by Basic Science Research Program through National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (2009-0081059), by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2011 (00045420-1), and by the Technology Innovation Program, Industrial Strategic Technology Development Program(Development of SR(Super Resolution) Image Scaler for 4K UHD), under Grant K10041900.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Directionally Adaptive Image Interpolation Based on Bicubic Interpolation

The proposed directionally adaptive image interpolation method is shown in Fig. 3. Given the estimated edge direction θ in the previous subsection, the intensity value of a pixel to be interpolated (the red pixel inside the square shown in Fig. 3) is computed by weighted averaging of P_1 and P_2, which are interpolated using a one–dimension (1D) vertical and horizontal cubic interpolations, respectively.

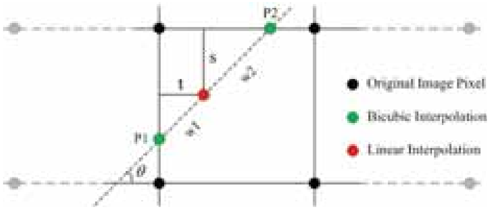


Fig. 3. The proposed directionally adaptive image interpolation algorithm.

C. Image Restoration Using Directionally Adaptive TCLS Filter

The proposed restoration filter is based on the constrained least squares (CLS) filter, and a practical truncation method has been proposed for realizing a finite impulse response (FIR) filtering structure in [4]. In this digest, four directional high pass filters of size 3×3 are used for the constraint function, $c(x, y)$, of the CLS filter. The frequency response of the proposed CLS filter is expressed as

$$R^c_{CLS}(u,v) = \frac{H^{c*}(u,v)}{|H^c(u,v)| + \lambda \, |C^\theta(u,v)|}. \quad (3)$$

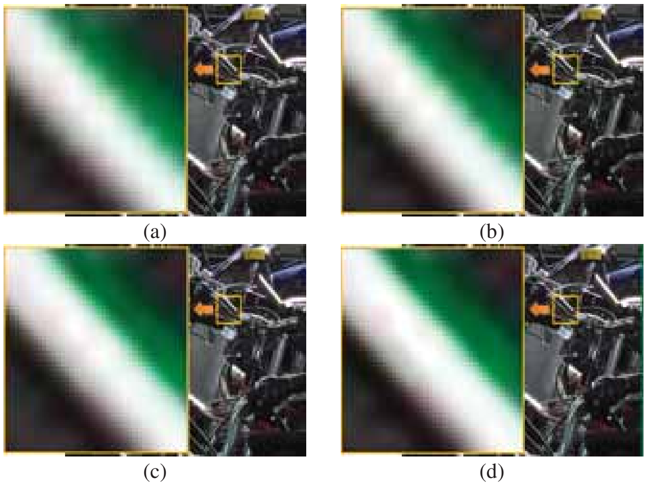
where C^θ represents the directional constraint, the superscript C the color channel, $C \in \{R, G, B\}$, $H^c(u,v)$ the frequency response of the point-spread-function (PSF) determined by the zooming ratio, and $\theta \in \{0°, 45°, 90°, 135°\}$. We truncated the CLS filter using a raised cosine window to obtain the truncated CLS (TCLS) filter. For the implementational issue in a low-cost mobile devices, the size of the FIR restoration filter is confined to 5×5. The degradation artifact caused by the in interpolation process is removed by using the directionally adaptive TCLS filter based on the optimally estimated edge direction in the image.

III. EXPERIMENTAL RESULTS

The proposed method gives acceptable performance in the sense of both suppressing the jagged edges and removing blur compared with the existing state-of-the-art image interpolation method as shown in Fig. 3, Furthermore, peak-to-peak signal-to-noise ratios (PSNR) value and MSSIM are given in Table 1.

IV. CONCLUSIONS

In this digest, we proposed a real-time digital zooming method based on directionally adaptive image interpolation and restoration. The proposed method analyzed the edge direction using computationally efficient steerable filters and weighted smoothing of edge directions. The selective use of bi-cubic and bi-linear interpolations according to the estimated edge direction can enhance image quality with reduced computational load. Various types of degradation caused by interpolation are removed by employing the directionally adaptive TCLS filter. Experimental results show that the proposed method can provide digitally zoomed images without interpolation artifacts such as aliasing and jagged edges. The proposed method can be applied to real-time image processing in the form of built-in hardware because of its FIR filtering structure, and is suitable for the digital zooming functions of various mobile imaging devices.



(a) (b)

(c) (d)

Fig. 3. Experimental results of digital zooming by using four different methods; (a) bilinear interpolation, (b) bicubic interpolation, (c) ICBI [2] interpolation, and (d) the proposed method.

TABLE I
PSNR AND MSSIM COMPARISON OF DIFFERENT METHOD

Image Type (2048X2048)	**Interpolation Type**	**PSNR/ MSSIM**	**Zoom Ratio**		
			X2	**X4**	**X8**
	Bilinear	PSNR	27.7406	21.4036	18.4182
		MSSIM	0.9933	0.9287	0.7984
	Bicubic	PSNR	27.7833	20.9402	17.8920
		MSSIM	0.9935	0.9262	0.7839
	ICBI [2]	PSNR	31.1814	25.5749	22.6127
		MSSIM	0.9942	0.9413	0.9122
	Proposed Method	PSNR	28.2445	23.7232	21.8733
		MSSIM	0.9887	0.9345	0.9098

REFERENCE

[1] X. Li and M. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Processing*, vol. 10, no. 10, pp.1521-1527, October 2001.

[2] A. Giachetti and N. Asuni, "Real-time artifact-free image upscaling," *IEEE Trans. Image Processing*, vol. 20, no. 10, pp. 2760-2768, October 2011.

[3] Freeman, T and Adelson, H, "The design and use of steerable filters," *IEEE TPAMI*, vol. 17, no. 5, pp. 488-499, September 1991.

[4] S. Kim, S. Jun, E. LEE, J. Shin, and J. Paik, "Real-Time Bayer-Domain Image Restoration for an Extended Depth of Field (EDoF) Camera," *IEEE Trans. Consumer Electronics*, vol. 55, no. 4, pp. 1756-1764, November 2009.

Fast Adjustment Method of Spherical Aberration and Focus Offset by Elliptic Equation

Y. Kanatake, T. Matozaki and N. Takeshita
Advanced Technology R&D Center, Mitsubishi Electric Corporation

***Abstract*—Fast adjustment method of spherical aberration and focus offset by elliptic equation was developed. In the case of Blu-ray disc™, spherical aberration is increased because of high NA optical system. In this paper, we introduce our new adjustment method of spherical aberration and focus offset and experimental result of the effectiveness of this method.**

I. INTRODUCTION

In order to realize high-density recording on optical disc, for example, Blu-ray Disc™, it is necessary to reduce an aberration of an objective lens. In order to make the light spot size small, it is necessary to increase a numerical aperture (NA) of an objective lens, but this leads to increase of spherical aberration [1, 2]. The spherical aberration is proportional to the fourth power of NA of the objective lens [3]. The amount of spherical aberration of Blu-ray Disc™ (NA of 0.85) is approximately 6.5 times as much as the one of DVD (NA of 0.6). If spherical aberration increases, a light spot of laser light irradiated onto the information recording layer of the optical disc changes in shape and reproduction performance is deteriorated. The amount of spherical aberration varies depending on a thickness of a layer disposed on the information recording layer of the optical disc. Accordingly, in order to maintain high reproduction performance, it is important to adjust so as to keep the amount of spherical aberration below standard one [4].

Furthermore, the light spot of the laser light irradiated onto the information recording layer of the optical disc also varies in shape depending on performance of focus servo which makes the objective lens follow an optimum position in a direction perpendicular to the information recording layer of the optical disc. An appropriate adjustment of focus offset of the focus servo enables to improve the focus servo performance and to have an appropriate profile of the light spot on the information recording layer of the optical disc. Therefore, it is also necessary to adjust focus offset [4].

II. Proposed Method of Adjustment of Spherical Aberration and Focus Offset

A schematic of the optical head is shown in Fig. 1. During data reproduction, laser beam emitted from the laser diode is onto the optical disc, through the collimator lens, objective lens, and so on. To adjust spherical aberration, collimator lens is adjusted slightly. And to adjust focus offset, objective lens is adjusted slightly.

In general, spherical aberration and focus offset are respectively adjusted so that an amplitude of RF signal is the maximum or jitter value of reproduced signal is the minimum.

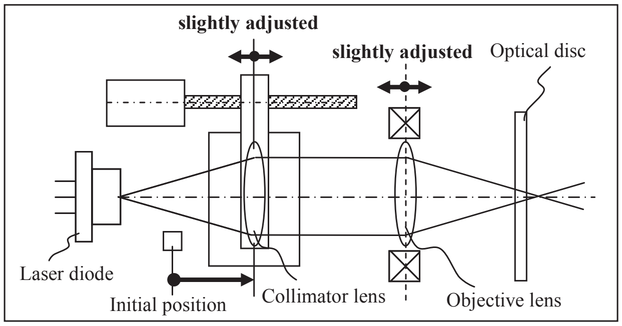


Fig. 1. A schematic of optical head. Collimator lens and objective lens are slightly adjusted.

A schematic of conventional method of adjustment of spherical aberration and focus offset is shown in Fig. 2. Originally, spherical aberration and focus offset are alternately and gradually adjusted so that an amplitude of RF signal is the maximum. In this method, it takes a long time to adjust them.

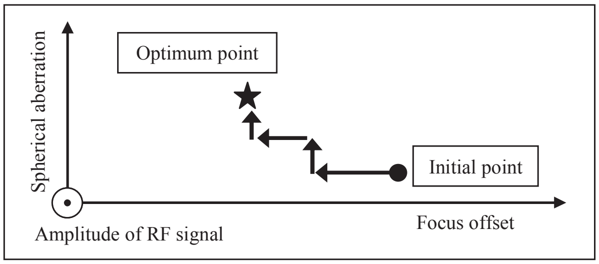


Fig. 2. A schematic of conventional method of adjustment of spherical aberration and focus offset. They are adjusted alternately and gradually so that an amplitude of RF signal is the maximum.

A diagram illustrating an example of a distribution of an amplitude of RF signal in relation to spherical aberration (y-coordinate) and focus offset (x-coordinate) is shown in Fig. 3. Fig. 3 presents a contour map of RF signal and the shape of the map is an ellipse. An optimum value of spherical aberration and focus offset is near a center point of the ellipse.

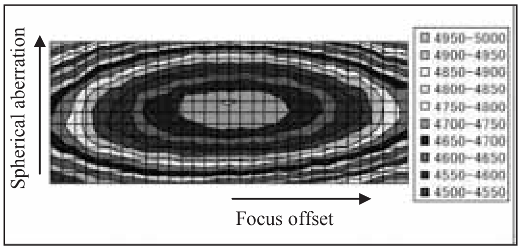


Fig. 3. A diagram illustrating an example of a distribution of an amplitude of RF signal in relation to spherical aberration (y-coordinate) and focus offset (x-coordinate).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

A schematic of proposed method of adjustment of spherical aberration and focus offset is shown in Fig. 4. The points of the same amplitude of RF signal as the amplitude of point D are calculated from point A~C (A: initial point) of an amplitude. The ellipse equation of the same amplitude of RF signal is the following.

$$\left(\frac{(x-x_0)^2}{a^2}+\frac{(y-y_0)^2}{b^2}\right)=1 \qquad (1)$$

Where x_0 is the optimum point of focus offset, y_0 is the optimum point of spherical aberration, and a and b is coefficient. x_0, y_0, a, and b are calculated from the points of the same amplitude of RF signal.

The ellipse slopes depending on the optical pickup models. The equation (1) sometimes changes according to the slope angle.

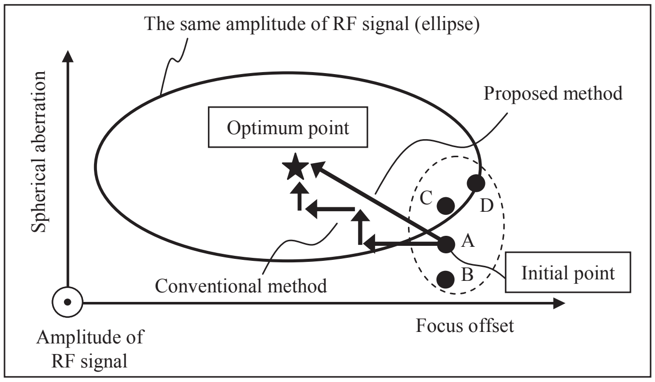


Fig. 4. A schematic of proposed method of adjusting spherical aberration and focus offset. They are adjusted from points A~D (A: initial point) of amplitude of RF signal by the ellipse equation.

III. Performance of Proposed Method

We experimentally confirmed the effectiveness of the proposed method of adjusting a spherical aberration and a focus offset using 45 different types of Blu-ray disc™. Evaluation conditions are shown in Table I.

Histogram of adjustment time of spherical aberration and focus offset by conventional method and proposed one is shown in Fig. 5. Adjustment time is shortened by 2.0 seconds (from 3.0 seconds to 1.0 seconds) compared with conventional method.

Jitter value by conventional method and proposed one is shown in Fig. 6. The difference of Jitter value between them is only 0.1 %. So, we confirmed that the difference of jitter value between conventional method and proposed one is trivial.

TABLE I
Evaluation condition

Items	Contents
Temperature	-20~+75 degrees
Disc type	BD-ROM/R/RE
Number of the layer	Single layer/Dual layer

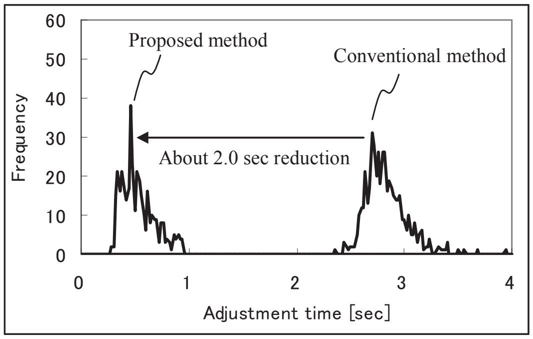


Fig. 5. Histogram of adjustment time of spherical aberration and focus offset by conventional method and proposed one. Adjustment time is shortened by 2.0 second compared with the conventional method.

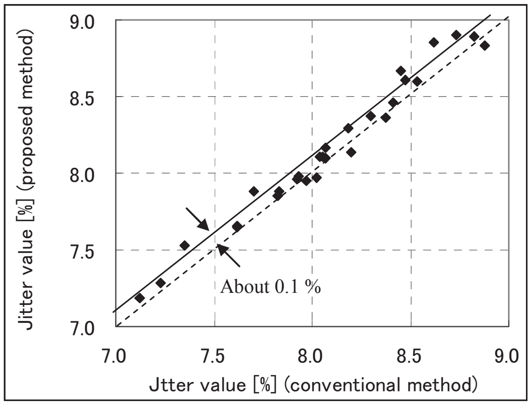


Fig. 6. Jitter value by conventional method and proposed one. The difference of Jitter value between them is only 0.1 %.

IV. Conclusion

New method to adjust spherical aberration and focus offset was proposed. Adjustment time is shortened by 2.0 second compared with conventional method. And the difference of Jitter value between them is only 0.1 %. Effectiveness of the method to adjust spherical aberration and focus offset was experimentally confirmed.

Acknowledgments

The authors sincerely thank our colleagues for their continuous effort and encouragement to complete this investigation.

Examples of Reference Styles

[1] N. Ogata, Y. Kanazawa, M. Horiyama, S. Nishioka, T. Miyake, Y. Nakata and Y. Kurata, "Spherical Aberration Error Detection for Blu-ray Disc Optical Pickups," *Jpn. J. Appl. Phys.*, vol. 45, pp. 5807-5809, 2006.

[2] H. Nakahara, D. Matsubara, T. Matozaki, N. Takeshita and T. Yoshihara, "New Spherical Aberration Compensator for Blu-Ray Disc," *Jpn. J. Appl. Phys.*, vol. 44, pp. 3405-3409, 2005.

[3] M. Itonaga, F. Ito, K. Matsuzaki, S. Chaen, K. Oishi, T. Ueno and A. Nishizawa, "Single Objective Lens Having Numerical Aperture of 0.85 for a High Density Optical Disk System," *Jpn. J. Appl. Phys.*, vol. 41, pp. 1798-1803, 2002.

[4] M. Moriya, T. Takizawa, H. Ishibashi, K. Watanabe and Y. Hino, "Blu-ray Disc Drive Development, " *Matsushita Technical Journal.*, vol. 50, No. 5, 2004.

How Costly are Secure Transactions on Handheld Devices?

Florina Almenares, Patricia Arias, Andrés Marín López, Daniel Díaz-Sánchez, Rosa Sánchez
Telematic Engineering Department, University Carlos III of Madrid, Leganés, Madrid (Spain)

Abstract— **Handheld devices are more and more powerful allowing to do most things people do on a desktop. Nevertheless, mobile device security follows being an open issue. We have performed the first study of the security support between native and OpenSSL-based libraries, in terms of energy consumption and time, about secure communication performance.**

I. Introduction

Handheld devices are more and more powerful allowing to do most things people do on a desktop. Due to the mobility, these devices are more and more interconnected, participating in diverse networks. Nevertheless, mobile device security follows being an open issue, because they are exposed to suffer same attacks than on desktop with a less security support and new challenges provided by the environment. They have been simply considered as HTTPS clients, but other applications requiring security support do not have it; for example, e-mails. For this reason, we have exhaustively studied the cost of using secure communications with X.509 certificates, that is, TLS/SSL. This protocol is used to protect the network activity; therefore, it requires underlying cryptographic algorithms. The main objectives of such study were to determine: 1) security level provided by symmetric, asymmetric, and hash algorithms that are supported on handheld devices, 2) efficiency of each one, including ciphering and signing. We have compared native and OpenSSL-based libraries to calculate the overhead added by this to the applications.

In order to fulfill with such objectives, we have evaluated the cost in terms of time and energy consumption, because energy is another very important factor for handheld devices, apart from response time. Other studies have been oriented to analyze only the performance of cryptographic algorithms [1] and security protocols [2]. [1] is a recent work, but they do not include the communication cost. They present the amount of overhead that security algorithms can introduce in a system. Authors in [2] use only the OpenSSL library. [3] analyzes the performance regarding energy consumption over different networks (i.e. 3G, GSM, and Wi-Fi), with the aim of optimizing the use of the network interfaces and battery saving. So, a study including different ciphers, data sizes and libraries is not available. This paper is only focused on the performance of SSL connections according to the supported ciphers by the browsers. The rest of this article is organized as follows. In section II, we describe the full study performed. Then, section III details the results obtained from secure transactions tests according to different ciphers, data sizes, and browsers. Finally, we briefly give some discussion and conclusions in section IV.

II. Energy Consumption and Time Analysis Description

We have performed an experimental test set, using a smart phone with a processor 264MHz ARM9, and battery capacity 950mAh, and a handheld device with a processor 330MHz ARM11, and battery capacity 1500mAh. Firstly, we start testing cryptographic algorithms (i.e. encryption and signing), and certificate management, because these are the underlying support for secure transactions. Then, we have tested secure communications through download of pages from an Apache server, using different data sizes and ciphers. We have used two different suites of algorithms and protocols: native algorithms implemented by proprietary operating systems, and OpenSSL-based library, because the trend in operative systems for handheld devices is to use systems based on Linux kernel.

In the secure communication test, the handheld device acts as a HTTPS client. We have used two browsers: a) the browser installed by default, and b) a commercial browser. We have firstly evaluated the download of a page from the server, using different ciphers for establishment of the secure communication. Then, we tested the set of ciphers supported by the server in order analyze the efficiency and overhead imposed by different data sizes (i.e. 1KB, 10KB, 100KB, and 1MB), compared with transmission without encoding.

The device and server were connected through a Wi-Fi network, which was created only for them. So, the measures obtained are free of external noise. We have obtained both time and software-based energy consumption measures such as is explained in the following section.

III. Cost of Secure Communications

Fig.1. and Fig. 2 represent the consumption in terms of energy and time, respectively, for the establishment of a SSL connection according the testbed explained previously. The measurements have been carried out for each of the *cipher suites* available in the native WebKit-based browser installed in the smart phone, and varying the size of the web page. It is to say that from the 28 *cipher suites* available in the openSSLv0.98g used for setting up the secure server, only a subset of 13 *cipher suites* were supported by the native browser and 9 *cipher suites* by the non-native Presto-based browser in the best case.

The measurements in the native browser are only showed, because lineal upward trend is the same in both cases; with the difference that in the non-native browser, the execution time is 30% faster. On the contrary, Presto-based browser does not support any of the *cipher suites* recommended by NIST for maximum security: RSA/DSA authentication with ephemeral Diffie-Hellman key agreement (e.g. DHE-DSS-AES256-CBC-SHA). It does not support AES for some operating systems; it only supports more insecure ciphers such as DES and RC4.

This work was partially supported by the Spanish Ministry of Science and Innovation under the project CONSEQUENCE (TEC2010-20572-C02-01), and by the State of Madrid (Spain) under the contract number S2009/TIC-1650 (e-Madrid).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

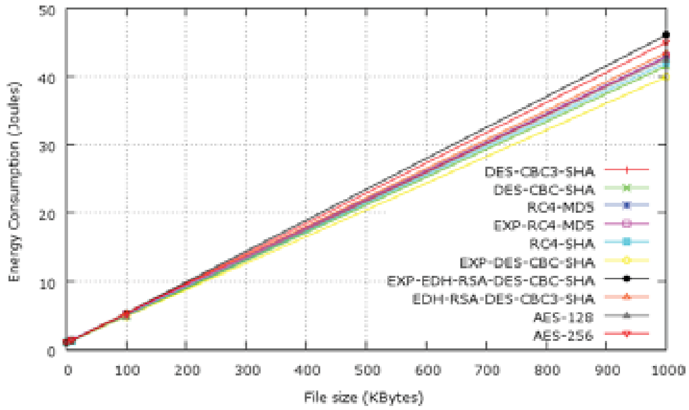


Fig.1. Energy consumption in establishing a SSL connection to transfer a web page between the phone's native browser and a secure server.

We observed that both magnitudes increase linearly with the size of the transferred page. The variations between the different *cipher suites* are more significant for bigger file sizes, especially in regard to energy consumption.

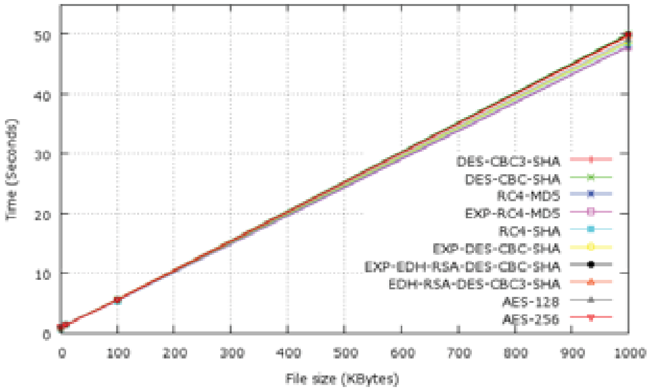


Fig.2. Time consumption in establishing a SSL connection to transfer a web page between the native browser and a secure server.

To better analyze the correlation between time and energy consumption (i.e. power), the graph in Fig. 3 shows the evolution of energy regarding time. In this case, size of the web page was the greatest one, 1MB. In this graph, we can conclude that algorithms that consume more energy are not the ones that consume more time. Likewise, we can observe that despite the difference in the consumption seems irrelevant, indeed for longer transactions time the battery energy saving is 12%, which is a relevant number.

Extracted from the graph in Fig. 1, Table I summarizes the energy consumption overhead for the best and worst *cipher suites* (i.e. the cheapest and the more expensive from the point of view of energy consumption versus time ratio) with respect to the case where no security is applied. Measures correspond to the 1MB file size. According to this table, the EXP-EDH-RSA-DES-CBC-SHA is the most energy consuming algorithm; and EXP-DES-CBC-SHA is the less consuming *cipher suite*. However, these *cipher suites* are not the slower and faster, as shown in Fig. 2, but their time performance is very similar. Thus, there is not a direct general correlation between the performance in time of an algorithm and the associated energy consumption, i.e. the slowest *cipher suite* is not necessary the most expensive in terms of energy.

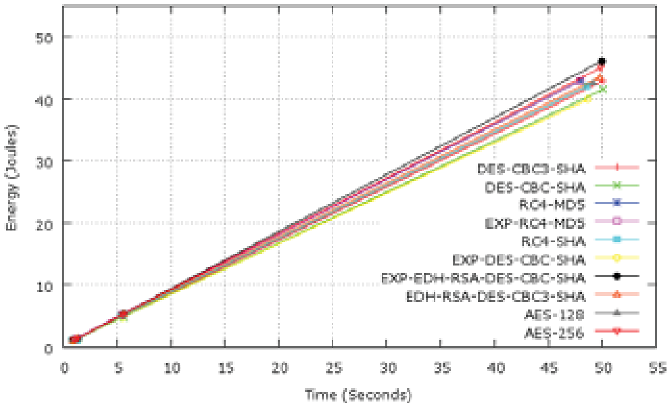


Fig. 3. Energy consumption versus time consumption in establishing a SSL connection to transfer 1MB.

The energy cost depends on the operations underlying the particular *cipher suite* and thus evolves differently with time.

TABLE I
ENERGY OVERHEAD FOR THE BEST AND WORST CIPHER SUITES

Cipher suite	Energy Consumption	Overhead
No cipher suite	37,22059095 J	0%
EXP-DES-CBC-SHA	39,97908126 J	7,41%
EXP-EDH-RSA-DES-CBC-SHA	46,11309850 J	23,89%

IV. DISCUSSION AND CONCLUSIONS

According to the results showed, we can conclude that OpenSSL-based or/and native WebKit-based browsers offer a more high level of security to applications using HTTPS, because these support the *cipher suites* recommended for maximum security, as well as having a greater support of algorithms. Besides, these do not send all the user's certificates stored in the device when client-side authentication is required.

Regarding overhead added by the security algorithms, this reaches about 24% in the worst case. Using the processor ARM11 330MHz is could become lesser, 17%.

Finally, it is worth to mention that the result from tests performed with the same ciphers directly through desktop applications with different libraries showed that native library is better for greater data sizes, but the performance obtained from OpenSSL-based functions is better for smaller file sizes.

REFERENCES

[1] H. Rifà-Pous, and J. Herrera-Joancomartí, "Computational and Energy Costs of Cryptographic Algorithms on Handheld Devices", *Future Internet*, vol. 3, pp. 31-48, Feb. 2011.

[2] N.R. Potlapally, S. Ravi, A. Raghunathan and N.K. Jha, "A Study of the Energy Consumption Characteristics of Cryptographic Algorithms and Security Protocols," *IEEE Transactions on Mobile Computing.*, vol. 5, no. 2, pp. 128-143, Feb. 2006.

[3] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications", in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference* (IMC '09). ACM, New York, NY, USA, 280-293. Nov. 2009.

Transparent Fast Resynchronization for Consumer RAID
Digest of Technical Papers

Sung Hoon Baek, *Member, IEEE*
Dept. of Computer System Engineering, Jungwon University, Korea

Abstract–**Consumer RAID without a power-fail-safe component suffers from the problem of time-consuming, scan-based, resynchronization after a sudden power-off. It is very inconvenient for consumers to wait for that the very long resynchronization completes as a penalty for a careless power-off. This paper presents a fast resynchronization with negligible overhead for consumer RAID. The proposed scheme that was implemented as a software RAID driver in Linux shortens 200 minutes of the resynchronization process to 5 seconds with 2% of overhead in an experiment**.

I. Introduction

Small Office and Home Office (SOHO) or consumer RAID systems are aimed to reliably store user's precious data and serves as a network attached storage with several (2~10) disks in a home or office intranet.

High-end RAID systems utilize uninterruptible power supply (UPS) or battery-backed RAM to achieve both reliability and performance. Such power-fail-safe devices are not applicable to consumer RAID at an affordable cost.

SOHO RAID systems without expensive UPS suffer from several hours of parity resynchronization process after a sudden power-off. Consumers may frequently cause ungraceful power-off like home appliances by turning off the switch of a power strip. The system must scan all data for parity resynchronization after the unexpected power-off, which may generate inconsistent stripes where the parity block is inconsistent with the data blocks.

When a write is issued to a RAID-5 array, two or more disks for parity and data must be updated in a consistent manner. If a sudden power-off occurs after the data are written but before the parity is updated, the stripe is left in an inconsistent state.

Because the system does not know which stripe is in an inconsistent state after a crash, it must scan the entire volume to search for inconsistent stripes that are not recoverable when a disk fails in the future [1]. Scanning all stripes takes several hours and makes consumers inconvenient to use the storage system.

To significantly reduce the expensive resynchronization time sacrificing performance, Multi-device (MD) of Linux kernel provides an intent bitmap scheme [2], which sets and records the bit of the location of a pending block before it is issued to a disk, and it clears and records the bit after the block is written to the disk.

An alternate approach eliminates the time-exhaustive resynchronization with a small overhead by extending the journaling scheme of file systems [3], but it needs to modify file systems and other file system tools such as *fsck*. This approach involves development cost for each file system and each version of file systems.

The proposed scheme of this paper is transparent to (independent of) file systems while removing expensive resynchronization process with a small overhead.

II. Intent Bulk Log-Based Resynchronization

The proposed scheme utilizes a write-ahead log that lists the locations of aggregated multiple blocks before they are written. The log gives hints for the stripes that were being written when the power went off.

The proposed scheme supports a write cache without UPS or battery-backed RAM because of employing "write barrier" [4], which is a kernel mechanism used to ensure that file system metadata are correctly written and ordered on persistent storage. Write barrier in journaling file systems such as EXT3, EXT4, NTFS, XFS, and BTRFS ensures that the file systems are consistent even when storage devices with volatile write caches lose power.

Fig. 1 shows how to process the write-ahead log that is called intent bulk log in this paper. The proposed scheme employs two kinds of write cache, which are called Write Cache 1 (WC1) and Write Cache 2 (WC2). At first, let WC1 be Buffering Write Cache (BWC) and WC2 be Destaging Write Cache (DWC). (Step 1) All blocks that are requested to write them are buffered to BWC. (Step2) If DWC becomes empty by destaging all its buffered data and BWC is not empty, BWC is swapped with empty DWC. (Step 3) A list containing all stripe numbers buffered in DWC is logged in a designated location. The list is called intent bulk log. (Step 4) Data buffered only in DWC can be issued to disks only after

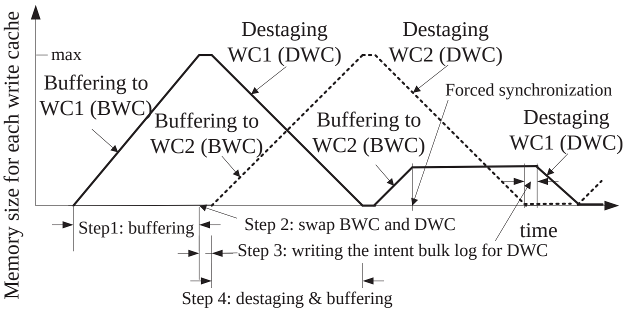


Fig. 1. The size of WC1 and the size of WC2 are shown for each step. The solid line and dot line indicate WC1 and WC2, respectively.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

the logging completes. After all blocks in DWC are evicted to disks, we repeat Step2 to Step 4.

Only DWC can destage its blocks to disks because the block addresses contained only in DWC were recorded in the last intent bulk log. None of the blocks in BWC can move to DWC because blocks that are not described by the intent bulk log cannot be updated to disks. Blocks in DWC, however, can move to BWC to apply a destaging scheme such as least recently written (LRW) policy when cache hits occur for the blocks.

BWC and DWC share the memory resource. The size of BWC can increase if the size of DWC decreases, and vice versa. Memory that is evicted from DWC by destaging can be allocated to BWC.

In the booting stage, if the system detects an unclean shutdown, it searches for the latest intent bulk log and rebuilds the parity only for the stripes that are recorded in the latest log. The number of targeted stripes for resynchronization is bounded within the write cache size. Hence, the synchronization process can complete within limited time.

The intent bulk log requires an additional write access for thousands of blocks, but the intent bitmap causes two writes to set and clear the bit for every data in the worst case. The intent bulk log and the intent bitmap dramatically remove the time-consuming resynchronization process, but for normal I/Os, the intent bulk log employs much less overhead than the intent bitmap method.

III. Experimental Results

The *intent bulk log* scheme was implemented in a software RAID driver, Layer Of RAID Engine (LORE) [5], in a 64-bit Linux kernel 2.6.35. The storage system in this experiment is a RAID-5 array that runs LORE with a 3.2GHz i5 processor, 2GB of main memory, and five 1TB 7200rpm SATA3 hard disk drives.

Multi-device (MD) is a software RAID driver included in Linux and employs the *intent bitmap* scheme. We compared the overheads of the intent bitmap and the intent bulk log using MD and LORE.

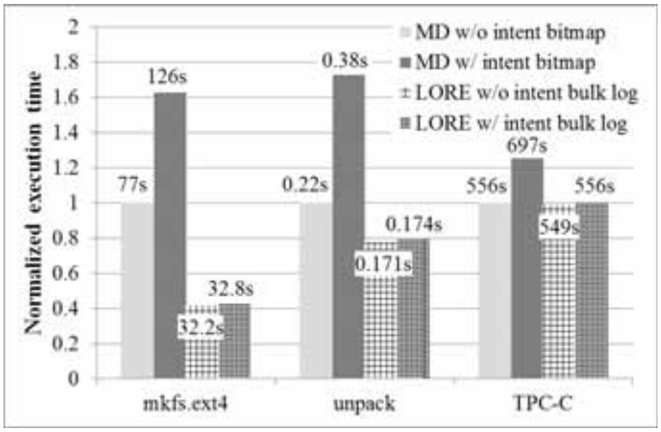


Fig. 2. It compares the normalized execution time of building an ext4 file system (mkfs.ext4), unpacking an openssh package (unpack), and replaying a TPC-C I/O trace (TPC-C) as compared to 'MD without the intent bitmap'.

TABLE I
Resynchronization Time

	Entire scan	Intent bitmap	Intent bulk log
Resync. time	200 minutes	< 5 seconds	< 5 seconds

Fig. 2 compares the normalized execution time of building an ext4 file system (mkfs.ext4), unpacking an openssh package (unpack), and replaying a TPC-C I/O trace (TPC-C) as compared to MD that scans the entire volume for resynchronization.

Adding the intent bitmap to MD employs 62%, 72%, and 25% degradation for mkfs.ext4, unpack, and TPC-C, respectively. Whereas, the intent bulk log scheme exhibits 2%, 2%, and 1% overhead, respectively. The intent bulk log is significantly superior to the intent-bitmap.

TABLE I compares the resynchronization time to fix inconsistent stripes after an unclean shutdown between the entire scan and the intent bulk log. The conventional method requires 200 minutes to resynchronize the 4TB volume. It is very inconvenient for consumers to wait for that the long resynchronization completes. The proposed scheme, intent bulk log with dual write caches shortens 200 minutes of the resynchronization process to 5 seconds with about 2% of overhead.

IV. Conclusion

Unlike high-end RAID system that uses UPS or battery-backed RAM to protect buffered data. The proposed scheme targets to a consumer RAID system without a power-fail-safe component so that the system is affordable to consumers. Low cost RAID systems must include a solution for inconsistent stripes after a sudden power-off.

Scanning the entire volume for the solution exhibits good performance but it forces users to wait for several hours as a penalty for a careless shutdown. The intent log employs short resynchronization time but significantly sacrifices the performance of the normal I/O. The proposed scheme, intent bulk log with dual write caches includes negligible performance degradation and the very short resynchronization process. The proposed scheme provides great advantage against traditional solution for consumer RAID.

V. References

[1] D. Teigland and H. Mauelshagen, "Volume Managers in Linux", *In Proc. of the USENIX Annual Technical Conference*, June 2002.

[2] P. Clements and J. Bottomley, "High Availability Data Replication", *In Proc. of the 2003 Linux Symposium*, June 2003.

[3] T. E. Denehy, A. C. Arpaci-Dusseau, and R. H. Arpaci-Dusseau, "Journal-guided Resynchronization for Software RAID", In Proc. of the 4th USENIX Conf. on File and Storage Technologies", Dec. 2005.

[4] J. Bacik, K. Dudka, H. Goede, D. Ledford, D. Novotny, N. Straz, ... , "Write Barriers", *in Red Hat Enterprise Linux 6 Storage Administration Guide*, pp149-151, 2011

[5] S.H. Baek and K.H. Park, "Striping-aware sequential prefetching for independency and parallelism in disk array with concurrent accesses", *IEEE Transactions on Computers*, Vol.58, No.8, Aug 2009, pp1146-1152.

Fast Coding Unit Decision Algorithm Based on Inter and Intra Prediction Unit Termination for HEVC

Hyang-Mi Yoo and Jae-Won Suh, *Member, IEEE*
Chungbuk National University, Cheong-ju, Korea

Abstract—**To provide highly efficient video coding standard, the new high efficiency video coding (HEVC) standard has adopted a recursive quad-tree structured coding unit (CU) and a transform unit (TU). Although the HEVC obtains a very high coding efficiency but it considerably increases the computational complexity because the encoder tests every possible CU in order to estimate the coding performance of each CU. In this paper, we propose a fast CU decision algorithm based on Inter and Intra PU termination for the HEVC to reduce the computational complexity. The proposed algorithm checks code block flag (CBF) value and rate distortion (RD) cost for inter PU prediction. If these two values are satisfied with our present conditions, the next PU process is terminated in the current CU. Experiment result shows that the proposed algorithm reduces the encoding time average 49.56% and 37.3% according to the weighted condition of our proposed algorithm.**

I. INTRODUCTION

Currently, the content growth of HD and the HD broadcasting service are offering users to enjoy high quality and high resolution. In the near future, the content growth of UHD and the UHD broadcasting service also will be offered to users following the demands to needs for higher quality and higher resolution. For these high quality and high resolution video, the HEVC standard, the next generation of video coding standard, is developing on the JCT-VC team. This HEVC is aimed to have two times higher coding efficiency than H.264/AVC, and now HEVC has the 40~45% data compression ratio when compared to the H.264/AVC using HEVC Test Model (HM) 4.0 [1].

In the HEVC, the input video is encoded by recursive quad-tree structured CU. This complex quad-tree structure makes HEVC coding more efficient, but it also makes the HEVC have several times higher complexity and encoding time than the H.264/AVC. In order to reduce this computational complexity and the encoding time in HEVC, several fast mode decision algorithms were suggested. Choi [2] proposed a fast coding unit decision method. It states that if SKIP mode is determined as the best mode in current CU depth, then the rest of sub-tree CU prediction processing in all the next lower CU depths is skipped. Gweon [3] also proposed early termination of CU encoding to reduce complexity in HEVC. If all CBF values for luma and two chromas are zero after the prediction unit (PU) process, then the next PU processes in current CU depth are skipped. Then move to the next CU depth's prediction process.

II. MODE DECISION FOR HEVC

Mode decision of the HEVC is started with a unit of largest CU (LCU). If the maximum CU size is set as 64, then the block in the Figure 1 (a) is represented as the LCU and this CU's depth is 0. CU depth is increasing by splitting the one CU into four CUs by quad-tree structured CU system. As a result, the CU prediction process is performed by the number of blocks in Fig. 1 until CU could no longer be split down into smaller CUs. Whenever the four CU predictions are finished in the same CU depth, the sum of these four CU cost values are compared to the corresponding one CU cost value of one step upper CU depth. Then, the smaller value of these two cost values is chosen as the best mode cost.

One CU is composed of the various PU processes, such as Merge 2N×2N, Inter 2N×2N, Inter N×2N, Inter 2N×N, Inter 2N×nU, Inter 2N×nD, Inter nL×2N, Inter nR×2N, Intra 2N×2N, Intra N×N. Intra N×N prediction is performed only when current CU size N is 8. RD cost for each PU is compared to decide the best mode within the current CU.

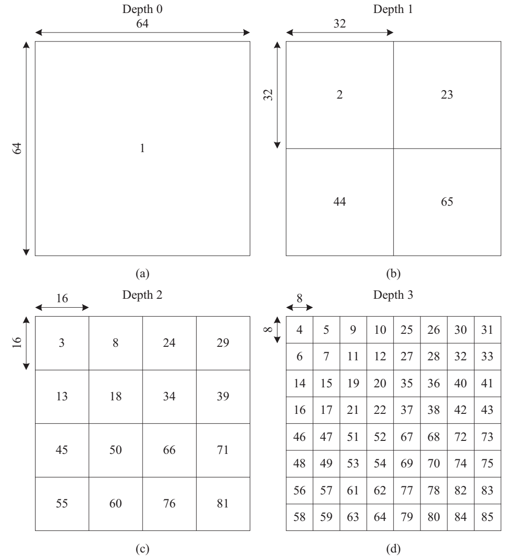


Figure 1. Inter CU prediction order with one LCU (64×64)

III. PROPOSED ALGORITHM

In this paper, we suggest a fast CU decision algorithm based on Inter and Intra PU termination. The whole flow chart is shown in Fig. 2.

First, we utilize the CBF and RD cost for Inter PU termination. We do not use the asymmetry partition for Inter PU, such as Inter 2N×nU, Inter 2N×nD, Inter nL×2N, Inter nR×2N. Whenever a CU is predicted in an Inter PU, the CBFs and RD costs for Inter 2N×2N, Inter 2N×N, Inter N×2N are

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0011097).

examined. If CBF of the Inter PU is zero for one luminance and two chrominances and RD cost of the Inter PU is lower than the weighed moving averaged SKIP mode RD costs as in (1) and (2), the next PU process is terminated in the current CU.

$$Th_{\text{avg SKIP mode cost}} = \frac{\Sigma_{\text{n}}^{\text{n - 4}}\text{RD Cost}_{\text{skip mode}}}{5} \times \alpha \qquad (1)$$

$$RD\ Cost_{\text{Cur Inter mode}} < Th_{\text{avg SKIP mode cost}} \qquad (2)$$

In (1), $Th_{avg\ SKIP\ mode\ cost}$ is independently calculated according the CU size. This value is updated after finishing all PU processes are done in current CU depth.

In order to illustrate our algorithm, we use the predefined block number as shown in Fig. 1. If Inter 2N×2N PU of the first block "2" in the Depth 1 satisfy the proposed two conditions, then we can skip the Inter 2N×N, Inter N×2N and Intra prediction within the "2" block. In addition, we can also skip the remaining further splitted CUs, such as "3", "8", "13", "18" in Depth 2 and "4-7", "9-12", ″14-17", and "19-22" in Depth 3.

Next, we use the Inter PU information to skip Intra PU. During the Inter 2N×2N, Inter N×2N, and Inter 2N×N, we can obtain the four N×N CBF values for each Inter PU. We call it sub CBF. In order to determine whether to skip Intra PU or not, these sub CBFs are inclusive OR operated by according to the spatial position, which is expressed

$$C_b(i,j) = \begin{cases} 0 & if\, P_{2N\times 2Nb}(i,j) = 0 \quad OR \quad P_{2N\times Nb}(i,j) = 0 \quad OR \quad P_{N\times 2Nb}(i,j) = 0 \\ 1 & otherwise \end{cases} \qquad (3)$$

$$number\ of\ zero\ in\ C_b(\text{i, j}) \geq 3 \qquad (4)$$

$P_{2N\times 2Nb}$, $P_{2N\times Nb}$, and $P_{N\times 2Nb}$ represent the 2×2 data consisted of 4 sub CBF values which is obtained by Inter 2N×2N prediction, Inter 2N×N, and Inter N×2N, respectively. $C_b(i,j)$ represents the combined sub CBF map. If the number of zero in $C_b(i,j)$ is three or more, then we skip the Intra PU.

IV. Simulation Results and Conclusions

In order to evaluate the performance of our proposed algorithm, we use HM5.0 reference software and compared our experiment results with Choi [2] and Gweon [3]'s experiment results which are already implemented in HM 5.0. We performed computer simulations on various test video sequences, Class A-D (Calss A: 5 seconds, Class B-D: 10 seconds) recommended by JVT with QP 22, 27, 32, 37. We simulated the following configuration; we used the random access configuration file which we received when we downloaded the HM 5.0. We did not use the asymmetry partition by modifying this part in the configuration file.

We summarized the experiment results in Table I. We found out that in the 1.5 value of α, the average bit increment was 0.33% compared with HM5.0 reference software and average PSNR loss was -0.07dB. The proposed algorithm reduced the total average encoding time by -49.56%. In the 1.0 value of α, the average bit increment was -0.08%, but average PSNR loss was -0.02dB and the total average encoding time is reduced by -37.3%. If α value is changed, then we can control the PSNR, bit rate and encoding time in different situations. These results prove that this proposed algorithm is a great solution as a fast mode decision algorithm.

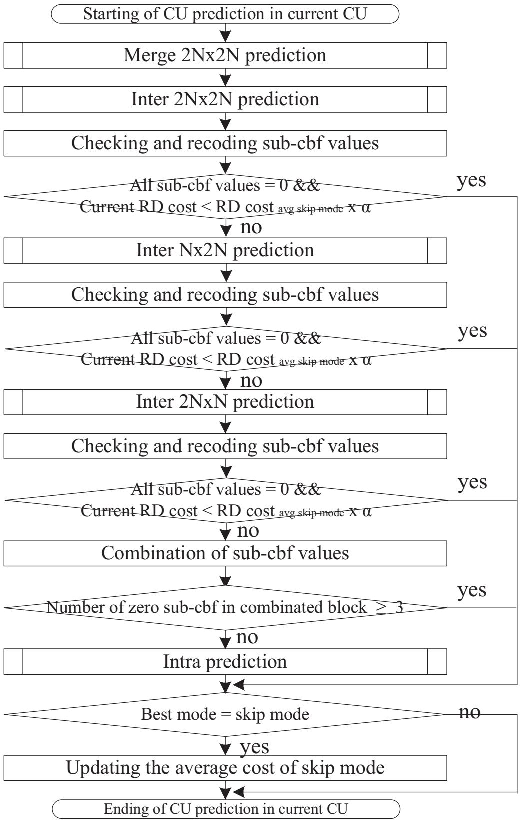

Starting of CU prediction in current CU
Merge 2Nx2N prediction
Inter 2Nx2N prediction
Checking and recoding sub-cbf values
All sub-cbf values = 0 &&
Current RD cost < RD cost avg skip mode x α
yes
no
Inter Nx2N prediction
Checking and recoding sub-cbf values
All sub-cbf values = 0 &&
Current RD cost < RD cost avg skip mode x α
yes
no
Inter 2NxN prediction
Checking and recoding sub-cbf values
All sub-cbf values = 0 &&
Current RD cost < RD cost avg skip mode x α
yes
no
Combination of sub-cbf values
Number of zero sub-cbf in combinated block ≥ 3
yes
no
Intra prediction
Best mode = skip mode
no
yes
Updating the average cost of skip mode
Ending of CU prediction in current CU

Figure 2. Flow chart of the proposed algorithm

Table I. The experiment results

	Choi [2]			Gweon [3]		
	ΔPSNR	ΔBits	ΔTime	ΔPSNR	ΔBits	ΔTime
Class A	-0.06	-0.92	-43.25	-0.05	-0.42	-44.19
Class B	-0.03	-0.83	-45.93	-0.02	-0.48	-42.19
Class C	-0.04	-0.71	-33.21	-0.05	-0.52	-35.48
Class D	-0.04	-0.69	-30.86	-0.07	-0.72	-34.80
Average	-0.04	-0.79	-38.31	-0.05	-0.53	-39.17
	Proposed with α (1.0)			Proposed with α (1.5)		
	ΔPSNR	ΔBits	ΔTime	ΔPSNR	ΔBits	ΔTime
Class A	-0.02	0.21	-42.48	-0.08	-0.15	-52.83
Class B	-0.01	-0.04	-43.22	-0.05	-0.55	-57.45
Class C	-0.02	0.15	-33.57	-0.06	-0.16	-45.28
Class D	-0.02	0.006	-29.93	-0.09	-0.47	-42.66
Average	-0.02	0.08	-37.30	-0.07	-0.33	-49.56

References

[1] Bin Li, Sullivan, Gary J. Sullivan and Jizheng Xu "Comparison of Compression Performance of HEVC Working Draft 4 with AVC High Profile," JCT-VC document, JCTVC-G399, November, 2011

[2] Kiho Choi, Sang-Hyo Park and Euee S. Jang "Coding tree pruning based CU early termination," JCT-VC document, JCTVC-F092, July, 2011

[3] Ryeong Hee Gweon ,Yung-Lyul Lee, Jeongyeon Lim "Early Termination of CU Encoding to Reduce HEVC Complexity," JCTVC-F045, July, 2011

Multidimensional Workload-Performance Analysis on the NAND Flash-based SSD Array Systems

Brian Myungjune JUNG, Jupyung LEE, Boncheol GU, Jungmin SEO, Hyun-Jung SHIN, and Eunsoo SHIM
SAIT, Samsung Electronics, Korea

***Abstract*-- We performed the workload-performance analysis on the NAND flash-based SSD array systems, especially regarding the parallel I/O workload. Lessons learned from this study are: (a) SSD array system has sweet spots, the workload patterns which can maximize the overall system performance. This sweet spot approach is possible due to the distinguishing performance characteristics of each member SSD. We showed that the SSD array performance can be fairly improved by transforming I/O workload close to the sweet spot workload pattern. (b) It is possible to use mathematical optimization to find optimal workload patterns when the performance characteristics and the system-wide workload conditions are given. This is very helpful method to decide which workload patterns are the best, or the next best choices in current situation.**

I. Introduction

NAND flash-based distributed SSD (solid-state drive) array systems are getting more attention than ever with the increasing demands on the faster storage systems to handle larger data at higher speed. But the actual situation is not that simple because SSD array facing real I/O workload does not always show expected performance. From where did things go wrong? Due to the changeable and complicated performance characteristics of the SSD? What about the heterogeneity of SSD internal hardware and software design come from different manufacturers? To answer these questions properly, let us start discussion from the single SSD's performance issues and then extend our discussion to the distributed SSD array system, because the overall performance of SSD array is closely dependent on each member SSD's performance.

II. Performance Degradation Problem of the SSD

Many studies revealed that performance degradation of the SSD is getting worse as the random write to the SSD continues [1][2]. This random write makes the SSD get experienced more 'write amplification' which reduces its life and degrades the performance. The things would go easy if the amount of disadvantageous I/O workload pattern such as random writes could be decreased, but reality is distinct from our hope. On the contrary, current IT trends such as multi-core, multi-tenant cloud computing, and virtual desktop infrastructure (VDI) makes things harder. The increase of the number of processor cores makes parallel I/O dominant, which consequently generates more random I/O. Wider adoption of VDI in large companies and multi-tenant cloud computing make things harder because it brings about intensive I/O blending which also makes I/O randomized [3].

III. Multidimensional Workload Analysis

Although the existing approaches to handle random write works well to some extent, there still remains the room to improve the performance by more detailed and comprehensive analysis and manipulation on the I/O workload. We decided to use uFLIP benchmark suite [4] as a basis for our experiments and analysis because it provides nine different types of micro-benchmarks defined over various I/O workload patterns including, but not limited to locality effect test, pause effect test, parallelism effect test.

Experiments are designed so that performance impact by various aspects of I/O workload such as location, timing, and concurrency can be tested for a given SSD. We also consider the several models of real SSD products come from different manufacturers to examine and compare the performance characteristics of different SSDs (SSD product 1 [5] and SSD product 2 [6]) against the same type of workload pattern. We performed all the tests with random write based workloads. Due to the space limitation, we chose two benchmark results to discuss, among nine different workload-performance benchmarks results.

Fig. 1 shows the different performance characteristics of SSD product 1 and SSD product 2 in case of pause-time workload test. In the SSD product 1 case, 100us pause between I/Os lowers latency from 386us to 217us. Considering the cost of pause time by 100us, this actually means 121.8% improvement in IOPS. In the SSD product 2 case, 100us pause between I/Os lowers latency from 5,996us to 5,694us, which actually means just merely 103.6% improvement in IOPS considering the pause time cost. This result shows the amount of performance improvement by pause time change is different according to the SSD product model.

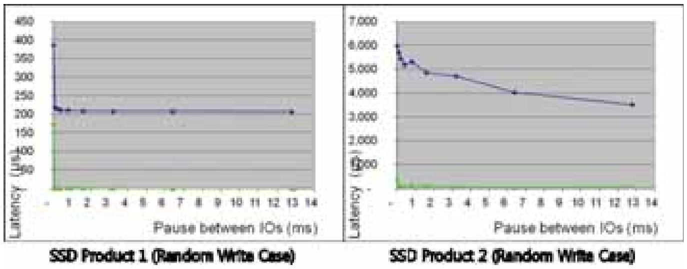


Fig. 1. Performance characteristics of SSD product 1 and SSD product 2 for pause-time workload test. There is almost no improvement of performance in SSD product 2 case, while there is a little improvement in SSD product 1 case.

Fig. 2 shows fairly different performance characteristics between SSD product 1 and SSD product 2 in case of parallel workload test. SSD product 1 shows fairly good performance up to 4 parallel I/Os in terms of low latency value and performance sustainability. But when the number of parallel I/Os exceeds over 4, the performance drops sharply (latency increases rapidly). But SSD product 2 has bad performance

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

characteristics across all test ranges. Its latency value at 1 parallel I/O is 6,178us, and latency value at 16 parallel I/Os is 6,624us. But one remarkable characteristics of SSD product 2 is that there is almost no change although the number of parallel I/O varies from 1 to 16.

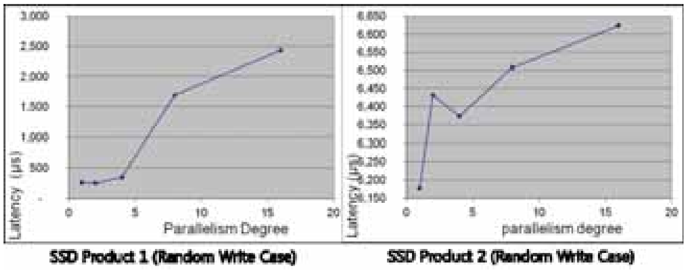


Fig. 2. Performance characteristics of SSD product 1 and SSD product 2 for parallel workload test. SSD Product 1 shows fairly good performance until 4 parallel I/Os. But SSD product 2 shows bad performance characteristics across all test ranges.

IV. Performance Improvement by Exploiting Heterogeneous Performance Characteristics

We start this section with the I/O traffic handling method in storage systems comprised of multiple storage devices. Let's assume that there is a distributed storage system comprised of N multiple storage devices connected over the storage networking protocol such as iSCSI. It is not easy to guarantee that all storage devices remain in the same model through the life of that storage system. What traffic distribution policy should be considered if the distributed storage array is comprised of more than single types of SSDs? What makes the situation more complex is the caprice of the SSD performance, which is usually affected by the workload history and the current workload being delivered to the SSD [7][8]. This is the reason why we need adaptive I/O traffic handling and identification of the optimal I/O workload patterns, to get the maximum I/O performance of the SSD array system.

In terms of the parallel I/O workloads, is it optimal to evenly distribute the parallel I/O traffic to each SSD? Should that still be the only way to handle the I/O traffic to get maximized overall performance of storage array system even in SSD case? For ease of discussion, assume that there are total 24 persistent parallel I/O streams in the target system, and we have an SSD array system comprised of three different SSD models, of which performance characteristics are presented in the Table 1 and Fig. 3.

Table 1. Exemplified Latency Characteristics of Three SSDs

# of Parallel I/O Streams	SSD A	SSD B	SSD C
1	290	750	6,000
2	290	800	6,100
4	300	1,000	6,200
8	3,000	2,000	6,300
16	4,000	2,500	6,400

Latency are represented in microseconds. Actual unit of each cell value is microseconds per unit I/O operation, i.e., us/io.

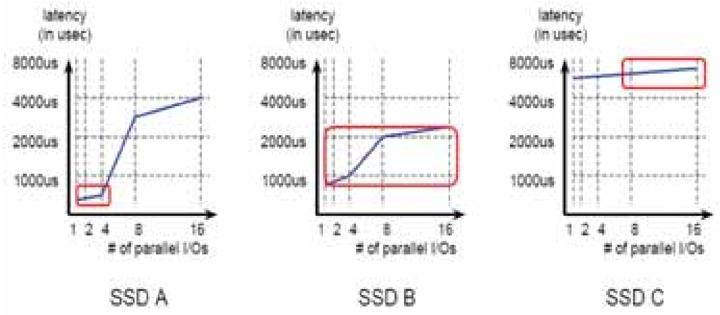


Fig. 3. Latency Characteristics Graph for Three SSDs. The latency characteristics model of SSD A actually reflects the Intel 320 SSD, while the SSD C reflects a certain SSD in an unusual state with an unexpected performance problem. The characteristics pattern of SSD B is defined to have somewhat in-between characteristics.

If we configure the I/O traffic to be distributed evenly throughout three SSDs, then each SSD will get the 8 I/O streams. Then, referring to the values of Table 1, the average latency of each SSD shows is respectively, SSD A: 3,000us, SSD B: 2,000us, and SSD C: 6,300us. This case is depicted in Fig. 4. But otherwise, what will happen if we configure the I/O traffic is adaptively distributed according to the performance characteristics of each SSD? We can easily grasp that SSD A handles very well up to four parallel I/O streams, and the performance of SSD C is very low but the range of latency fluctuation is considerably small. Suppose that we decide to send up to four parallel I/O traffic to SSD A, and to send the remained I/O streams, which originally has to be handled by SSD A, to the SSD C. Fig. 5 represents the expected performance that each SSD will show.

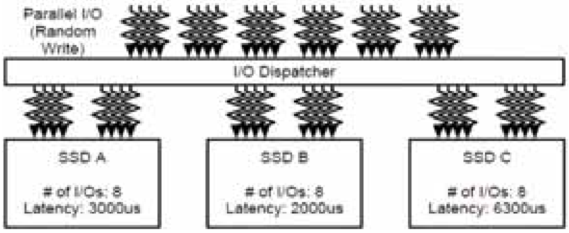


Fig. 4. An SSD array comprised of three SSDs with evenly distributed I/O traffic regardless of each SSD's performance characteristics.

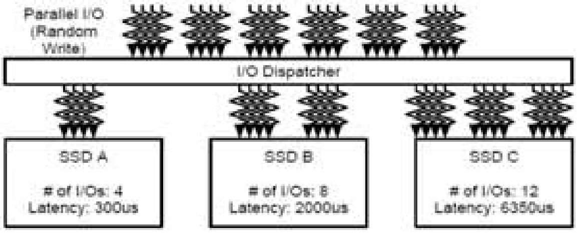


Fig. 5. An SSD array comprised of three SSDs with adaptively distributed I/O traffic according to each SSD's distinguishing performance characteristics.

Comparing the latency represented in Fig. 5 with that in Fig. 4, we can see that there is considerable performance improvement in the SSD A, while the amount of performance degradation of SSD C by having more I/O traffic is relatively small.

Quantitative Performance Metric

We need to measure the overall performance gain quantitatively. We decided to use the reciprocal of average latency value, because latency is the elapsed time to process a unit I/O operation by definition. So we get the IOPS metric of each SSD by take the reciprocal of each average latency value, and also we get the aggregated IOPS of the SSD array by summing up the each SSD's IOPS value when number of parallel I/O streams is given. Table 2 presents the definition of aggregated IOPS metric.

Table 2. Aggregated IOPS Metric

Notation	Description	Eq.
$\mathrm{Nio_i}$	Number of parallel I/O streams delivered to the ith SSD	(1)
$\mathrm{Lat_i(Nio_i)}$	Average latency metric that ith SSD shows when Nio_i parallel I/O streams are given to the ith SSD	(2)
$\frac{1}{\mathrm{Lat_i(Nio_i)}}$	IOPS metric that ith SSD shows	(3)
$\sum_{i=1}^{N}\frac{1}{\mathrm{Lat_i(Nio_i)}}$	Aggregated IOPS metric	(4)

We can calculate the aggregated IOPS with (4). Let's recalculate the aggregated IOPS metric for both cases of Fig. 4 and Fig. 5.

$$\sum_{i=1}^{N}\frac{1}{\mathrm{Lat_i(Nio_i)}} \text{ : aggregated IOPS for Fig. 4 case}$$

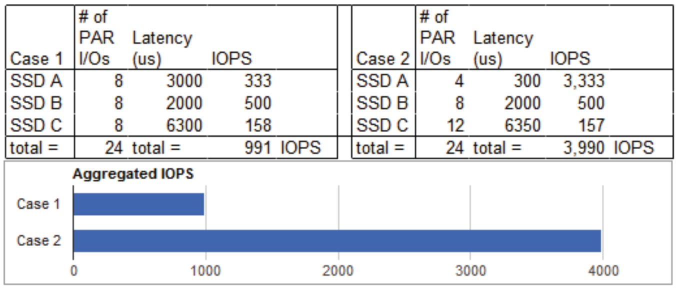
$$= \frac{1}{3{,}000\ {}^{\mathrm{us}}/_{\mathrm{io}}} + \frac{1}{2{,}000\ {}^{\mathrm{us}}/_{\mathrm{io}}} + \frac{1}{6{,}300\ {}^{\mathrm{us}}/_{\mathrm{io}}} \qquad (5)$$

$$= (333\ + 500\ + 158)\ \text{IOPS} = 991\ \text{IOPS}$$

$$\sum_{i=1}^{N}\frac{1}{\mathrm{Lat_i(Nio_i)}} \text{ : aggregated IOPS for Fig. 5 case}$$

$$= \frac{1}{300\ {}^{\mathrm{us}}/_{\mathrm{io}}} + \frac{1}{2{,}000\ {}^{\mathrm{us}}/_{\mathrm{io}}} + \frac{1}{6{,}350\ {}^{\mathrm{us}}/_{\mathrm{io}}} \qquad (6)$$

$$= (3{,}333\ + 500\ + 157)\ \text{IOPS} = 3{,}990\ \text{IOPS}$$



Case 1	# of PAR I/Os	Latency (us)	IOPS
SSD A	8	3000	333
SSD B	8	2000	500
SSD C	8	6300	158
total =	24	total =	991 IOPS

Case 2	# of PAR I/Os	Latency (us)	IOPS
SSD A	4	300	3,333
SSD B	8	2000	500
SSD C	12	6350	157
total =	24	total =	3,990 IOPS

Fig. 6. Aggregated IOPS as a Quantitative Performance Metric. Aggregated IOPS metric is used to quantitatively measure the overall system performance. This graph clearly shows significant performance improvement slightly over fourfold.

V. Mathematical Optimization and Sweet Spots

We start this section with the definition of mathematical optimization. Mathematical optimization is the selection of a best element from some set of available alternatives regarding some criteria or constraints [9]. In the simplest case, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function. We need to find the set of optimal numbers of parallel I/O streams to go to each member SSD, which can be denoted as NIO in (7).

$$\mathrm{NIO} = \left\{\mathrm{Nio_1, Nio_2, \ldots, Nio_N : maximing} \sum_{i=1}^{N}\frac{1}{\mathrm{Lat_i(Nio_i)}}\right\} \qquad (7)$$

We can use mathematical optimization method to find one or more optimal solutions NIO when SSD characteristics are given as Table 1. There are some constraints we can use for mathematical optimization.

$$\text{(8a) } \mathrm{Nio_1 + Nio_2 + Nio_3 = 24}$$
$$\text{(8b) } \mathrm{Nio_1, Nio_2, Nio_3\ \in Z}$$
$$\text{(8c) } \mathrm{Nio_1 \geq 0, Nio_2 \geq 0, Nio_3 \geq 0} \qquad (8)$$

And optionally, for ease of manual optimization,

$$\text{(8d) } \mathrm{Nio_1, Nio_2, Nio_3\ \in \{0, 1, 2, 4, 8, 16\}}$$

Table 6. Exhaustive Search Space for NIO.

$\mathrm{Nio_1}$ (for SSD A)	$\mathrm{Nio_2}$ (for SSD B)	$\mathrm{Nio_3}$ (for SSD C)
4	4	16
4	16	4
8	8	8
8	16	0
16	4	4
16	8	0

Thanks to the constraint (8d), we could get this simple search space table. But it is surely possible to rebuild this search space table without the constraints (8d). Then the size of the search space will grow larger than this because other integer value other than 1, 2, 4, 8, and 16 can be considered. In that case, we can use extrapolation or interpolation method.

Table 6 shows the search space made using all the constraints listed in (8). The rightmost column for $\mathrm{Nio_3}$ (SSD C) can be removed away using the constraints (8a), because we have an assumption that total number of parallel I/O streams is persistently 24. So the function for the aggregated IOPS (will be abbreviated as A_IOPS, later on) can be described with just two parameters $\mathrm{Nio_1}$ and $\mathrm{Nio_2}$ as the following.

$$\text{A\_IOPS} = f(\mathrm{Nio_1, Nio_2}) \qquad (9)$$

For ease of demonstration, we use MINITAB program to find the optimal solution. We made an input table for

MINITAB surface plot using the definition of A_IOPS described in (4), the relationship between A_IOPS and Nio_1 and Nio_2 defined in (9), and the performance characteristics defined in Table 1. The resultant graph from MINITAB, which provides simple mathematical optimization, is depicted in Fig. 7.

Sweet Spots - Set of Workload Patterns to Maximize the System Performance

The surface plot created by MINITAB explains the relationship between the aggregated IOPS of the given SSD array system and the controlling parameters Nio_1 and Nio_2. This tells that the smaller number of parallel I/O streams to SSD A and SSD B, the bigger aggregated IOPS of the SSD array system. More specifically, the aggregated IOPS will be maximized when Nio_1 is 4 and Nio_2 is also 4, which means Nio_3 is 16. In terms of parallel I/O workloads, this set NIO = { 4, 4, 16 } is the optimal pattern for this SSD array, we call it 'sweet spot' workload pattern. If possible, we can get the maximized (or at least fairly high) performance by trying to manipulate the parallel I/O streams close to the sweet spot workload pattern(s) for the SSD array system given.

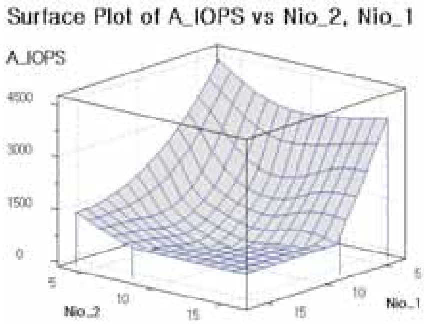


Fig. 7. Surface Plot which shows sweet spot, where A_IOPS is maximized at Nio_1 is 4, and Nio_2 is also 4. This graph shows which combinations of Nio_1 and Nio_2 can maximize the aggregated IOPS (A_IOPS) of the SSD array system. The optimal solution in this case, what we call 'sweet spot' is NIO = {4, 4, 16}, which means Nio_1 is 4 and Nio_2 is also 4.

We can extend this 'sweet spot' research to other axes of I/O workload patterns to find more opportunities to maximize the I/O system performance. In addition to the parallel I/O workloads, we may further consider another types of workloads, such as random-sequential interleaving patterns, and continuity duration patterns of massively random I/O, and so on.

VI. Conclusions

We performed the workload-performance analysis on the NAND flash-based SSD array systems, especially regarding the parallel I/O workload. Lessons learned from this study are:

(a) SSD array system has sweet spots, the workload patterns which can maximize the overall system performance. This sweet spot approach is possible due to the distinguishing performance characteristics of each member SSD. We showed that the SSD array performance can be fairly improved by transforming I/O workload close to the sweet spot workload pattern.

(b) It is possible to use mathematical optimization to find optimal workload patterns when the performance characteristics and the system-wide workload conditions are given. This is very helpful method to decide which workload patterns are the best, or the next best choices in current situation.

Future Work

We will extend this 'sweet spot' research to other axes of I/O workload patterns to find more opportunities to maximize the I/O system performance. In addition to the parallel I/O workloads, we may further consider another types of workloads, such as random-sequential interleaving patterns, and continuity duration patterns of massively random I/O, and so on.

In addition to sweet spot analysis, we have a plan to evaluate the performance improvement by the newly implemented I/O dispatcher engine which handles the I/O workload adaptively based on the performance characteristics of SSD array systems.

References

[1] N. Agrawal, V. Prabhakaran, T. Wobber, J. Davis, M. Manasse, and R. Panigrahy, “Design tradeoffs for SSD performance,” in USENIX Annual Technical Conference, 2008, pp. 57-70.

[2] F. Chen, D. Koufaty, and X. Zhang, “Understanding intrinsic characteristics and system implications of flash memory based solid state drives,” in Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems. ACM New York, NY, USA, 2009, pp. 181-192.

[3] Pure Storage, "The I/O Blender", http://www.purestorage.com/blog/the-io-blende/

[4] L. Bouganim, B. Jonsson, and P. Bonnet. "uFlip: Understanding flash IO patterns," CIDR, (2009)

[5] Intel SSD 320 Series (120GB, MLC), http://ark.intel.com/products/series/56553

[6] Samsung SSD 830 Series (120GB, MLC), http://www.samsung.com/us/computer/memory-storage/MZ-7PC128N/AM

[7] D. Ajwani, I. Malinger, U. Meyer, and S. Toledo, “Characterizing the performance of flash memory storage devices and its impact on algorithm design,” MPI-I-2008-1-001, Tech. Rep., 2008.

[8] M. Moshayedi and P. Wilkison, “Enterprise ssds,” Queue, vol. 6, no. 4, pp. 32-39, 2008.

[9] Mathematical Optimization, Wikipedia, http://en.wikipedia.org/wiki/Mathematical_optimization

A Pseudo Metamesh Approach for 3D Mesh Morphing

Bogdan MOCANU and Titus ZAHARIA

***Abstract*—The effect of morphing from one object to another has become a popular trend in computer graphics due to its applicability in various domains as architecture, gaming, cinema... In this paper, we propose a novel 3D mesh morphing algorithm for genus-0, closed manifold models. The technique firstly embeds the two 3D source and target objects into a common, spherical domain. For this we employ a modified version of the Gaussian curvature that returns a locally flattened version of the original models with a convex structure which can be simply projected onto the unit sphere. By overlapping the two embeddings and warping them in a suitable manner with the aid of RBF functions, we can establish a correspondence between the models. We also introduce a new method to create the supermesh model that share the topology of both input objects and which can easily be transformed from the source model into the target.**

I. INTRODUCTION

Morphing methods are today extensively used in computer graphics to simulate the transformation between two completely different objects or to create new shapes as a combination of other existing shapes. Morphing has a variety of applications ranging from special effects in film industry and other visual arts to medical imaging and scientific purposes.

The problem of constructing a smooth transition between two objects was firstly addressed in the 2D case [1], [2]. Such approached take advantage of the common and regular topology of the 2D images, defined on rectangular grids of pixels. However, in the 3D case, mesh models can exhibit highly irregular topologies/connectivities and are in most of the cases defined with different numbers of vertices/edges. Thus, the mature 2D approaches cannot be extended in a straightforward manner to 3D meshes.

Elaborating advanced and efficient 3D morphing methods could have a strong economical impact on the graphics industry, specifically within the framework of content/special effects production. In this paper we aim to present a 3D morphing method which can ensure high quality, smooth and as gradual as possible transition sequences, consistent with respect to both geometry and topology, and visually pleasant.

The rest of this paper is organized as follows: After a brief recall of the most important methods dedicated to 3D mesh morphing, Section III details the steps of the proposed algorithm. Section IV presents and discusses the experimental results obtained. Finally, Section V concludes the paper and opens some perspectives of future work.

II. RELATED WORK

The first step in mesh morphing is to establish a bijective correspondence from each vertex of the source to a point (vertex or point on an edge/face) on the target surface. This makes it possible to construct a set of in-between models, by interpolating corresponding points from source to target positions.

The issue of mesh correspondence is treated in an indirect manner with the help of parameterization techniques, which consists of constructing a one-to-one mapping between the 3D mesh surface and a common parametric 2D domain. Various parameterization techniques are today available, including planar [3]-[5], and spherical mappings [6]-[8].

In [3], authors use harmonic mappings to embed two models onto the unit disk. The parameterizations are merged and a new mesh with connectivity inherited from both models is created. Furthermore, for closed genus-0 objects, different approaches [9], [10] describe a user specified technique to cut the models into patches and then use a similar method as [3] to obtain the embeddings.

Since closed genus-0 models are topologically equivalent with the sphere, Alexa [6] propose to directly establish the correspondence in the spherical domain. The mesh vertices are projected onto the unit sphere and a relaxation process is optimizing their corresponding position. However, the final embedding is not guaranteed to be valid in all cases.

An alternative to those methods is presented by Lee *et al.* [11] which utilize a multiresolution parameterization algorithm (MAPS) to generate coarse models where only the feature vertices are kept. The correspondence of the original models is computed by constructing a map between the source and target coarse versions and exploits harmonic maps. This method can lead to fold-overs and the user is required to manually fix the problem. A different approach is proposed in [12] which directly maps the connectivity of the source mesh onto the target mesh without partitioning or flattening the models onto 2D plane/3D sphere domain. Based on a mean-value Laplacian fitting scheme they manage to compute a shape preserving correspondence directly in 3D space.

Once correspondence is established, linear interpolation is frequently preferred due to its simplicity. The main drawback is related to the self-intersections that might appear. Different approaches attempt to improve the quality of the morphing sequence. Alexa *et al.* [13] propose to interpolate Laplacian coordinates instead of Cartesian coordinates in order to determine the vertices trajectory during the transformation. Since the Laplacian coordinates are not invariant under rotation and scaling the results is not satisfactory in all cases. Better results can be obtained using the dual Laplacian coordinates [14], or the pyramid coordinates technique [15].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

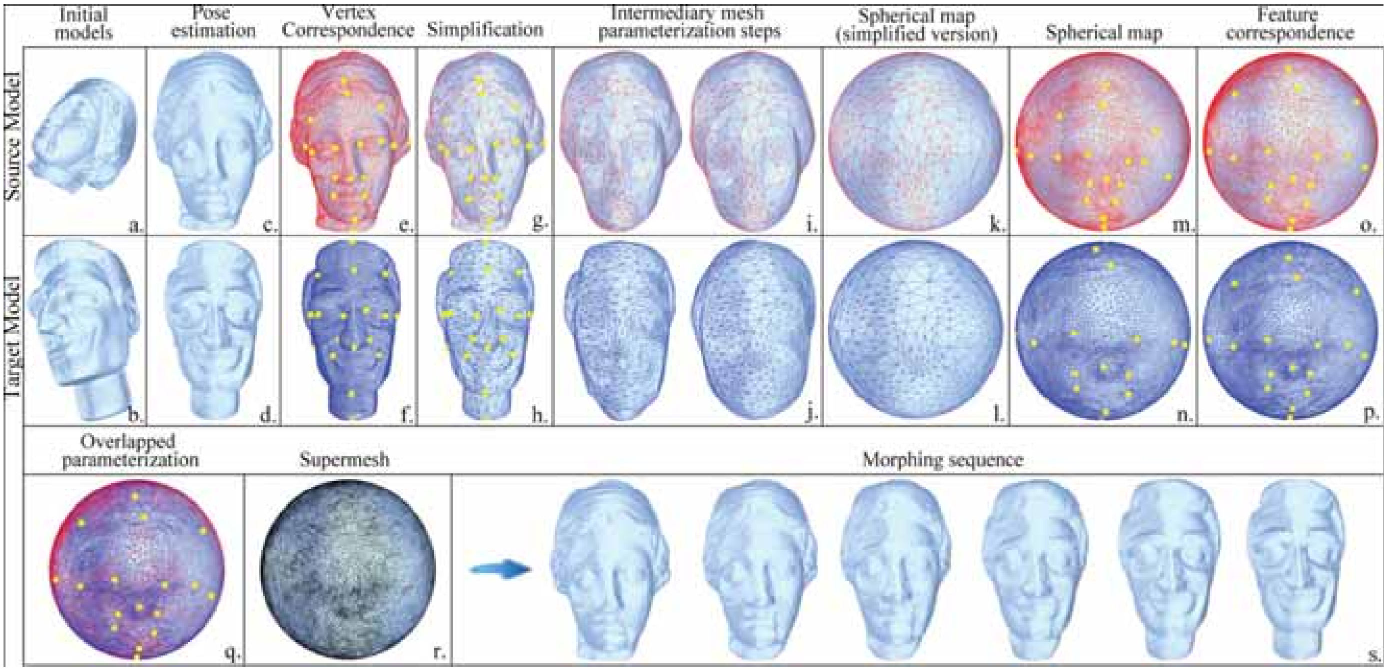


Fig. 1. Steps involved in our morphing process.

III. SPHERICAL MAP-BASED MESH MORPHING

A. Overview and Contributions

Fig. 1 illustrates the proposed 3D mesh morphing process, with the various stages involved. Our algorithm starts with two pre-processing phases, which are PCA pose normalization and mesh simplification. Then, we employ our previous work presented in [16] that introduces a Gaussian curvature driven spherical parameterization method for 3D genus-0, two-manifold meshes.

Next, the mesh structure is reconstructed through a progressive mesh sequence which optimally reinserts the vertices removed in the simplification process. Then, we employ a warping scheme based on RBF functions in order to align the main features of the two models. Moreover, we introduce additional constraints to maintain all vertices on the sphere and to avoid triangles overlapping.

Once the two spherical parameterizations are obtained and the feature vertices are aligned, we can overlap the embeddings and create the supermesh. We introduce here a novel and simple method of overlapping that creates a mesh structure capable to approximate both the source and target topologies and geometries.

The various stages involved are detailed in the following sections.

B. Mesh Simplification

We propose a mesh simplification scheme as an intermediate step in the morphing process in order to simplify as much as possible the parameterization operation that normally require important computational resources for meshes with a high number of vertices. The goal of this step is to produce coarser versions of the input mesh by iteratively reducing the number of vertices and triangles.

The proposed technique is based on the simple edge collapse operator introduced in [17]. However, instead of using their measure to establish the order for the removed vertices, which suffers from a high computational complexity, we have adopted the quadric error metric proposed by Garland and Heckbert [18].

In contrast with the original technique presented in [18], which allows the contraction of arbitrary pairs of vertices, in our case we join solely vertices that define an edge in the mesh. This constraint is useful for preserving the original mesh topology.

In addition, the decimation process is controlled such that the simplified mesh obtained still preserve the main geometrical features of the initial model. More precisely, the decimation stops when the mean geometry deviation between two simplified versions M' and M" of the original model exceeds a pre-established threshold T_{err}.

In our experiments, we have chosen empirically a value of 0.0025 for the T_{err} parameter and a value of 100 collapsed edges between M' and M". These values provide a good compromise between quality of the resulting simplified meshes and the computational time required.

The proposed decimation algorithm yields high quality results that preserves the initial model's shape and topology (Fig. 2) even for drastic simplification rates, in a relatively short time.

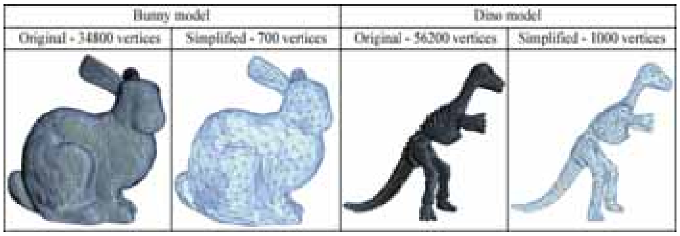

Bunny model
Dino model
Original - 34800 vertices
Simplified - 700 vertices
Original - 56200 vertices
Simplified - 1000 vertices

Fig. 2. Two examples of mesh simplification.

C. *Gaussian Curvature-based Spherical Parameterization*

The employed parameterization algorithm is based on one of previous work presented in [16] and consists in the following three steps:

1) *Iterative Flattening based on Gaussian Curvature*
2) *Spherical Projection*
3) *Vertex Split Sequence*

Fig. 3 illustrates various intermediate steps involved in the spherical parameterization process.

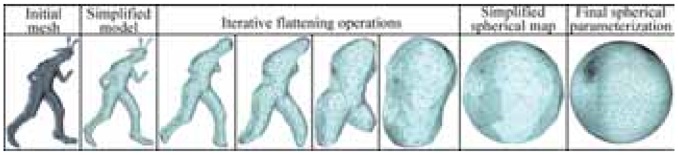

Initial mesh
Simplified model
Iterative flattening operations
Simplified spherical map
Final spherical parameterization

Fig. 3. Spherical parameterization process.

D. *Mesh Mapping Warping and Feature Correspondence*

The warping stage aims at aligning the corresponding features of the two models. This problem is solved in the spherical domain, with the help of radial bases function. Based on the results in [19], we have adopted the CTPS C^2_a radial basis function since it ensures the best compromise between the displacement accuracy and the level of distortions. CTPS C^2_a function can be mathematically expressed as:

$$\phi_r(\xi)=\phi(x/r)=1-30\xi^2-10\xi^3+45\xi^4-6\xi^5-60\xi^3\log(\xi) \quad (1)$$

where r represents the support radius that control the influence of a vertex displacement.

The displacement d can be approximated by the sum of basis functions:

$$d_i(x)=\sum_{i=1}^{n_c}\alpha_i\phi(\|p_j-p_{ci}\|),\quad j=1,\dots,N \qquad (2)$$

where N is the total number of vertices, p_{ci} represents the known vertices spatial positions and n_c the number of these vertices. The weights α_i can be estimated using linear least squares fitting.

We chosen to successively decompose the RBF in a number of steps (between 10 and 100), since directly determining a global RBF deformation would lead to a mesh surface which would not lie on the unit sphere. In addition, such an approach avoids fold-overs and self intersections. In order to guarantee that the final embedding remains valid we propose projecting the mesh back onto the unit sphere after each step.

E. *Supermesh Construction*

Previous works [6], [9], [7] treat the supermesh construction problem by overlapping the two maps of the models, followed by an iterative operation of edge insertion. However, this approach proves to be very challenging due to numerical instabilities when computing intersections between source and target edges. Furthermore, the number of vertices of the resulting metamesh will drastically increase.

In order to overcome this drawback we propose further a simple technique to create pseudo supermesh that avoids tracking the edge intersections. We initialize first the supermesh structure with the one of the target connectivity.

Then for each source map vertices we establish the supermesh triangle in which it can be projected. Each face that contains at least one vertex is subdivided into a 1-to-4 scheme as illustrated in Fig. 4. The process is repeated until all source vertices are included in the supermesh structure. When a source vertex p^s is included in a triangle $f = (p_i, p_j, p_k)$ on the parametric domain, the final 3D position p is computed using the barycentric coordinates (α, β, γ):

$$p=\alpha\cdot p_i+\beta\cdot p_j+\gamma\cdot p_k \qquad (3)$$

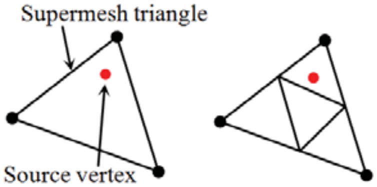


Fig. 4. 1-to-4 subdivision scheme.

F. *Geometry Interpolation*

Since the supermesh structure is able to approximate both the source and target models, the intermediate shapes at various time steps of the morphing transformation can be now calculated. At the moment *t = 0 (t = 1)* the vertex positions with respect to the source (resp. target) object are known. The simplest way to interpolate between these points is a linear interpolation:

$$M^t=(1-t)M^S+tM^T \qquad (5)$$

IV. Results

Based on the methods described in the previous section, we developed a prototype application that allows user to interactively operate with 3D models and control the morphing process. The intuitive interface permits user to select the correspondent vertices in the two models or to save the processed meshes at any time. Fig. 5 presents a snapshot of the user interface. The left window display the source model, while the right one displays the target object.

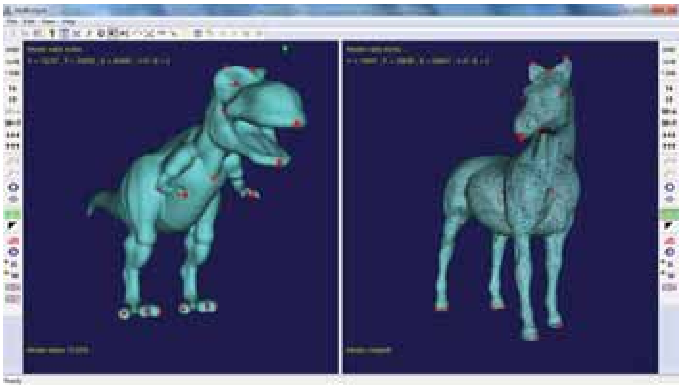


Fig. 5. Graphical user interface.

Fig. 6 illustrates 2 different cases of metamorphosis: one between the Armadillo model (15100 vertices and 30196 faces) and the Man model (17530 vertices and 35056 faces) and the other one between the Cow model (11610 vertices and 23216 faces) and the Triceratops model (2832 vertices and 5660 faces). All models are 3D closed genus-0 objects selected.

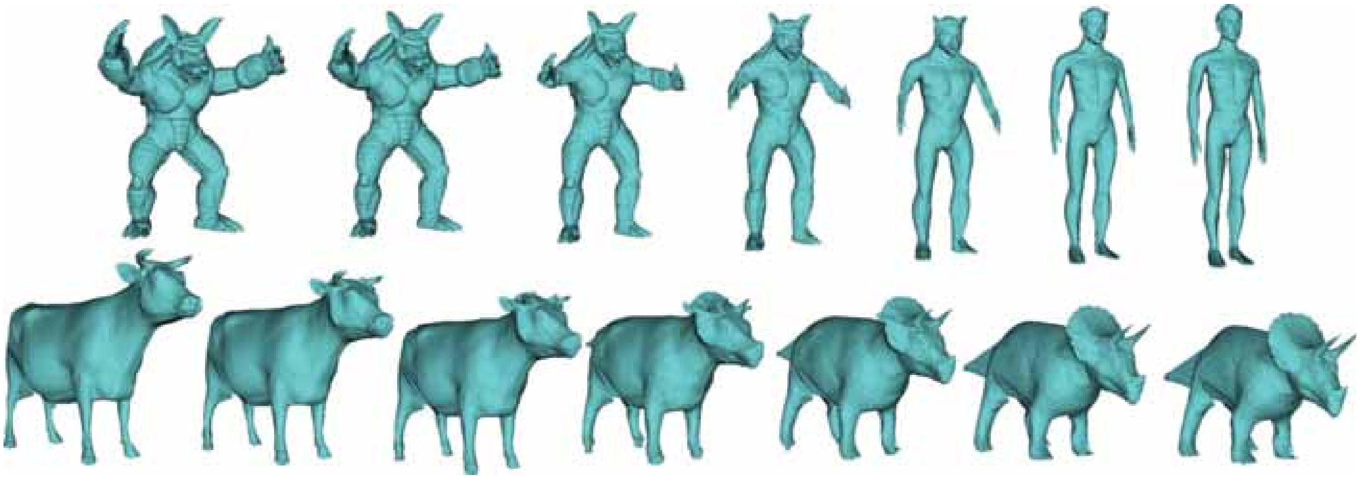


Fig. 6. 3D mesh morphing examples.

We can observe that in both cases the resulting morphing sequences ensure a gradual and visually pleasant transition between source and target models. In addition, the pseudo supermesh considered is able to adapt to both source and target shapes.

V. CONCLUSIONS AND PERSPECTIVES

In this paper, we proposed a novel and complete morphing technique that permits smooth and natural transformations between 3D closed genus-0 meshes. The technique involves a minimum user intervention through a dedicated GUI to specify only some vertices of correspondence.

A spherical mapping has been considered as intermediary step, followed by a feature matching that employs a mesh warping scheme based on radial basis function. Furthermore, we have introduced a simple, but efficient technique of mesh merging to create a pseudo supermesh structure that approximates well both source and target 3D shapes.

Perspectives of future work concern: (a) the interpolation mechanism: linear interpolation should be replaced with other alternatives in order to avoid self-intersection at the level of in-between models; (b) the re-triangulation of the pseudo supermesh which leaves room for optimization.

REFERENCES

[1] S. M. Shontz and S. A. Vavasis, "A mesh warping algorithm based on weighted Laplacian smoothing," Proceedings of the Tenth International Meshing Roundtable, Sandia National Laboratories, Santa Fe, pp. 147–158, 2003.

[2] M. T. Rahman, M. A. Al-Amin, J. B. Bakkre, A. R. Chowdhury, M. A. Bhuiyan, "A novel approach of image morphing based on pixel transformation", Computer and information technology, pp.1-5, 2007.

[3] T. Kanai, H. Suzuki, and F. Kimura. "Three-dimensional geometric metamorphosis based on harmonic maps", The Visual Computer, vol. 14(4), pp. 166–176, 1998.

[4] T. Kanai, H. Suzuki and F. Kimura, "Metamorphosis of Arbitrary Triangular Meshes", IEEE Computer Graphics and Applications, pp. 62-75, 2000.

[5] L. Chao-Hung and L. Tong,"Metamorphosis of 3D polyhedral models using progressive connectivity transformations", IEEE Transaction on visualization and computer graphics, vol. 11(1), pp. 2-12, 2005.

[6] M. Alexa, "Merging polyhedral shapes with scattered features", The Visual Computer, vol. 16, pp. 26-37, 2000.

[7] Z. J. Zhu and M. Y. Pang, "Morphing 3D Mesh Models Based on Spherical Parameterization", International Conference on Multimedia Information Networking and Security, pp. 309–313, 2009.

[8] T. Athanasiadis, I. Fudos, C. Nikou and V. Stamati, "Feature-based 3D morphing based on geometrically constrained sphere mapping optimization", 25th ACM Symposium on Applied Computing (SAC'10), Sierre, Switzerland, pp.1258-1265, 22-26 March 2010.

[9] R. Urtasun, M. Salzmann and P. Fua, "3D Morphing without user interaction", EPFL Technical report 2004.

[10] J. B. Yu, J. H. Chuang, "Consistent mesh parameterizations and its application in mesh morphing", Proc. Computer Graphics Workshop, Hualian, 2003.

[11] A. Lee, D. Dobkin, W. Sweldens and P. Schröder, "Multiresolution Mesh Morphing", Proceedings of SIGGRAPH 99, pp. 343-350, 1999.

[12] H.Y. Wu, C. Pan, Q. Yang, and S. Ma, "Consistent correspondence between arbitrary manifold surfaces". In ICCV, pp. 1–8, 2007.

[13] M. Alexa, "Local control for mesh morphing", Proceedings of Shape Modeling International, pp. 209-215, 2001.

[14] J. Hu, L. Liu and G. Wang, "Dual Laplacian morphing for triangular meshes", Computer Animation and Virtual Worlds, vol.18(4/5), pp. 271-277, 2007.

[15] A. Sheffer and V. Kraevoy, "Pyramid Coordinates for Morphing and Deformation", Proc. 3D Data Processing, Visualization and Transmission Conference (3DPVT), pp. 68-75, 2004.

[16] B. Mocanu, T. Zaharia, "Direct spherical parameterization of 3D triangular meshes using local flattening operations", 7th International Symposium on Visual Computing, pp. 611–622, Las Vegas, USA, 2011.

[17] H. Hoppe, "Mesh Optimization", In Proceedings of ACM SIGGRAPH, pg. 19-26, 1993.

[18] M. Garland and P. S. Heckbert, "Surface Simplification Using Quadric Error Metrics", In 24th Annual Conference on Computer Graphics and Interactive, pg. 209-216, 1997.

[19] A. Boer, M.S. Schoot and H. Bijl, "Mesh deformation based on radial basis function interpolation", Computers & Structures, vol. 85, pp. 784-795, 2007.

Motion Estimation Algorithm for Periodic Pattern Objects based on Spectral Image Analysis

Seung-Gu Kim, Tae-Gyoung Ahn, and Se-Hyeok Park
Samsung Electronics, Suwon, Gyeonggi, Korea

***Abstract* – In this paper, we propose a motion estimation algorithm for periodic pattern objects that conventional block matching based motion estimation algorithms cannot give reliable results. The proposed algorithm is based on spectral image analysis and statistical object motion calculation. The proposed algorithm doesn't significantly increase computational complexity because it uses Fast Fourier Transform and statistical methods rather than complex object segmentation and pixel matching calculations. Experimental results show effectiveness of proposed scheme with significantly reduced defects in image processing results.**

I. INTRODUCTION

Motion Estimation (ME) is widely used and even an essential part in various applications such as MC based Frame Rate Up-Conversion (FRUC) algorithms and various video coding algorithms. The recursive Block Matching Algorithm (recursive-BMA) is one of ME algorithms that is widely adopted in many applications [1], [2]. However, the recursive-BMA cannot offer accurate ME results under situations of deformable objects, variation in light, and periodic pattern objects that consist of periodically repeated parts.

This paper gives focus on improving accuracy of recursive-BMA ME results of pattern objects because human eyes are more sensitive to defects on high frequency and these defects significantly degrade subjective quality of result images of many image processing algorithms.

To improve ME results on pattern objects, many approaches have been proposed. A concept of 'anchor vector' that can represent a motion of a pattern object is proposed in [3]. In this study, Motion Vectors (MVs) of a pattern object are refined with the anchor vector under spatial locality assumption. In other study, analysis results on periodicity of a pattern object are used for MV correction of the pattern object [4]. In this paper, a MV in pattern objects is replaced with a MV of local minimum SAD (Sum of Absolute Difference) point closest to zero in static case and is replaced with a MV of a neighbor block in moving case.

The proposed method tries to find true MVs of pattern objects by guiding ME procedure to conduct Full Search (FS) in contrast to previous studies replace MVs of pattern objects with MVs of selected blocks or artificially adjusted MVs.

It consists of three parts. Pattern blocks that form pattern objects are recognized first; then, initial MVs for recursive-BMA are calculated based on occlusion estimation and statistical pattern object motion calculation, and finally, search range for FS is adaptively controlled according to characteristics of pattern objects.

Detailed algorithm of proposed method is presented in section II. In section III, experimental results are shown and discussed. Finally, section IV concludes this paper.

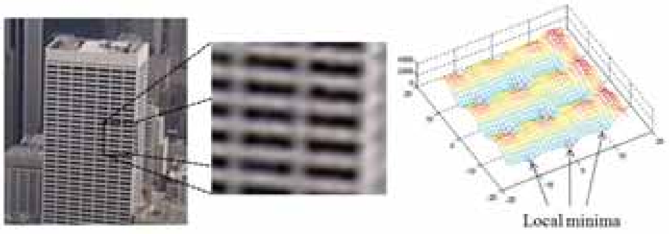
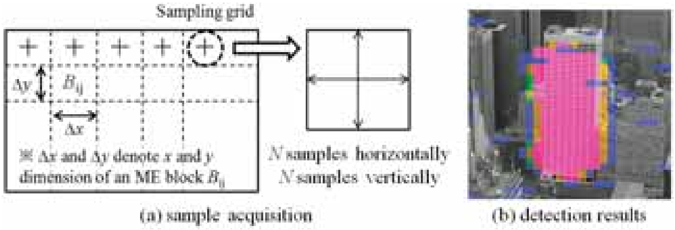


Fig. 1. An example of pattern object and its SAD surface



(a) sample acquisition (b) detection results

Fig. 2. Pixel sample acquisition and pattern block detection results

II. PROPOSED METHOD

A. A Periodic Pattern Object

A periodic pattern object or pattern object is an object in an image that has periodic structure inside it and further divided into smaller pattern blocks to find motion of each block using block matching ME. However, a BMA hardly gives reliable ME results because pattern blocks that have similar shape make multiple minimum SAD points shown in Fig. 1.

B. Spectrum Analysis on Periodic Pattern Objects

Pattern blocks that form pattern objects are detected using spectral image analysis [5]. At first, pre-determined number of pixels is acquired at the center point of each ME block B_{ij} and neighbor pixels. The pixel samples (p_{ij}(n), n is spatial domain sample index) are transformed to frequency domain components (P_{ij}[k], k is frequency domain coefficient index) using FFT. Finally, periodicity intensity of a block B_{ij} is calculated by

$$\textit{Periodicity intensity, } I(B_{ij}) = \max_{k\in S}\left(\left|P_{ij}[k]\right|\right) / \sum_{k\in S}\left|P_{ij}[k]\right|, \qquad (1)$$

where S is set of frequency component P_{ij}[k]. The sample acquisition process and pattern block detection results are shown in Fig. 2 where N is number of pixel samples.

C. Pattern Object Boundary Detection

Boundaries of pattern objects are distinguished because occlusion areas are often appear near object boundaries where different objects meet. Boundaries of pattern objects are detected using derivatives of periodicity intensity among pattern blocks because change of periodicity intensity has

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

maximum value at a boundary of a pattern object and a non-pattern object. The amount of change in periodicity intensity can be defined by gradient $\nabla I(B_{12})$

$$\nabla I(B_{12}) = \frac{I(B_1) - I(B_2)}{\sqrt{(x_{B_1} - x_{B_2})^2 + (y_{B_1} - y_{B_2})^2}}, \tag{2}$$

where B_1 and B_2 are two candidate blocks, $I(B_x)$ is periodicity intensity of a candidate block B_x; x_{Bx}, and y_{Bx} are x and y position of the candidate block B_x. If the amount of the gradient $|\nabla I(B_{12})|$ is larger than a threshold, the candidate block B_1 and B_2 are marked as object boundary blocks.

D. Initial MVs for Pattern Blocks

Initial MV calculation for detected pattern blocks is done in three steps. At first, MVs in $W_{N1}\times W_{M1}$ block window are collected as candidate MVs for initial MVs of the center block in the $W_{N1}\times W_{M1}$ window. Next, MVs of boundary blocks are excluded because they are potential occlusion blocks. Finally, candidate MVs that have larger distance than a threshold value in comparison to the statistically calculated MV of a pattern object are excluded.

The statistical calculation of MV of a pattern object is done in three steps. At first, MVs of detected pattern blocks in another $W_{N2}\times W_{M2}$ block window are collected as candidates. Then, *k-means clustering* is applied to the candidates [6]. Finally, a median MV of a major group from clustering result is selected as an MV that represents a motion of $W_{N2}\times W_{M2}$ area of a pattern object.

E. Search Range Control for Pattern blocks

Search range for pattern blocks in recursive-BMA is adaptively controlled to minimize number of SAD minima points during block matching. It is controlled by

$$Search\ \ Range = \pm\left(\frac{N}{\arg\max\limits_{k\in S}\left(\left|P_{ij}[k]\right|\right)}\right)/2, \tag{3}$$

where N is number of pixel samples used in FFT.

III. SIMULATION RESULTS

In this section, we describe some experimental results of the proposed method. Recursive-BMA combined with proposed algorithm is used for ME between two adjacent images. The block size is 16x12 pixels and search range is adaptively controlled for each initial MV point. In addition, simple-MCI (Motion Compensated Interpolation) scheme is used for FRUC.

Fig. 5 shows FRUC results on test sequence *City*. This test sequence includes many buildings that consist of periodically repeated parts in its scenes. In addition, the test sequence also has occlusion areas due to camera panning. In Fig. 5, a building is being occluded by a foreground building but the proposed algorithm gives clear FRUC result in contrast to conventional method that has broken patterns.

Fig. 6 shows FRUC results on test sequence *BQ-Terrace*.

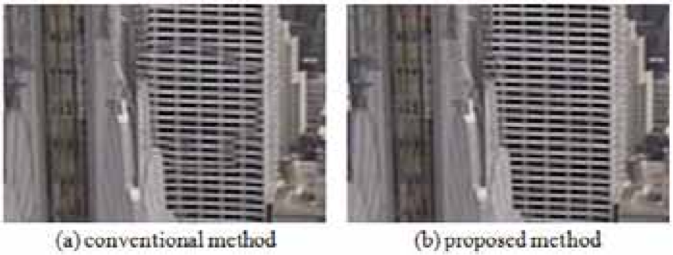

(a) conventional method (b) proposed method

Fig. 5. Results on test sequence *City*

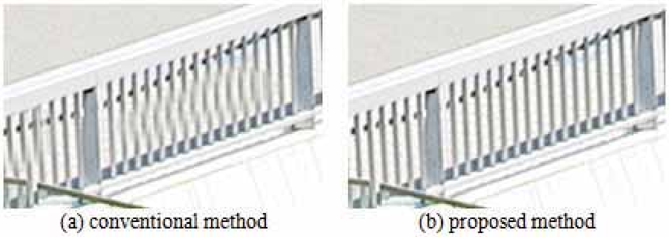

(a) conventional method (b) proposed method

Fig. 6. Results on test sequence *BQ-Terrace*

In this test sequence, the intensity of locality of pattern blocks is lower than the test sequence *City* in Fig. 5. Thus, it is hard to calculate motion of a pattern object with statistical method because the number of pattern block samples within a limited $W_{N2}\times W_{M2}$ block window is relatively small. However, the proposed method gives reliable ME results and the structures of pattern objects in interpolated frames are preserved in contrast to conventional method is fail to find true motion of pattern blocks and make broken patterns in interpolated frames.

IV. CONCLUSION

An ME algorithm for periodic pattern objects is proposed in this paper. The proposed algorithm is based on spectral image analysis and statistical method that do not significantly increase computational complexity because it only needs FFT computations and use previous ME results without any further matching calculations among segmented objects. Experimental results show that the proposed ME algorithm contribute to make high quality interpolated frames in various test sequences.

REFERENCES

[1] A. Heinrich, C. Bartels, R.J. van der Vleuten and G. de Haan. "Robust Motion Estimation Design Methodology," in *Proc. Conf. on Visual Media Production*, pp. 49-57, Nov. 2010.

[2] Mert Cetin and Ilker Hamzaoglu, "An Adaptive True Motion Estimation Algorithm for Frame Rate Conversion of High Definition Video and Its Hardware Implementations," *IEEE Tran. on CE*, Vol. 57, No. 2, pp. 923-931, May, 2011.

[3] Jun-ichi Kimura and Naohisa Komatsu, "Accurate Motion Estimation For Image of Spatial Periodic Pattern," in *Proc. 28th Picture Coding Symposium*, Nagoya, Japan, pp. 250-253, Dec. 2010.

[4] Sung-Hee Lee, Ohjae Kwon, and Rae-Hong Park, "Motion Vector Correction Based on the Pattern-Like Image Analysis," *IEEE Tran. on CE*, Vol. 49, No. 3, pp. 479-484, Aug. 2003.

[5] Surapong Lertrattanapanich and Yeong Taeg Kim, "System and Method for Periodic Pattern Detection for Motion Compensated Interpolation," US Patent, Pub. No. 2008/0317129, Dec. 2008

[6] J. MacQuenn, "Some methods for classification and analysis of multivariate observations," in *Proc of the 5th Berkeley Symposium on Mathematics, Statistics, and Probabilities*, pp.281-297, 1967.

Music Classification Applying Prime Form and Interval-Class Vector

Takashi Maekaku and Hiroyuki Kasai, The University of Electro-Communications, Tokyo, Japan

Abstract **—This paper proposes music classification considering melody transition for browsing system to recommend tunes depending on users' preference. Especially, we apply the Prime form and the Interval-Class Vector to Melodies Markov Model proposed in [1][1].**

I. Introduction

Now that it is available to play lots of songs using a portable music player or a cellular phone, there is more and more demand on finding the music suiting users 'taste that they have never heard before. For users who listen to many kinds of music, the recommendation system should have flexibility covering any kind of music such as a traditional music, indie music and so on as well as popular songs. Since such a minor music does not always have metadata, we concern content-based music information retrieval. To interrelate these music and popular songs, we think it is important to classify unknown music into similar artists especially. Various music classification techniques have been proposed and these methods mostly depend on acoustic features. However, these methods need much time to calculate, we rather use the lower dimension feature vector with score. Here, we propose the method improving the proposal focusing on every composer's preference for how to connect melody pieces [1] about the representation of music and the criteria about zero-frequency problem.

II. Melody Classification

We first construct the Melodies Markov model [1] considering melody transitions by Markov chain and classify music to calculate transition probability. Here we adopted Pitch-Class Set [2] as the state of melody, which has advantage of lower computational complexity.

A. Classification by Markov Chain

As described in [1], we use simple Markov process assuming that a current melody depends on previous one mostly and construct Markov Model for a set of training music letting each measure feature be a state in advance. Given an event sequence (a list of measure summary) $w_1, \ldots, w_m$, which represent music d the probability of the transition from 1 to m, $P(w^m{}_1)$ is described as the product of each transition probability: $P(d)=P(w^m{}_1)=\prod^m_{j=1}P(w_j|\ w_{j-1})$. To classify music d to a class $c_k \in C$, we calculate the following maximum likelihood probability.

$$\begin{aligned} c_k^d &= \operatorname{argmax} P(c_k \mid d) \\ &= \arg\max \pi \prod_{i=1}^{N} P_{c_k}(w_{i+1} \mid w_i) \end{aligned} \quad (1)$$

, where $c^d{}_k$ is a maximum likelihood class for music d, π is a occurrence probability of w_1, N is the number of state, and P_{ck} is a conditional probability of model c_k. We discuss the definition of a state representation w in section 3.

B. Zero-frequency Problem

If the Markov Model is sparse, there is the possibility that conditional probability $P_{ck}(w_{i+1}|w_i)$ on the way of calculating equal zero and the result does not make sense. To solve this problem, Y. Yoshihara and T. Miura [1] adjust the transition probability $P(w_{i+1}|w_i)$ into $P'(w_{i+1}|w_i)$ by summarizing the probability multiplied with cosine similarity: $Sim(w,g)=\boldsymbol{w}\cdot \boldsymbol{g}/\|\boldsymbol{w}\|\cdot\|\mathbf{g}\|$ (where w is a feature vector of w and $\|\boldsymbol{g}\|$ is that of possible state g over all g with w_i defined as follows: $P'(w_{i+1}|w_i)=\sum_g P(w_{i+1}|g)\times sim(g,w_i)$. However, the probability might be larger relatively and it does not seem to compare the case which has the sparseness problem and the other case which does not have correctly by this formula. Therefore, we solve the problem using another method with interval-class vector instead of it as we mention next section.

III. Application of Prime Form and Interval-Class Vector

A. Prime Form

Each pitch (C, $C\#$, D, ..., B) is assigned to the integer from 0 to 11 and the numbers is called "pitch class", and a set of pitch class is called to pitch-class set. Here, octave doublings and displacements are ignored. For example, pitch-class set of C, E, G, and C is [0, 4, 7]. A" Prime Form "represents a group of similar sets that have only differences by transposition or inversion each other. Any pitch-class set has a single prime form and the form can be calculated by the algorithm in [3]. We used this form per measure as a state of Markov chain. Using this form, we can reduce states from 4095 to 208.

B. Interval-Class Vector

A pitch interval is simply the difference between two pitch-class numbers. Considering pitch-class intervals larger than six are equivalent to their complements mod 12 (1= 11, 2 = 10,3=9,4=8,5=7,6=6), there exists six interval classes. Interval-Class Vector is a histogram of the all intervals in a set and is presented as a string of six numbers. It shows how the set sounds. For example, the interval-class vector of the set [0, 4, 7] is

[1] This work is supported in part by the National Institute of Information and Communications Technology (NICT), Japan, under the new generation network R&D program for innovative network virtualization platform and its application.

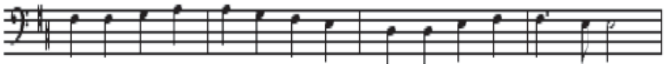
978-1-4673-1363-6/13/$31.00 ©2013 IEEE

<001110>. If two sets have similar interval-class vectors each other, they sound alike together. Fig. 1 shows the melody description using prime form and interval-class vector as a state of Markov Process.

C. Melody Classification using Prime form

As a state of Markov Chain, *pitch spectrum* that consists of only *n* biggest durations' pitch elements considered as a chord in a measure is adopted in [1]. However, the problem is that the *n* tones of the highest pitches is selected if there exists more than *n* candidates. By doing this, we think that the information as the melody feature might be distorted because we cannot mention the highness of pitch unless we know the root of chord in a measure and the highness of pitch has nothing to do with the representative note element. Therefore, the way of choice is not very appropriate as the feature. Hence we determine the class using Melodies Markov Model applying Prime form as a each measure state in order to preserve all pitches with state reduction. There is also possibility that a Pitch-Class in a test set of melody does not exist as a state in Melodies Markov Model. To avoid this zero-frequency problem, we find a set w′ that is most similar to a current set w from Melodies Markov Model and multiply the similarity between the interval vector of set w_i and the vector of w_i' using cosine similarity. Hence, if P_{ck} $(w_{i+1}|w_i) \neq 0$, the weighted probability is described as:

$$P_{c_k}(w_{i+1} \mid w_i) = p_{c_k}(w'_{i+1} \mid w_i) \cdot \max_{w' \in \text{model}} Sim(w_{i+1}, w'_{i+1}). \quad (2)$$



Pitch-Class Set : [6,7,9] → [4,6,7,9] → [2,4,6] → [4,6]
Prime From : (0,1,3) → (0,2,3,5) → (0,2,4) → (0,2)
Interval Vector : <111000> → <122010> → <020100> → <010000>

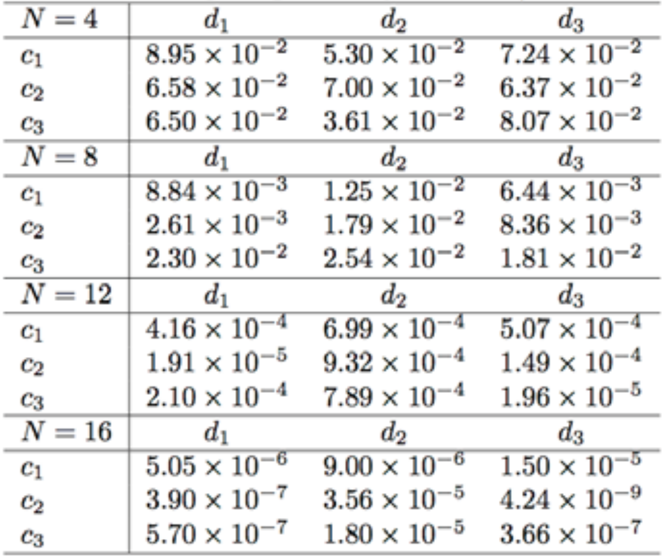
Fig. 1. Melody Description.

IV. Results

We experimented our method using the following some classical music from MIDI file, Brahms (class c_1): "Waltzes Op. 39-1", "Waltzes Op. 39-5", "Rhapsody Op. 79-2", "Variations on a Theme by Haydn", Debussy (class c_2) : "Deux arabesques", "Clair de Lune", "Reflets dans l'eau", "Reverie", and Chopin (class c_3) : "Waltzes Op. 18 in E♭", "Prelude Op. 28", "Nocturnes Op. 9, No.1", "Ballades Op. 38 in F" as a training collection of music, and Brahms : "Hungarian Dances No. 1 (d_1)", Debussy : "Prelude (d_2)", and Chopin : "Etudes Op. 10, No. 1 (d_3)" as a set of test music (letting each desirable class of d_1, d_2 and d_3 be c_1, c_2 and c_3). We used 16 bars in three tunes for creating each model due to 4-fold cross validation and 4, 8, 12, 16 bars for each test data. The procedure is as follows:

- Calculate a prime form for each measure as a state of Markov Model from a training collection of music and generate melodies model.
- Given unknown music d, we calculate the formula (1) and determine the class.
- If a Prime form of test data does not belong to a model as a state, calculate the similarity (2) between a current state and each state of model using interval-class vector and choose one as a next state that makes the similarity maximum.

TABLE I. Weighted probability: Pr (class | music)



$N=4$	d_1	d_2	d_3
c_1	8.95×10^{-2}	5.30×10^{-2}	7.24×10^{-2}
c_2	6.58×10^{-2}	7.00×10^{-2}	6.37×10^{-2}
c_3	6.50×10^{-2}	3.61×10^{-2}	8.07×10^{-2}
$N=8$	d_1	d_2	d_3
c_1	8.84×10^{-3}	1.25×10^{-2}	6.44×10^{-3}
c_2	2.61×10^{-3}	1.79×10^{-2}	8.36×10^{-3}
c_3	2.30×10^{-2}	2.54×10^{-2}	1.81×10^{-2}
$N=12$	d_1	d_2	d_3
c_1	4.16×10^{-4}	6.99×10^{-4}	5.07×10^{-4}
c_2	1.91×10^{-5}	9.32×10^{-4}	1.49×10^{-4}
c_3	2.10×10^{-4}	7.89×10^{-4}	1.96×10^{-5}
$N=16$	d_1	d_2	d_3
c_1	5.05×10^{-6}	9.00×10^{-6}	1.50×10^{-5}
c_2	3.90×10^{-7}	3.56×10^{-5}	4.24×10^{-9}
c_3	5.70×10^{-7}	1.80×10^{-5}	3.66×10^{-7}

Results are shown in Table 1. we can see that d_1, d_2 and d_3 and d_1 and d_2 where N = 16 bars are classified appropriately. We can also say that preferences for composing become clear to some extent to consider melody transition.

V. Conclusion

We presented the music classifying method based on melody transition using Markov model and Prime form and it seems to perform better in terms of computational complexity. Future work includes implementations of the algorithm considering the note duration and arrangement.

References

[1] Y. Yoshihara, and T. Miura, "Classifying Polyphony Music Based on Markov Model," Intelligent Data Engineer- ing and Automated Learning (IDEAL), 2006.

[2] [Allen Forte, *The Structure of Atonal Music,* Yale University Press, 1973.

[3] Joseph N. Straus, *Introduction to Post Tonal Theory,* Prentice Hall, 2005.

Speaker Dependent Visual Speech Recognition Using Extended Curvature Gabor Filters

Jeongwoo Ju[1], Heechul Jung[2], and Junmo Kim[3], *Member, IEEE*
{veryju[1], heechul[2]}@kaist.ac.kr, junmo@ee.kaist.ac.kr[3]
Division of future vehicle, KAIST, Daejeon, Korea[1]
Dept. of Electrical Engineering, KAIST, Daejeon, Korea[2,3]

***Abstract*—Performance of a speech recognition system often degrades severely under low SNR environment. To overcome this difficulty, the visual signal is also considered as an additional aid these days. In this paper, we address speaker dependent visual speech recognition problem using Extended Curvature Gabor (ECG) wavelet. First, lip image sequences are filtered using the ECG, because the variation of the filter response well represents the lip movement. Next, the distance between the output and training data is calculated using the Multi Dimensional Dynamic Time Warping (MDDTW) with new cost matrix. Finally, the lip sequences are classified into the corresponding utterance. In this process, the parameters of ECG must be selected appropriately, where we compare a simple greedy selection method and selection scheme based on AdaBoost**

I. INTRODUCTION

Humans utilize both visual and audio information simultaneously to interpret speech signal by their nature. While both cues being processed, inconsistency between them can cause perception confusion. This phenomenon is called the McGurk effect [1] suggesting importance of visual cue in speech recognition.

The audio carries more useful information for speech recognition than visual information does. However, audio speech can be easily contaminated by acoustic noise. For example, making a sense of speech inside a car with lots of noise source or recognizing a person's speaking under noisy environment is difficult task. In this case, it is hard to obtain reliable information from audio signal leading to decrease in recognition accuracy. As one solution, visual cue can be combined with audio one to reduce the recognition error rate [2]. Similar to human speech perception mechanism, most researchers have developed a variety of methods to integrate visual and audio signals to advance automatic speech recognizer. As a result, it has become widely known that visual cue helps improve the performance of automatic speech recognition. Since human speech production depends on vibration of the vocal tract as well as partly visible articulators such as teeth, tongue and lip, visual data extracted mostly form lip can provide supplementary information.

In this paper, we propose a new method for visual speech recognition which can have enormous potential applications depending on user’s preference. As in previous studies [3],[4], we aim to classify a few numbers of isolated utterances.

In this work, the sum of the output response to Extended Curvature Gabor (ECG) filter plays a role as representing a single lip image frame.

Although two utterances belong to the same class, due to the individual speaker's speech habit and speed, they may have a different number of frames. Therefore, dynamic time warping (DTW) is used to measure the similarity between lip image sequences and find the nearest sequence in the database to classify the test utterance. To increase performance, we extract multi-dimensional information from images using multiple parameters of ECG. Accordingly, we convert DTW to multi-dimensional DTW (MDDTW). Moreover, due to huge number of parameters of ECG filter, it would require exhaustive search in order to find parameters that show best performance. Two approaches are adopted to avoid this case, multi class AdaBoost and a greedy parameter selection method.

II. ECG-BASED VISUAL SPEECH RECOGNITION

A. Extended Curvature Gabor Wavelet

Hwang et al.[5] have extended banana wavelet to Extended Curvature Gabor (ECG) wavelet to capture curvature components efficiently.

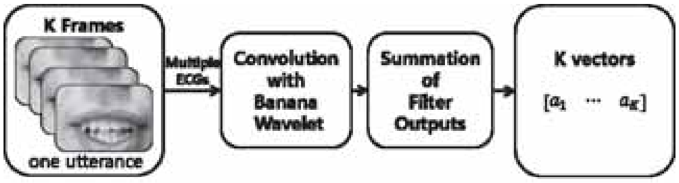


Fig 1. Feature Extraction Process

Most of the lip shapes are composed of the combination of curved and straight lines rather than single component. This fact emphasizes the necessity of utilizing banana wavelet to analyze lip shapes effectively. We assume that the output of image filtered by banana wavelet may vary according to the state of mouth (open, lightly open and completely close). For instance, opened mouth may contain more curved components.

We extract multi-dimensional feature from an utterance using multiple ECG filters with different parameters. The Schematic of feature extraction is illustrated in Fig.1.

B. Multi-Dimensional Dynamic Time Warping

Visual information is contained in not only lip shape but also lip movement. Due to its intrinsic time-bearing information, it is essential to select metric measuring time series similarity. To measure this similarity, MDDTW is utilized. A reader may refer to the literature [6] for detailed review on dynamic time warping. In general, multi-

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

dimensional information plays an important role to enhance performance. Hence, it is necessary to build new cost matrix which is adequate for multi dimensional visual speech input. During cost matrix construction, we follow the same procedure as in [7] and normalize each element by the corresponding 'mean of within class variance' to pay no attention to feature's variation.

Given the test data, we perform the MDDTW between all training data and test data. And then the class of training data having the smallest distance is assigned to the test data. This is called one nearest neighbor MDDTW classification method (1NN MDDTW).

C. Parameter Selection of ECG

Parameters of ECG do not equally contribute the performance. Consequently, parameter selection is required to increase the discriminative power. We approach this problem in two ways, multi-class AdaBoost and a greedy selection strategy.

Detailed proofs and explanations on multi-class AdaBoost can be found in [8]. We adopt multi-class AdaBoost algorithm in two scenarios. In scenario1, 1NN DTW classifiers corresponding to selected set of parameters are strong classifiers given as linear combinations of weak classifiers with individual parameter in the parameter set as in [8]. In scenario2, we build a single 1NN MDDTW classifier corresponding to the chosen parameters.

The greedy parameter selection algorithm is easy to implement and starts by selecting one parameter that shows best performance. Next, second parameter is chosen so as to achieve the best recognition rate when linked together with the firstly chosen one. Similarly, third parameter is selected so that it has best recognition accuracy when associated with the two previously chosen parameters. We repeat the same procedure until the number of chosen parameter reaches to a specified value K. The overall algorithm is illustrated in Fig. 2.

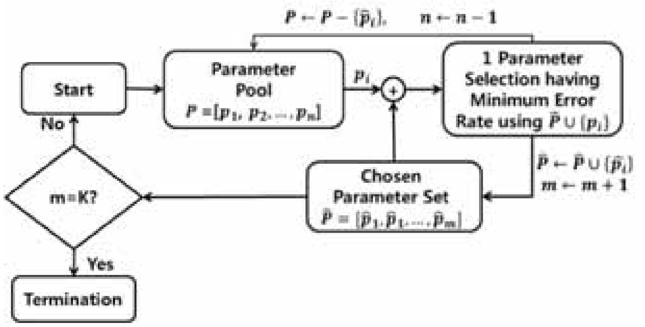

$P \leftarrow P - \{\bar{p}_i\}$, $n \leftarrow n - 1$
Start
Parameter Pool
$P = [p_1, p_2, \cdots, p_n]$
p_i
+
1 Parameter Selection having Minimum Error Rate using $\bar{P} \cup \{p_i\}$
$\bar{P} \leftarrow \bar{P} \cup \{\bar{p}_i\}$
$m \leftarrow m + 1$
No
m=K?
Chosen Parameter Set
$\bar{P} = [\bar{p}_1, \bar{p}_1, \cdots, \bar{p}_m]$
Yes
Termination

Fig 2. The schematic of the greedy parameter selection algorithm

III. EXPERIMENTAL RESULTS

Evaluation of proposed method was conducted on the most recently constructed OuluVS Database [3]. Depending on mouth localization, there are two types of image sequences, automatic sequences and manual sequences. Our experiment was carried out on automatic ones. Due to few samples for single person in OuluVS Data, leave one utterance out strategy was utilized for cross validation; one sequence is left for test and the remaining sequences are used for training. Overall recognition rate using training data are shown in Fig.3.

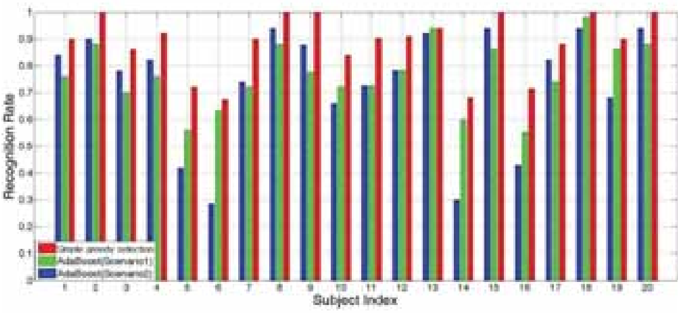


Fig 3. Recognition Rate using Training Data

IV. CONCLUSION

In this paper, ECG is utilized to deal with speech recognition problem. Since parameters in ECG have a great effect on recognition performance, a greedy parameter selection method and multi-class AdaBoost are utilized to find parameters that show reasonably good performances. Moreover, due to the time series nature of the lip image sequences, we constructed cost matrix which is suitable for multi-dimensional time sequences. In comparison with traditional MDDTW, we normalized each features using its mean of within class variance to equalize its effect on constructing cost matrix. As a result, our method achieved promising recognition rate using training data.

ACKNOWLEDGMENT

This work is funded by Korea Institute for Robot Industry Advancement (NIPA-2012-H1502-12-1002)

REFERENCES

[1] H. McGurk and J. MacDonald. "Hearing lips and seeing voices". 1976.

[2] C. Neti, G. Potamianos, J. Luettin, I. Matthews, H. Glotin, D. Vergyri, J. Sison, A. Mashari, and J. Zhou. "Audio-visual speech recognition." In Final Workshop 2000 Report, volume 764, 2000.

[3] G. Zhao, M. Barnard, and M. Pietikainen. "Lipreading with local spatiotemporal descriptors.", Multimedia, IEEE Transactions on, 11(7):1254-1265, 2009.

[4] E.J. Ong, R. Bowden, and G. GU27XH. Learning sequential patterns for lipreading. Procs. of BMVC To Appear, Dundee, UK, Aug, 2011.

[5] W. Hwang, X. Huang, K. Noh, and J. Kim. "Face recognition system using extended curvature gabor classifier bunch for low-resolution face image". In Computer Vision and Pattern Recognition Workshops (CVPRW), 2011.

[6] P. Senin. "Dynamic time warping algorithm review", Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA, 2008.

[7] GA Ten Holt, MJT Reinders, and EA Hendriks. Multi-dimensional dynamic time warping for gesture recognition. In In Proc. of the conference of the Advanced School for Computing and Imaging (ASCI 2007), 2007.

[8] J. Zhu, S. Rosset, H. Zou, and T. Hastie. Multi-class adaboost. Technical report, Stanford Univ, 2006.

A New Low Energy IMF based Audio Stenographic Technique

Saif alZahir, *Member, IEEE* and Md. Wahedul Islam, *Student Member, IEEE*
University of N. British Columbia, British Columbia, CANADA

ABSTRACT

We present a new audio steganographic technique based on empirical mode decomposition and Hilbert Transform. The audio signal is decomposed into several intrinsic mode functions to be the addressee for the payload of a QR code. Our results show that the proposed method is robust against common audio processing attacks.

I. INTRODUCTION

Multimedia information security in consumer electronic devices and systems have been realized via two different technologies: encryption and digital steganography. Steganography is the art of hiding secret data in an innocent looking container called cover data. This cover data may be any digital media such as digital image, audio, movie file etc. Usually the embedded secret data is called payload. Embedding information into audio sequences is more complicated task than that of images, due to superiority of the human auditory system (HAS) over human visual system (HVS) [1]. In addition, multimedia message passing in smart phones is rapidly increasing and getting more popular day by day and hence, sending secret messages with stego-audio would be an interesting addition. Due to variations in different digital audio file formats, steganography methods have been developed to exploite such file formats. MP3Stego developed by Petitcolas [2] is a good example. However, the most popular special domain approach is the least significant bit (LSB) substitution approach which was improved by several authors. The literature is scares on EMD based audio steganography methods and has non (to our knowledge) with Quick Response code (QR Code). According to International Federation of Phonographic Industry (IFPI) and STEP2001 [3], to hide data in audio signal, it must have at least 20 dB SNR to be acceptable.

II. EMPIRICAL MODE DECOMPOSITION (EMD)

Researchers have realized that a complex signal should consist of some simple signals, each of which involves only one oscillatory mode at any time instance. Based on this realization, Haung et.al.[4] proposed a new method called EMD which can adaptively decompose a signal into a number of zero mean oscillating components referred to as intrinsic mode function IMFs. Basically, EMD can be represented as an iterative process for a given signal f(t), given as follows:

$$f(t) = r_m(t) + \sum \mathrm{IMF}_m(t), \qquad (1)$$

where $r_m(t)$ is the residual "trend" and $IMF_m(t)$ represents mono-component signal, called IMFs. Furthermore, EMD decomposes the signal into its mono-components according to the signal itself and the number of IMFs is finite. IMFs give meaningful identifications of the instantaneous frequencies that make EMD highly efficient for non-stationary signal analysis and superior to Fourier and wavelet transformations [4]. Although all IMFs contain energy from both the original signal (e.g. audio) and the noise, the amount of energy distribution is different. However, low energy IMFs make it possible to insert payload in an audio that will remain safe even after time scale modification, TSM, and Gaussian attack. Due to its simplicity and elegancy among different methods for computing EMD we choose "Sifting process" for our method. Sifting EMD process is based on the idea of subtracting the component with the longest period from the data till an IMF is obtained or until the final residue is a constant, a function with only one maxima and one minima from which no IMF can be derived. Therefore, the 1st IMF will have the highest oscillating components; the component with highest frequency. In other word, the higher the order of the IMF, the lower its frequency content will be.

III. QUICK RESPONSE (QR) CODE

QR code was invented by Denso Wave in 1994. It is one kind of 2-dimensional code with control points which makes it easier to be interpreted by scanning equipment such as iPhone, Digital Camera and hand held scanner. Moreover, the error correction capability of QR code makes it ideal one for steganography. For different version of QR code there are different module configurations where modules refer to the black and white dots which construct the QR Code. The largest standard QR Code is a Version 40 symbol that 177x177 modules in size and can hold up 4296 characters of alphanumeric data (theoretically) compared to 25 characters for a Version 1 QR Code. In QR code each codeword is 8 bit long and use the Reed–Solomon error correction algorithm for four error correction levels which can resist upto 7%, 15%, 25% and 30% distortion [5]. In this research we used QR code Version 2 with 7% error correction level to host our secret message. Fig 1 illustrates the structure of a Version -2 QR code.

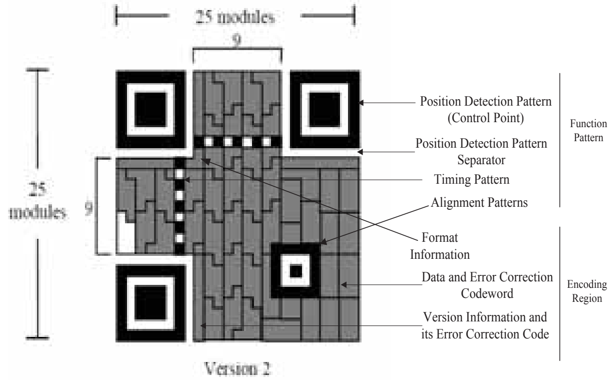

25 modules
9
25 modules
9
Position Detection Pattern (Control Point)
Position Detection Pattern Separator
Timing Pattern
Alignment Patterns
Format Information
Data and Error Correction Codeword
Version Information and its Error Correction Code
Function Pattern
Encoding Region
Version 2

Fig 1 Structure of a Version 2 QR Code Symbol [5]

IV. PROPOSED METHOD

In this section, we present our new stenographic technique using EMD transform and QR code. We use a QR code

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

generator to produce a payload (secret message) which is converted into a one dimensional vector as a sequence of 0.5's and -0.5's. We opted to use the idea of embedding our payload into the significant IMFs of the audio signal frames containing the lowest energy. The proposed new method proceeds as follows: (i) we divide the host signal into 625 numbers of frames ; (ii) each frame is decomposed into a finite and often small number of IMFs that acknowledge well-behaved Hilbert transform; (iii) finally, we calculate the energy of every IMF and then we pick the IMF containing the lowest energy to embed our payload. Number of frame depends on what version of QR-code needs to be hided in the signal.

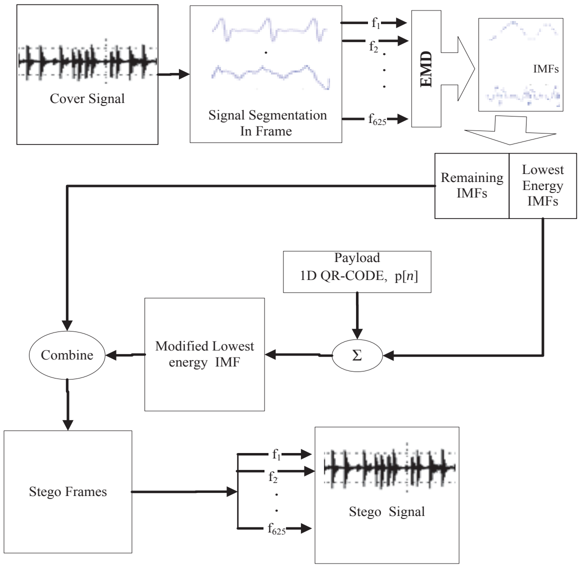


Fig. 2 Payload Embedding process

Once the IMF is chosen, a scaled version the 1D-Sequence of Version 2 QR code sequence, $p[i]$, is added to the selected IMF (lowest energy)by using the following equation:

$$IMF_m'[t] = IMF_m^i[t] + \alpha p[i] \,, \qquad (2)$$

where, $IMF_m^i[t]$ is the selected m^{th} IMF of i^{th} frame, $p[i]$ is the corresponding mark value of a stretched version of the stego signal, and $IMF_m'[t]$ is the stego IMF. The scaling factor α is chosen to be as high as possible while remaining inaudible. Here α is taken as 0.02.

In the extraction process, the original signal $x[t]$ and stego signal $x'[t]$ are transformed into HHT domain. Then the payload is extracted simply by using the following equation:

$$p^i[i] = y_i'[t] - y_i[t]/\alpha, \qquad (3)$$

where, $y_i'[t]$ is the HHT transformed stego audio signal, $y_i[t]$ is the HHT transformed original audio signal of i^{th} frame, and $p'[i]$ is the extracted module (0 or 1) of the i^{th} frame. The resulted watermark $p'[i]$ is averaged over the frame.

V. RESULTS AND DISCUSSIONS

To test the performance of the proposed stego method, we used 5 different types of audio signals of average of 18 seconds length and sampled at 44.1 KHz. Fig. 3 shows the original audio signal, stego audio signal and the QR code that contains the secret message: 'Demo Computer Science'.

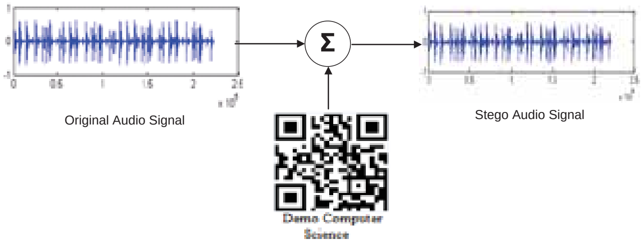


Fig. 3 Hiding Version 2 QR code in drum music.

Table – I shows the SNR and Segmental SNR of the stego audio samples used in the proposed method.

Table - I Results after payload Insertion

AUDIO NAME	SNR	Segmental SNR
Piano Music (6 Sec)	26.9	28.1
Drum Music (8 Sec)	23.8	16.8
Jazz Music (26 Sec)	31.6	31.9
Classical (29 Sec)	32.6	29.6
Svega [2] (20 Sec)	21.4	13.7

This table shows that stego audio signals are near identical to the original audio signal. Also, the stego-audio has passed the requirement of 20dB SNR as referenced earlier. To further verify that our stego method can withstand audio processing attacks, we tested it against Gaussian noise (with mean =0 and variance = 1), delay effect, down sampling (sampled at 22 KHz), and +20 time scale modification with SOLAF algorithm. In every case, we were able to extracted the QR code from the attacked stego signal and hence getting the secret message. However, this method did not withstand MP3 compression test and to do so, some adjustment to our parameters was required.

VI. CONCLUSIONS

The proposed method uses EMD of the audio signal which is the most powerful adaptive transformation technique in signal processing domain with perfect time-frequency localization properties. Using the minimum energy IMFs , we were able to address the most inaudibile portion in the stego-audio with sufficient robustness. Our method shows a great deal of robustness and could resist TSM of +20%, filtering and Gaussian noise to name a few. Finally, employing versions of QR code as payload, for the first time, will allow for larger amount of text to be embedded and provide a new avenue for information hiding.

REFERENCES

[1] I. J. Cox, M. L. Miller, and J. A. Bloom, Digital Watermarking, Morgan Kaufmann Publishers, San Francisco 2002.

[2] F. A. P. Petitcolas, "MP3Stego." http://www.petitcolas.net/fabien/steganography/mp3stego/ accesed on June 01 2012.

[3] STEP2001, [Online], Available: http://www.trl.ibm.com/projects /RightsManagement/datahiding/dhstep_e.html accesed on June 01, 2012.

[4] N. E. Haung, Z. Shen And S. R. Long, "The Empirical Mode Decomposition and Hillbert Spectrum for nonlinear and Non-stationary time series analysis". *Proceedings of the Royal Society of London*, A (454):903-995, 1998.

[5] http://www.denso-wave.com/qrcode/index-e.html. accesed on June 01, 2012

Hardware Trojan for Security LSI

M.Yoshikawa, *Member, IEEE,* R.Satoh, and T.Kumaki, *Member, IEEE*

***Abstract*--A hardware Trojan has become a serious problem of consumer electronics in recent years. The present study proposes a new hardware Trojan in a countermeasure circuit and evaluates the weakness of the countermeasure circuit through simulation experiments.**

I. INTRODUCTION

Large scale integration (LSI) is mounted on almost all consumer electronics, including mobile phones and personal computers. When designing LSI, reducing development cost and shortening the design period are important from the viewpoint of cost effectiveness in the business strategy. To achieve these goals, LSI is designed using intellectual property (IP) in many cases. In this instance, it is important not only to use IP developed by one's own company but to purchase IPs developed by other companies as well. Since IP is often purchased from all over the world, many unspecified engineers may be engaged in designing an LSI circuit. Consequently, a hardware Trojan has become a serious problem of consumer electronics in recent years [1].

A hardware Trojan is a circuit that is incorporated into the LSI by a malicious designer upon designing or manufacturing the LSI. When an LSI has been infected with a hardware Trojan and is used in a consumer electronics, the LSI acts according to its specifications in normal operations. However, when predetermined conditions are satisfied, subversive activities, such as the leaking of in-system information and system shutdown, can occur. A software Trojan can be theoretically removed. However, since a hardware Trojan is physically incorporated into the LSI, it can be difficult to find from the outside and cannot be removed. In previous studies on a hardware Trojan, its basic principle of operation was reported [1].

Cryptographic circuits, which are incorporated into credit and cash cards, are LSI circuits familiar to us. The safety of cryptograms used for cryptographic circuits is computationally secured. Recently, it was reported that confidential information (key information) could be revealed using secondary information, including power consumption and leaked electromagnetic waves during the encryption processing. As a result, several circuits containing measures against improper attacks (hereinafter referred to as countermeasure circuits) have been developed.

The present study proposes a new hardware Trojan in a countermeasure circuit and evaluates the weakness of the countermeasure circuit through simulation experiments. No studies have been performed on a hardware Trojan in a countermeasure circuit as proposed in the present study.

II. PROPOSED HARDWARE TROJAN

The present study uses threshold implementation (TI)[2] as a countermeasure circuit against power analysis attacks. TI has the highest resistance at present. TI of the advanced encryption standard (AES) is a measure against correlation power analysis (CPA). It uses random masks based on secret sharing and parallel computation. Figure 1 shows a circuit block diagram of TI. As shown in this figure, the circuit structure adopts a four-wire system for input. Using values input to TI and values output from a random number generator, TI is divided into four pieces and is unmasked at the XOR section in the final round.

In the present study, secret keys are not directly derived for the countermeasure circuit. Rather, a hardware Trojan, which enables power analysis attacks, is introduced by nullifying random masks. The value of the data register is forcedly reset to 0 immediately after encryption processing is completed. Random numbers created by the random number generator are forcedly changed to 0 as well. Since the plain text (PT) and the value of the reset register (the initial value of '0') are already known by an attacker, PT^K0 can be specified by estimating (trying) K0 in the 8th power of two ways (=2^8). Using the correlation between the hamming weight and power consumption, keys can be revealed.

In the proposed method, the conditions for the Trojan trigger to actuate a hardware Trojan can prevent a hardware Trojan from being found by persons other than the attacker by using two states as shown in Figure 3. In case where all the most significant eight bits of PT to be input are 1, the normal state is transited to the state wherein a hardware Trojan can be actuated. In addition, when the encryption processing is completed, the hardware Trojan is actuated. However, to actuate the hardware Trojan, PT must be in the Trojan state and the least significant bit of PT to be input must be 1.

Otherwise, the hardware Trojan will be unsustainable. Except for the above-mentioned conditions, the state, in which the hardware Trojan can be actuated, is transited to the normal state. Thus, in the proposed method, LSI acts based on the specifications (the encryption processing) provided in the functional test after the manufacturing process. In resistance evaluation against power analysis attacks (tamper resistance verification), the hardware Trojan is difficult to detect.

This work is supported by VDEC and CREST, JST.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

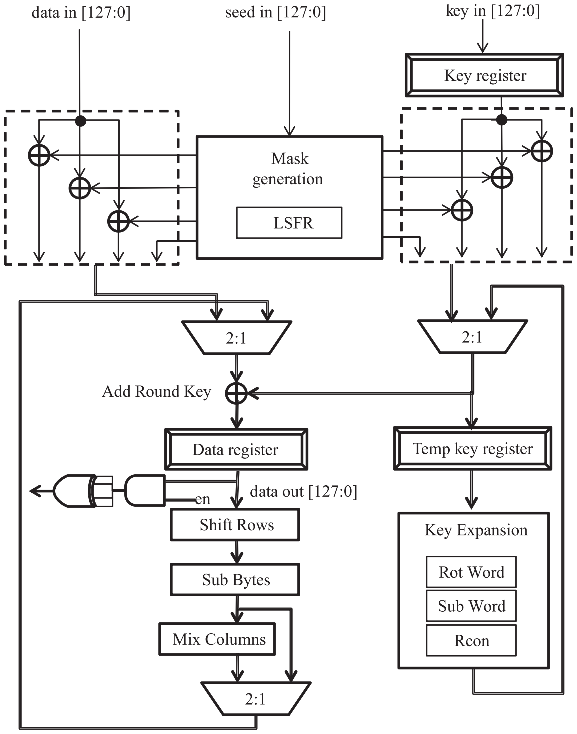


Fig.1Example of the threshold implementation

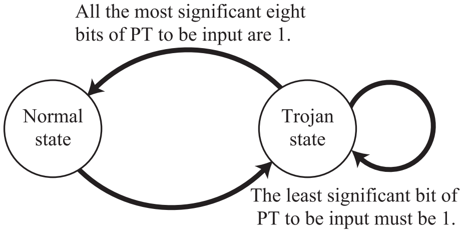

All the most significant eight bits of PT to be input are 1.
Normal state
Trojan state
The least significant bit of PT to be input must be 1.

Fig.2Two state for the Trojan trigger

III. EVALUATION EXPERIMENTS

A. Experimental Conditions

In order to evaluate the threat of the proposed hardware Trojan, experiments were performed. In the experiments, TI, into which the hardware Trojan was incorporated, was described using 0.18um CMOS Technology. Fig.3 shows the prototype chip for the hardware Trojan. Moreover, two resistance evaluations were performed as follows: (1) similar to the conventional tamper resistance verification, random PT was used; and (2) PT, by which a hardware Trojan was always actuated, was used.

B. Experimental Conditions

Fig.4 shows the results obtained by the evaluation experiments. In this figure, the horizontal axis represents the number of waveforms used in the experiments while the vertical axis represents the number of correct keys. Moreover, "PTp" represents the results obtained by performing CPA against random PT and "PTp_tro" represents the results obtained by performing CPA against PT, by which a hardware Trojan was always actuated. As shown in this figure, when random PT was used, the largest number of derived correct keys was 1. Therefore, resistance of TI, in which random PT was used, was similar to that of the TI reported in other papers[2][3]. When the hardware Trojan was actuated, all keys could be specified. Thus, when an attacker obtains a circuit into which a hardware Trojan proposed in the present study is incorporated, keys are easy to specify. Moreover, a hardware Trojan cannot be detected in conventional functional verification or tamper resistance verification.

IV. CONCULUSION

The present study proposed a hardware Trojan for countermeasure circuits against power analysis attacks and verified the threat of the proposed hardware Trojan through evaluation experiments. In the future, we will examine a method to detect a hardware Trojan.

. REFERENCES

[1] R.S.Chakraborty, S.Narasimhan, S.Bhunia, "Hardware Trojan: Threats and emerging solutions", Proc. of IEEE International High Level Design Validation and Test Workshop, pp.166-171, 2009.

[2] S.Nikova, C.Rechberger, V.Rijmen, "Threshold Implementations Against Side-Channel Attacks and Glitches", Proc. of ICICS 2006, LNCS 4307, pp.529-545, 2006.

[3] H.Mimura, T.Matsumoto, "Security Comparison among AES Cryptographic Circuits with Different Power Analysis Countermeasures", Proc. of SCIS 2011, 3D4-5, pp.1-8, 2011.

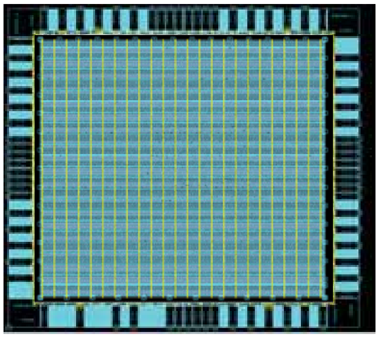


Fig.3 Prototype chip for the hardware Trojan

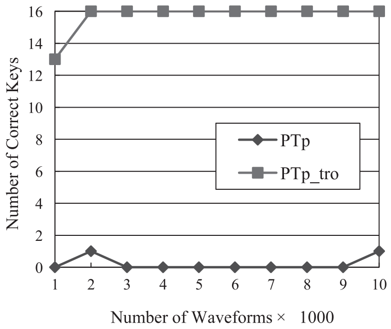


Fig.4 CPA results of normal AES circuit and that with Trojan circuit

Hybrid Immersive Audio Architecture Based on 3D Object Layer

Young Woo Lee, Sunmin Kim
Digital Media & Communications R&D Center, Samsung Electronics Co.Ltd., Korea

***Abstract*--This paper proposes the new 3D audio architecture which can give listeners an impression of 3D immersive audio, irrespective of various playback configurations of consumer electronics. The hybrid immersive audio architecture based on 3D Object Layer provides the high spatial fidelity of 3D effect while preserving the sound quality.**

I. INTRODUCTION

As 3D video industry is growing up, consumers would want to listen to spatially enhanced sound. The various multichannel surround systems are proposed to reproduce an immersive sense of presence and reality [1][2]. However, most of consumer electronics such as television and mobile device don't have enough channel configurations to reproduce the higher spatial audio quality. Therefore, it would be required to develop the new audio architecture which can faithfully reproduce the original sound scene, irrespective of reproduction systems. In this paper, we propose the new 3D audio architecture which can provide the functionality of distance control and format-independent rendering, and can maintain the backward compatibility.

II. CONVENTIONAL 3D AUDIO ARCHITECTURE

Parametric multi-channel audio coding such as MPEG Surround [3] was developed for the need of low-bit-rate transmission or storage. However, they are designed in a format specific way and can only be decoded as the same audio format. To overcome the limitation of channel-based coding, parametric scene-based audio architectures such as Directional Audio Coding [4] and Spatial Audio Scene Coding [5] have been discussed. Because those approaches can't separate the audio objects, the spatial fidelity of the audio reproduction may be limited. To provide the flexibility at the reproduction side, MPEG Spatial Audio Coding (SAOC) [6] is developed. SAOC provides the rendering of multiple audio object signals and the format independence. However, the backward-compatibility is limited to stereo reproduction, therefore not suitable for multichannel audio format. Furthermore, the quality of ambience is limited by transmitting a limited number of reverb objects. Recently, hybrid spatial audio coding architecture was proposed by Jot, et al [7]. Selected object audio signals are included in multi-channel base mix signals and the object mix cues of those and encoded object signals are transmitted to separate the object signals at the decoder.

III. PROPOSED 3D AUDIO ARCHITECTURE

A. *3D Object Layer*

Our basic idea of the proposed hybrid audio architecture is to get the mixing information of selected objects that mixing engineers apply the 3D control during the contents creation. The most important thing for this is simply applicable architecture without replacement of existing professional audio mixing console or digital audio workstation (DAW) system. We introduce the 3D Object Layer (3DOL) to separate the object to apply the 3D control such as distance, elevation, and so on, as shown in Fig. 1. 3DOL is used to mix the 3D objects only, while the other objects such as instruments and ambiences except the 3D objects are mixed to Target Layer (TL). At the mixing side, there is no need to change the mixing technology because the speaker assignment between 3DOL and TL is exactly same. This architecture is simply applicable to existing systems using plug-in software that mixing engineers can select the 3D object among the audio tracks. Moreover, because the Transmission Channels are combination of TL and 3DOL signals, this architecture is backward-compatible to legacy format. The hybrid metadata is obtained by scene analysis of 3DOL and transmitted to faithfully localize the 3D object at the various playback systems.

B. *3D audio scene analysis*

Fig. 2 represents the proposed 3D audio scene analysis architecture consisting of 3D Object Cues Estimation, Position Estimation, Parameter Encoder and Conventional Audio Encoder module. The 3D object cues are defined by the spatial parameters between downmix signals of 3DOL and Transmission Channel. The format-independent position metadata of 3D object can be expressed as azimuth, elevation angle, and distance or depth index which is relative distance.

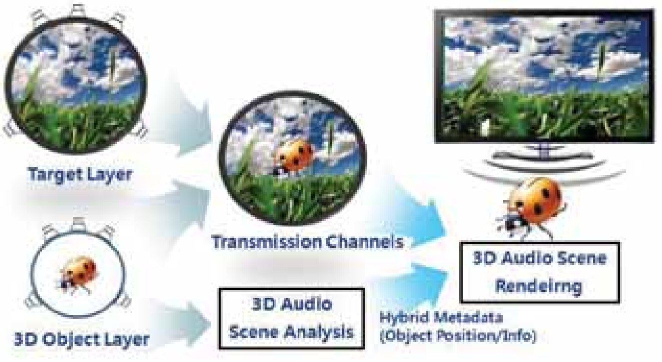


Fig. 1. 3D Object Layer concept for hybrid audio architecture.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

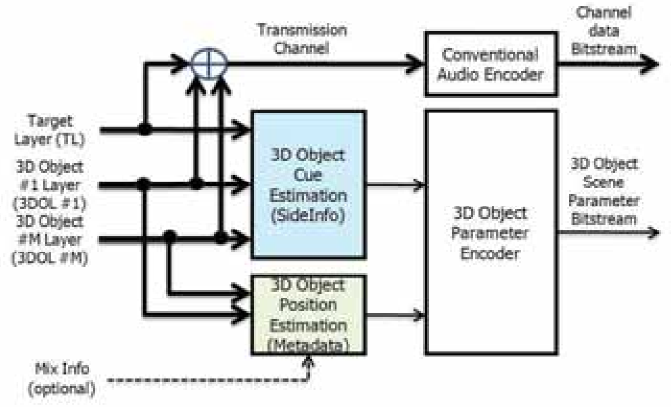

Transmission Channel
Channel data Bitstream
Conventional Audio Encoder
Target Layer (TL)
3D Object #1 Layer (3DOL #1)
3D Object #M Layer (3DOL #M)
3D Object Cue Estimation (SideInfo)
3D Object Parameter Encoder
3D Object Scene Parameter Bitstream
3D Object Position Estimation (Metadata)
Mix Info (optional)

Fig. 2. Block diagram of 3D audio scene analysis

Optionally, the divergence parameter of 3D object is defined as estimation of the energy distribution to target layout. These 3D object scene parameters are encoded and transmitted to 3D audio scene synthesis architecture.

C. 3D audio scene synthesis

As shown in Fig 3, the proposed 3D audio scene synthesis architecture receives the encoded channel data bitstream and 3D object scene parameter bitstream. TL-3DOL Separation module extracts the 3D object audio signal from decoded channel data using decoded 3D object cues. Target Layer which is residual signals are processed by the Layout Free Rendering module for reproduction in the target audio format. Separated 3D object audio signals are rendered by 3D Object Effect Rendering module using decoded 3D object position metadata. To provide the high spatial fidelity at various speaker configurations, the virtual audio rendering method suitable to the playback system is applied [8][9]. At the legacy decoder, the 3D object parameter is discarded or ignored, and therefore the backward compatibility is provided.

IV. DISCUSSIONS

The most important thing for new 3D Audio format is to reproduce the audio scenery faithfully with accurate sound localization and sound envelopment at any given loudspeaker configurations. Backward compatibility and new functionality such as interactivity are also important. The decoded contents based on the proposed architecture can provide the high sound quality which is the advantage of channel-based approach, because the transmitted signals are channel signals that the 3D object is mixed. Although Scene-based approach can reproduce the audio scenery in different channel configurations, it can estimate the direct-like signal instead of object signal by blind separation from channel signals. To provide the new functionality which is the advantage of Object-based approach, the proposed architecture can estimate the 3D object parameters consisting of object cue and position data. The approach proposed by Jot, et al has to transmit the encoded object signals to separate the object signals at decoder. Also, one object decoder should be added at the decoder side and one object encoder-decoder pair should be added at the encoder side. On the other hands, our proposed architecture doesn’t encode and transmit the object signals. Therefore, it’s simply applicable to existing mixing systems using plug-in software. Also, transmission of object cue instead of object signal enables the low-bit-rate transmission and there is no need to add the object decoder to separate the object signal from the transmitted data stream.

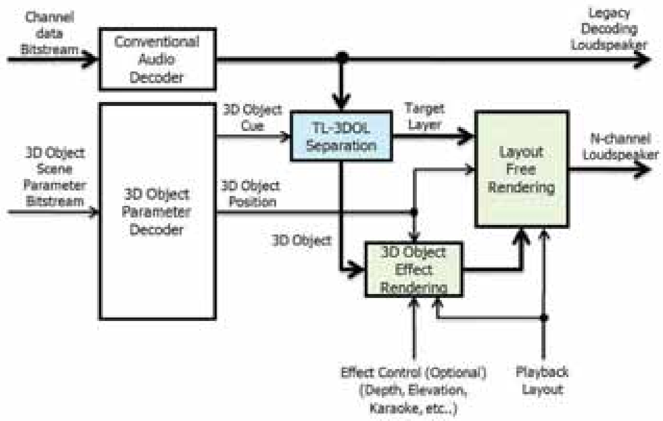

Channel data Bitstream
Conventional Audio Decoder
Legacy Decoding Loudspeaker
3D Object Cue
TL-3DOL Separation
Target Layer
Layout Free Rendering
N-channel Loudspeaker
3D Object Scene Parameter Bitstream
3D Object Parameter Decoder
3D Object Position
3D Object
3D Object Effect Rendering
Effect Control (Optional) (Depth, Elevation, Karaoke, etc..)
Playback Layout

Fig. 3. Block diagram of 3D audio scene synthesis

V. CONCLUSIONS

We proposed the new hybrid immersive audio architecture based on 3D Object Layer which is simply applicable to existing mixing systems. This architecture has the advantages in channel-based approach which enables the spatial reproduction with perceptually high quality, as well as maintaining the format-independence and the backward-compatibility. Also, it supports the interactivity of 3D object which is a merit of object-based using effect control. Therefore, this architecture can overcome the limitation of conventional approaches and provide the new 3D immersive audio format.

REFERENCE

[1] K. Hamasaki et al., “The 22.2 Multichannel Sound System and Its Appication”, 118th AES Convention, 2005.

[2] S. Kim, Y.W. Lee and V. Pulki, “New 10.2-channel vertical surround system (10.2-VSS); comparison study of perceived audio quality in various multichannel sound systems with height loudspeakers”, 129th AES Convention, 2010.

[3] J.Herre et at., “MPEG Surround – The ISO/MPEG Standard for Efficient and Compatible Mutlichannel Audio Coding”, J. Audio Eng. Soc. 56(11):932-955, 2008.

[4] Ville Pulkki, “Spatial Sound Reproduction with Directional Audio Coding”, J. Audio Eng. Soc. 55(6):503-516, 2007.

[5] M. Goodwin. J.-M. Jot, “Spatial Audio Scene Coding”, 125th AES Convention, 2008

[6] O. Hellmuth et al., “MPEG Spatial Audio Object Coding – The ISO/MPEG Standard for Efficient Coding of Interactive Audio Scenes”, 129th AES Convention, 2010.

[7] J.-M. Jot, Z. Fejzo, “Beyond Surround Sound – Creation, Coding and Reproduction of 3-D Audio Soundtracks”, 131th AES Convention, 2011.

[8] S. Kim, Y.W. Lee, “3D Audio Depth Rendering for Enhancing an Immersion of 3DTV”, 131th AES Convention, 2011.

[9] Y.W. Lee et al, “Virtual Height Speaker Rendering for Samsung 10.2-channel Vertical Surround System”, 131th AES Convention, 2011.

Integration of Face Recognition and Sound Localization for a Smart Door Phone System

Taewan Kim, Hyungsoo Park, and Yunmo Chung

Abstract--**This paper proposes a smart system using both face recognition and sound localization techniques to identify the faces of visitors from a door phone in much efficient and accurate ways. This system is effectively used to recognize the faces when their locations are out of the boundaries of the camera scope of the door phone. The smart door phone system proposed in this paper uses a visitor's voice source to rotate the camera to his face and then to recognize the face accurately. The integrated system has been designed with one FPGA(Field Programmable Gate Array) chip and tested for actual use in door phone environments.**

I. INTRODUCTION

With the importance of life safety and convenience, in recent, the use of door phones has been increasing along with the advent of security-related electronics, such as digital door locks, advanced video conversation devices, and wireless home security networks [1]. In general, a fixed camera has been used to identify visitors from traditional door phone systems. In this case, if their faces are located out of the boundaries of the camera scope, there may be a difficulty in recognizing the faces regardless of whether a good face recognition technique is used or not. As an example, we can consider the case that the faces are either lower or higher than the position of the camera.

To cope with this difficulty, if we can provide a way to know the location of a face in advance and then the door phone camera can be rotated to the front of the face, we can increase the success rate of identifying one's face. In this paper, we propose a smart door phone system by integrating both 4-channel microphone sound localization and face recognition techniques. To recognize a visitor's face, this system rotates the camera to the face by the sound localization technique based on the voice source from the visitor. And then the recognized visitor's information is displayed on the door phone screen.

The proposed system has been implemented with an FPGA chip and can be used in the near future for cell phones and smart phones if the rotation technique of a camera is well considered.

II. PROPOSED SYSTEM

Fig. 1 shows an outline of the system proposed in this paper. If a visitor's face is out of the boundaries of camera's scope even though somebody exists in front of the door phone, a host usually says "who is this?". Assuming that the visitor answers "It's me", the sound localization technique finds where the voice of the visitor comes from and the camera is automatically rotated to the direction of the visitor's face. In addition, the smart door phone system recognizes the visitor's face and then displays his information on the screen.

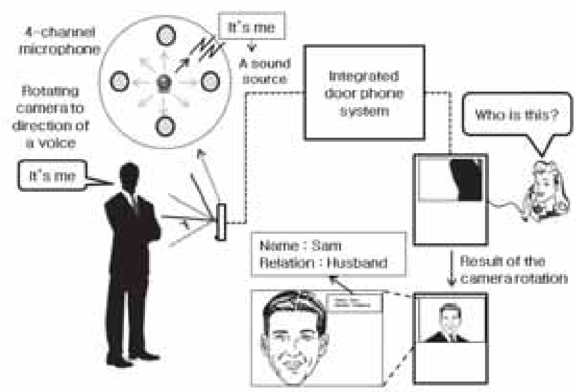

4-channel microphone
It's me
A sound source
Rotating camera to direction of a voice
Integrated door phone system
Who is this?
It's me
Name : Sam
Relation : Husband
Result of the camera rotation

Fig. 1. Proposed system outline

III. SOUND LOCALIZATION TECHNIQUE

The sound localization technique considered in this paper uses a 4-channel cross-typed microphone array as shown in Fig. 2 to detect a voice source location in the 360º range using the GCC(Generalize Cross Correlation) algorithm [2]-[3]. The technique has been modeled in Verilog HDL for hardware design as shown in Fig 3. A sound source sensed by each microphone is sent to GCC block, which estimates TDOA(Time Delay of Arrival) from microphones. And then TDOA data goes into GCC result table map block to determine the location of the source [4].

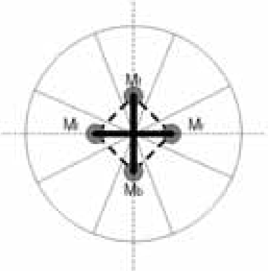


Fig. 2. 4-channel microphone array structure for front, back, left, and right microphones

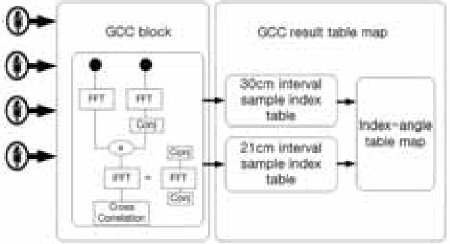


Fig. 3. 4-channel sound localization operation block diagram

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

IV. Face Recognition Technique

The algorithm converts the input images from a camera to LBP(Local Binary Pattern) images and then registers the face location candidate using the AdaBoost algorithm as shown in Fig. 4. The final location with highest reliability is determined among the registered face candidates [5]-[6]. Fig. 5 shows the performance results of the face detection algorithm with three people. Fig. 6 shows the recognition result based on the algorithm. The detected face feature is compared with features in the pre-stored database and the phone system shows the face information on the screen from the database.

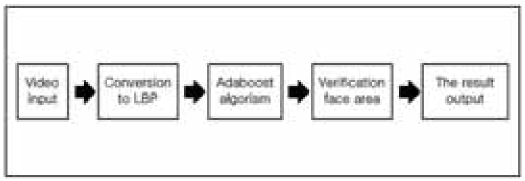
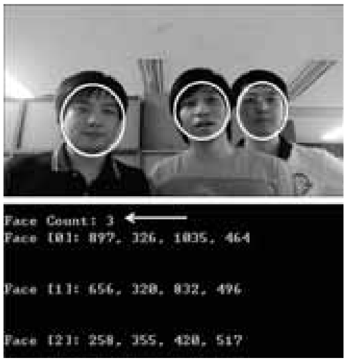


Fig. 4. Face detection algorithm

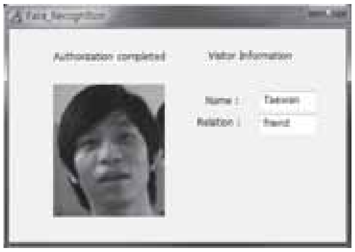


```
Face Count: 3
Face [0]: 897, 326, 1035, 464

Face [1]: 656, 320, 832, 496

Face [2]: 258, 355, 420, 517
```

Fig. 5. Face detection result



Authorization completed

Visitor Information

Name : Taewan

Relation : friend

Fig. 6. Face recognition result

V. Smart Door Phone System

In this paper, both sound localization and face recognition techniques have been designed and implemented for a smart door phone, which can give a great performance by recognizing the faces of visitors efficiently and accurately. Fig. 7 is the block diagram of an integrated smart door phone block diagram with one FPGA chip, in which the sound localization technique is designed by using a hardware IP(Intellectual Property) while the face detection and recognition technique is implemented by a software approach with a MCU(Micro controller Unit).

Fig. 8 shows the system implementation with one FPGA chip which includes a camera module, a motor module, a cross-typed microphone array, an FPGA board with LCD screen.

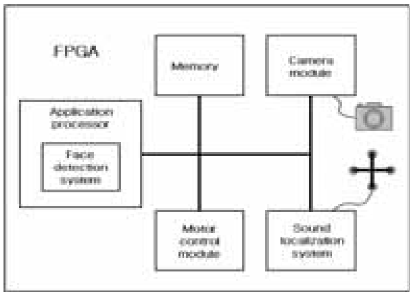


Fig. 7. Smart door phone integration block diagram

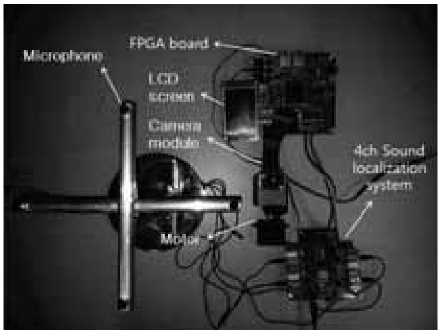


Fig. 8. Implementation of the proposed smart door phone system

VI. Conclusion

In this paper we are proposing a smart system to rotate the camera to the direction of a sound source to improve the face recognition rate in the traditional door phone system with a fixed camera. The system has been implemented in the integration of both sound localization and face recognition techniques with one FPGA chip to reduce power consumption and system size. According to the performance result, the system increases the success rate of face recognition and its convenience. Finally, the proposed system could be used for smart phones in the near future if we consider minimization and low power problems.

Reference

[1] Ching-Lung Chang and Han-Yu Tsai, “The design of Video Door Phoneand Control System for Home Secure Applications,” IEEE International Conference. Innovative Mobile and Internet Services in Ubiquitous Computing, Vol. 5, pp. 1-5, 2011

[2] Kai-Tai Song and Jian-Liang Chen, "Sound direction recognition using a condenser microphone array," IEEE International Symposium. Computational Intelligence in Robotics and Automation, Vol. 3, pp. 1445, Istanbul TURKEY, July 2003.

[3] Jindong Chen, Jacob Benesty, and Yiteng Huang, "Time Delay Estimation in Room Acoustic Environments: An Overview," EURASIP J. Appl. Signal Processing, Vol. 2006, pp. 19, 2006.

[4] Charles H. Knapp, "The Generalized Correlation Method for Estimation of Time Delay," IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. ASSP-24, No. 4, pp. 320-327, 1976.

[5] Y. Freund, and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", Journal of Computer and System Sciences, 1997, Vol. 55, pp. 119-139.

[6] Maenpaa and Topi, "Local binary pattern approach to texture analysis-extensions and applications", Univ. Oulu, 2003.

Acoustic Signal Based Abnormal Event Detection System with Multiclass Adaboost

Younghyun Lee, Hanseok Ko: Korea University, Seoul, Korea
David K. Han: Office of Naval Research, Arlington, VA, USA

***Abstract*—This paper addresses the problem of abnormal acoustic event detection in indoor security systems. We propose a multiclass Adaboost based acoustic context classifier for performance and speed improvements over the conventional and prominent GMM based classifiers.**

I. Introduction

For the last decade, there has been a slew of Intelligent Surveillance Systems (ISS) developed in the consumer market due to availability of inexpensive and compact sensors. Of all the modalities used by these sensors, acoustics based sensors offer certain advantages such as their ability of detecting events beyond 'Line of Sight' and the low cost. A particular set of problems within the acoustic sensor based ISS is classification of abnormal situations related to child safety. For enhanced security of their child, parents would be very much interested in automatically be informed by an ISS when their child is either crying or screaming. It is desirable for the ISS to be sensitive in accurately classifying those abnormal events while it exhibits a negligibly low false alarm rate. Conventional acoustic based systems have adopted Gaussian Mixture Model (GMM) based classifiers [2]-[3]. However, applying GMM for acoustic based classification may lead to high false classification rates. Additionally, GMM may not be the most suitable tool when near real-time processing is required due to its computational complexity for large dimensionality of the feature vector. Recently, the Adaboost method has become one of the most popular and effective classification tools in computer vision and pattern recognition due to its high-speed/low-complexity [4]. To overcome the shortcomings of the GMM, we propose to apply a multiclass Adaboost based method to develop an acoustic context classifier for a child safety application.

II. Abnormal Event Detection System

The proposed abnormal event detector consists of two steps. In the first step, acoustic features are extracted from the acoustic signals and they are classified into pre-defined context classes. In the second step, an abnormal event detector determines the acoustic state as either normal or abnormal from the accumulated context classes produced at the first step for a fixed period of time. The term 'Context' stands for the minimal length of acoustic signals for a person to recognize the event. 'Crying', 'Scream', 'Conversation', 'Empty', or 'Bell' are some examples of the context classes of interest. The term 'Event' stands for certain occurrence of interest that may occur during some time interval. Therefore, a combination of various contexts may compose a single event.

TABLE I
Adaboost algorithm for multiclass classification

```
Input: training samples $(\boldsymbol{x}_i, y_i)$, $y_i \in \{1, \cdots, C\}$, $i = 1, \dots, N$.
       maximum training iteration $T$.
Output: $\mathbf{F}(\mathbf{x}) = [\alpha_1 f_1(\boldsymbol{x}), \cdots, \alpha_T f_T(\boldsymbol{x})]^{\mathrm{T}}$, coding matrix $\mathbf{M} \in \{\pm 1\}^{C \times T}$.
Initialization
  The coding matrix $\mathbf{M} = []$, and the weight distribution $u_{i,c} = \frac{I(c \neq y_i)}{N(C-1)}$,
  $i = 1, \dots, N$, $c = 1, \dots, C$.
For $\boldsymbol{t} = \mathbf{1}{:}\boldsymbol{T}$ do
  a) Create $M(:, t) \in \{-1, +1\}^{C \times 1}$;
  b) Compute distribution for mislabels $\pi_i = \sum_{c=1}^{C} u_{i,c} I(M(c,t) \neq M(y_i, t))$;
  c) Normalize $\boldsymbol{\pi}$;
  d) Train a weak classifier $f_t(\boldsymbol{x})$ using $\boldsymbol{\pi}$;
  e) Compute $\epsilon = \sum_{i=1}^{N} \pi_i I\left(M(y_i, t) \neq f_t(\boldsymbol{x}_i)\right)$;
  f) Compute $\alpha_t = \frac{1}{4} \ln\left(\frac{1-\epsilon}{\epsilon}\right)$;
  g) Update $u_{i,c} = u_{i,c} \exp\left(-\alpha_t (M(y_i, t) - M(\mathrm{c}, t)) f_t(\boldsymbol{x}_i)\right)$;
  h) Normalize $\mathbf{u}$
```

A. Acoustic Context Classification

The acoustic signals are recorded at a sampling rate of 8 kHz using 16-bit quantization. We use 40-dim Mel-Frequency Cepstral Coefficients (MFCC) features with frame length of 50ms from an acoustic signal [2]-[3]. Acoustic contexts are classified into either normal or abnormal events. The normal events consist of 'Alarm bell', 'Door opening/closing', 'Announcement', 'Conversation', and 'Empty'. The abnormal events consist of 'Crying', 'Infant Scream', and 'Adult Scream'. In this paper, a multiclass Adaboost based classifier is employed for context classification. The multiclass problem can be reduced to multiple binary subproblems by using a coding strategy translating each label to a fixed binary string referred to as a codeword as in [5]. Then, weak classifiers can be trained at every bit position. The Adaboost based learning algorithm implemented here is summarized in Table 1. Training samples are given by $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$. $\boldsymbol{x}_i$ is a vector valued feature and y_i is a label from $\{1, \cdots, C\}$ for C classes. To reduce a multiclass problem into several binary subproblems, a coding matrix $\mathbf{M} \in \{\pm 1\}^{C \times T}$ is required. The c^{th} row of $\mathbf{M}$, $M(c, :)$, represents a T-length codeword for the class c. At each iteration, a randomly generated code is added into the matrix, which corresponds to a new binary subproblem being created. A decision stump as a binary weak classifier is trained for each column, where training samples have been relabeled into two classes. For a test sample $\mathbf{x}$, a classifier $\mathbf{F}(\mathbf{x})$ produces an unknown T-length codeword. Decision of the classifier $D(\mathbf{x})$ is as shown in (1). The "closest" row is identified as the predicted label.

$$D(\mathbf{x}) = \arg\max_{c} \left\{ \sum_{t=1}^{T} \alpha_t M(c,t) f_t(\mathbf{x}) \right\} \cdot \tag{1}$$

This research was supported by Seoul R&BD Program (WR080951)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Determination of Abnormal Events

Abnormal events are determined based on the results of the acoustic context classification. The reference frame is collected by accumulating results of the acoustic context classification for 1 second, and a histogram of the context is built from the reference frame. Classification of the reference frame R is by the maximum valued bin of context histogram H_{cont}, which would result 1 if a reference frame is abnormal, and 0 otherwise as in (2).

$$R = \begin{cases} 1, & \arg\max_i H_{cont}(i) \in \text{abnormal} \\ 0, & \text{otherwise} \end{cases} . \tag{2}$$

The abnormal event determination rule is defined simply based on the ratio of the number of abnormal frames as in (3)

$$D_n = \begin{cases} 1, & \text{if } \dfrac{1}{K} \displaystyle\sum_{k=n-(K-1)}^{n} R_k > T_D \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where, D_n is the final decision, which is equal to 1 if an abnormal event is detected and 0 otherwise at the n^{th} reference frame. T_D denotes the sensitivity parameter ranging from 0 to 1, meaning the threshold of the ratio of the number of abnormal frames to the total number of reference frames K.

III. Experimental Results

To train the acoustic context classifier and to test the performance of our proposed algorithm, we collected the acoustic signals from four different types of indoor space: three different elevator interiors and an office space. Fig. 1 shows the comparison of the classification methods by adjusting the parameter, T_D. Generally, the proposed method shows better performance than the GMM based methods proposed by [3]. Key advantages of the proposed method are the improvement in detection speed and low false alarm. As the dimensionality of feature space d increases, the numerical complexity of the GMM based classifier quickly becomes too expensive due to its inherent inverse matrix operations. However, the execution time of the Adaboost based classifier is independent of d. Table 2 shows the execution time of detecting an event from an acoustic record of 68 seconds divided into 2719 frames with the frame length of 50ms. The experiments were conducted by using a PC on a MATLAB platform with a 2.8 GHz CPU and 6GB RAM. For ensuring consistency of the result, the experiments were repeated ten times and the average total execution time was calculated. Reference [3] reported that the GMM classifier based abnormal detector can process test frames for acoustic signals in real time. However, for applications that require converged classifications of acoustic and visual signals as in [6], faster processing method as the one proposed here is more suitable.

IV. Conclusions

In this paper, we proposed and implemented an abnormal event detection system based on acoustic signals for indoor security applications. At first, MFCC acoustic features from a segment of acoustic signals were classified by a multiclass Adaboost based classifier into a normal event or an abnormal one. Then, through an accumulation process of these context classes for a fixed period of time, the acoustic event was finally determined to be either normal or abnormal. Through experimental results, the proposed algorithm was shown to be sufficiently accurate with minimal false alarm rate. Due to the expedient processing capability, the method is ideal for real-time processing applications such as commercial security/ surveillance systems.

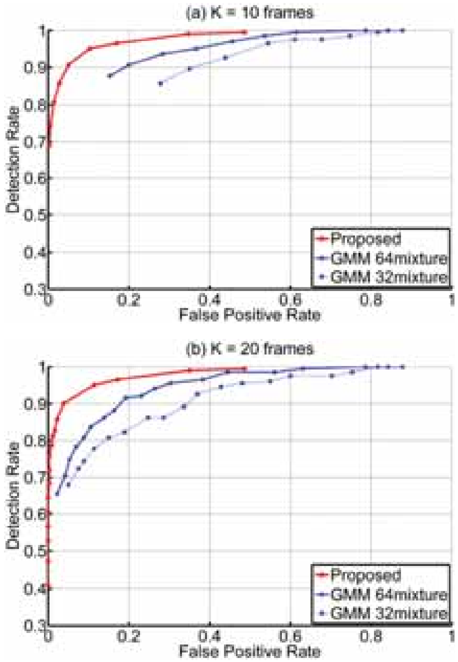


Fig. 1. Comparison of the performance of the classification methods.

TABLE II
EXECUTION TIME OF ABNORMAL EVENT DETECTION

	GMM (32 mix)	GMM (64 mix)	Proposed
Speed (s)	58.69	116.71	1.55

References

[1] J. L. Castro, M. Delgado, J. Medina, and M. D. Ruiz-Lozano, "Intelligent surveillance system with integration of heterogeneous information for intrusion detection", *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11182-11192, September 2011.

[2] W. Choi, S. Kim, M. Keum, D.K. Han and H. Ko, "Acoustic and visual signal based context awareness system for mobile application", *IEEE Trans. Consum. Electron.*, vol. 57, no. 2, pp. 738-746, May 2011.

[3] K. Kim and H. Ko, "Hierarchical approach for abnormal acoustic event classification in an elevator", in *IEEE Int. Conf. on Adv. Video and Signal Based Surveillance*, pp.89-94, 2011

[4] P. Viola and M. J. Jones, "Robust Real-Time Face Detection", *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137-154, 2004

[5] V. Guruswami and A. Sahai, "Multiclass Learning, Boosting, and Error-Correcting Codes", in *Proc. Annual Conf. Learn. Theory*, pp. 145-155, 1999

[6] Y. Lee, K. Kim, D. K. Han and H. Ko, "Acoustic and Visual Signal Based Violence Detection System for Indoor Security Application", in *IEEE Int. Conf. on Consum. Electron*., pp. 743-744 , 2012

3D Sound Rendering System Based on Relationship between Stereoscopic Image and Stereo Sound for 3DTV

Sunmin Kim, Young Woo Lee, and Yoon Jae Lee
DMC R&D Center, Samsung Electronics, Korea (sunmin21.kim@samsung.com)

Abstract— **This paper presents a new 3D sound rendering system based on hybrid 3D index estimation method. Spatial information of stereoscopic images and stereo sound are compared. When the spatial characteristics of video and audio object are matched, the rendering method generates the 3D sound effects which are stage expansion, distance control, and elevation rendering.**

I. Introduction

3D sound rendering method for 3DTV has received a great deal of attention in consumer electronics area as 3D video contents become popular. In general a 5.1-channel sound can be played back using a home theater system with multichannel loudspeakers. A listener who does not own a home theater system but has a 3DTV might still want to enjoy 3D sound using the two-channel TV loudspeakers. Various 3D sound rendering methods [1-6] have been developed to give the listener a 3D sound effect. Recently, a 3D depth rendering method based on depth index estimation using disparity information of stereoscopic images has been proposed by Kim [6], which can make a sound source close to the listener with respect to a distance of 3D video object.

In this paper, a new 3D sound rendering system, which consists of a hybrid 3D index estimation method and the index-based 3D sound rendering method, is proposed for 3DTV. The hybrid 3D index estimation method uses x-y position information of 3D video object as well as disparity value, and spatial characteristics of stereo sound are analyzed at the same time. The information obtained from analysis of both 3D video and stereo sound is used to estimate panning/depth/height indices which represent a 3D position of sound object and the three indices are used to control a 3D position of sound object matched with that of 3D video object. Stereo sound stage, distance, and elevation perception of sound object are realized based on the panning/depth/height indices, respectively.

II. Hybrid 3D Index Estimation

Sometimes, video object and audio object in a movie are not relevant to each other. Therefore, the 3D depth rendering method should use the information of video and audio, together. In order to get a matching score of video and audio object, positions of the video and audio object are compared.

A. *Disparity of Stereoscopic Images*

A disparity map can be obtained from the stereo image by comparing the difference of left and right images within a certain search range. Refer to [6,7] for details. In addition, the position index having maximum disparity among 9 zones is transmitted to audio processor (see Fig. 1). The position can be selected with multiple zones. The transmitted position of video object is compared to position of audio object obtained from spatial characteristics of stereo sound.

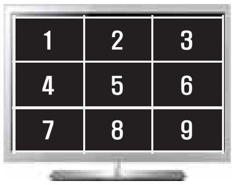


Fig. 1 Nine zones of position having maximum disparity.

B. *Panning Index*

A panning angle of audio object can be calculated from a power ratio of left and right signals, and the panning angle is compared to the horizontal position of video object having the maximum disparity. When the positions of video and audio objects are matched, the panning index value which ranges from 0 to 1 becomes 1.

C. *Depth Index*

In [6], the depth index is calculated without comparison of video and audio information. In case of 2D images, the depth index is estimated from only audio information. In case of 3D images, the depth index is obtained from only video information. In this paper, the depth index is estimated using both video and audio information. The hybrid depth index can be a meaningful value when the two depth indices are matched.

D. *Height Index*

In [6], there was no elevation rendering. In order to implement the elevation rendering, height index is obtained from the vertical position of the video object. The height index is dependent on only position of video object because the vertical position of audio object cannot be calculated from the stereo sound. Figure 2 shows a block diagram of hybrid 3D index estimation method which calculates three indices.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

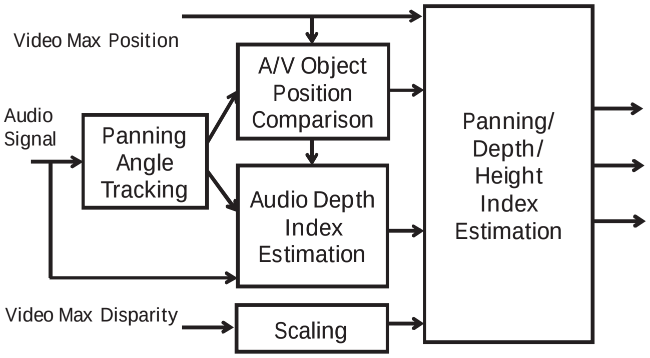

Video Max Position
Audio Signal
Panning Angle Tracking
A/V Object Position Comparison
Audio Depth Index Estimation
Video Max Disparity
Scaling
Panning/ Depth/ Height Index Estimation

Fig. 2. Block diagram of hybrid 3D index estimation.

III. 3D Sound Rendering Method

With the three indices, various 3D rendering methods can be realized: wide stereo, distance control, and elevation rendering.

A. *Dynamic Panning Method*

Sound stage is adaptively widened with respect to the panning index. Wide stereo algorithm [4] can be used for stage expansion. A dynamic panning method based on the panning index makes the sound stage more widened and dynamic. When audio object is panned in center, the virtual position of the audio object is still in center. However, the virtual position of the audio object is more shifted to left side, when the audio object is panned to the left side.

B. *Distance Control Method*

Distance control method [6] makes the audio object being close to a listener with respect to the hybrid depth index. The distance control method can yield the meaningful result even when the video and audio objects have no relationship.

C. *Elevation Rendering Method*

With respect to the height index, vertical position of audio object is controlled. The audio signals are convolved with a common head-related transfer function at 40 degrees of elevation. Refer to [8] for details of the common head-related transfer function which is calculated from an averaged spectral characteristics of many individual head-related transfer functions.

Figure 3 shows a block diagram of the 3D sound rendering system. The proposed system can provide various 3D sound effects.

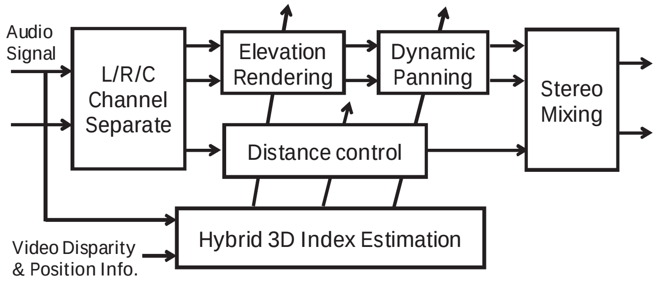


Fig. 3. Block diagram of 3D sound rendering method.

IV. Subjective Evaluation

Two kinds of listening tests were conducted for virtual elevation rendering effects. First evaluation term was sound quality, and second one was perceived angle. Ten subjects participated in normal listening room. Table 1 shows the results of experiment 1. Mean opinion score (MOS) was obtained with ranges from 1 to 5. MOS 5 mean that the sound quality of the processed sound is same as that of original sound. Two kinds of test materials were used: movie and music clip. The averaged score was 4.59. Table II shows the results of experiment 2. Helicopter sound clip was used for test materials. In order to accurately obtain a perceived angle for the subject, the angle was written on a measure, and the measure was vertically attached to TV. The averaged perceived angle was 39.5 degrees.

TABLE I
Test Results of Sound Quality Evaluation.

Test Program	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	#10	AVG
Movie (MOS)	5	4	4.2	4.5	4.5	4	4.2	4.3	4.5	4.5	4.59
Music (MOS)	5	5	4.0	4.5	4	4.5	4.8	4.9	4.8	4.5	

TABLE II
Test Results of Perceived Angle Evaluation.

Test Program	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	#10	AVG
Helicopter (Angle)	25	45	55	55	55	35	40	25	40	20	39.5

V. Conclusion

In this paper, 3D sound rendering system based on hybrid 3D index estimation method is proposed for 3DTV. The hybrid 3D index estimation method uses both information of video and audio object in order to consider the relation between stereoscopic images and stereo sound. The proposed 3D sound rendering system can provide three kinds of effects: wide stereo, distance control, and elevation rendering.

References

[1] Begault, D. R., 3-D sound for virtual reality and multimedia, pp 158-163, Academic Press, Inc., London (1994).

[2] A. I. Klayman, “Audio Enhancement System for Use in a Surround Sound Environment,” US5970152, SRS Labs, Inc., Irvine, CA (1999).

[3] D. S. McGrath, A. R. McKeag, G. N. Dickins, R. J. Cartwright, and A. P. Reilly, “Audio Signal Processing Method and Apparatus,” US6741706 B1, Lake Technology Limited, Ultimo, Australia (2004).

[4] Sunmin Kim, et al., “Virtual Sound Algorithm for Wide Stereo Sound Stage”, presented at 117th AES Convention, San Francisco, USA (2004).

[5] Sunmin Kim, et al., “Adaptive Virtual Surround Sound Rendering System for an Arbitrary Listening Position", Journal of Audio Engineering Society, Vol. 56, No. 4, pp 243-254 (Apr., 2008).

[6] Sunmin Kim, et al., “3D Audio Depth Rendering for Enhancing an Immersion of 3DTV”, presented at 131st AES Convention, New York, USA (2011).

[7] SCHARSTEIN, D., AND SZELISKI., R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. Journal of Comp. Vision 47, 1/2/3, 7–42.

[8] Young Woo Lee, et al., “Virtual Height Speaker Rendering for Samsung 10.2-channel Vertical Surround System”, presented at 131st AES Convention, New York, USA (2011).

A General Search Strategy for Multiple Reference Frame Motion Estimation

Chun-Su Park
Dep. of Information & Telecommunications Engineering, Sangmyung University, Cheonan, Korea
Email: cspark@smu.ac.kr

Abstract — Multiple reference frame motion estimation (MRFME) is one of main features of recent video coding standards such as H.264/AVC and HEVC. In this paper, we present a general search strategy for MRFME. The proposed method can accelerate the conventional full search algorithms without any penalty in the rate-distortion (RD) performance. Experimental results show that the proposed method can reduce the motion estimation (ME) time of the conventional algorithms up to 39.48%.

I. INTRODUCTION

With the generalization of mobile communication services, user-created video content attracts more attention since it enables users to share the personal experience with their friends and acquaintances. The captured video is compressed in a portable device and then the compressed video stream is transmitted over networks with limited bandwidth. A compression scheme, which can provide high quality video with low complexity, plays a key role in this application. Block-based motion estimation (ME) utilizing temporal redundancy between adjacent frames is one of the most significant features of state-of-the-art video coding standards such as H.264/AVC and HEVC. However, it is well-known that the ME process is the most time-consuming part of the video encoder.

Several fast search algorithms have been proposed to accelerate the ME process without sacrificing the coding efficiency [1], [2]. These methods establish inequality constraints and examine only the pixels satisfying the predefined inequality constraints. Basically, these methods are designed for single reference frame ME (SRFME). In multiple reference frame ME (MRFME), the block matching process is conducted using additional reference frames, thereby obtaining better prediction signal as compared to SRFME. The SRFME algorithms can be directly applied to the each reference frame, i.e., all reference frames are searched repeatedly. In this case, the complexity of the encoder increases significantly proportional to the number of searched frames.

We propose an effective search strategy for accelerating MRFME. We first analyze the available information after ME of the previous frames. Then we reduce the number of pixels to be examined by increasing the tightness of the inequality constraints of the conventional SRFME algorithms. In this paper, we focus on the spiral search of the H.264/AVC reference software [3]. However, the proposed strategy can be applied to most conventional full-search-equivalent ME algorithms [4].

II. PROPOSED ALGORITHM

The problem of finding the optimal MV for a given rate constraint can be formulated as finding the best point on the convex hull of all possible rate-distortion (RD) points. Lagrangian optimization is widely utilized for solving the problem. Let s and $c(m)$ be, respectively, the original block and its reconstruction obtained by using the MV m. For an inter-coded block, ME is performed to find the optimal MV by minimizing

$$J(m\,|\,s,\hat{m}) = D(s,c(m)) + \lambda \cdot R(m-\hat{m}) \qquad (1)$$

where $\hat{m}$ is the prediction for the MV and λ is the Lagrange multiplier. The distortion $D(s,c(m))$ is the difference between s and $c(m)$. The rate $R(m-\hat{m})$ specifies the number of bits required for encoding the difference between the original MV and the predicted one.

Let Ψ be a search window of size W and $m_k \in \Psi$, $0<k\le W$, be an MV indicating a possible pixel position in a spiral order. Then, Ψ is composed of the pixels within the search window, which can be represented as $\Psi=\{m_1,m_2,...,m_W\}$. In the ME process, after examining the sub-window $\Psi_k=\{m_1,...,m_k\}$, the encoder can obtain the local minimum RD cost J^c as follows

$$J^c = \min_{m_i\in\Psi_k}\{J(m_i\,|\,s,\hat{m})\} \qquad (2)$$

where $0<i\le k$.

The spiral search establishes an inequality relation to accelerate the search process while preserving the RD optimized solution for the MV. The Lagrangian cost function consists of two parts, the distortion and rate. From (1), we can simply derive an inequality constraint between the cost and rate by eliminating the distortion $D(s,c(m))$. This leads to the result that, after examining the sub-window Ψ_k, the spiral search needs to evaluate the cost function only at the pixels m_j's which satisfy the following criterion

$$R(m_j-\hat{m}) \le \frac{J^c}{\lambda} \qquad (3)$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

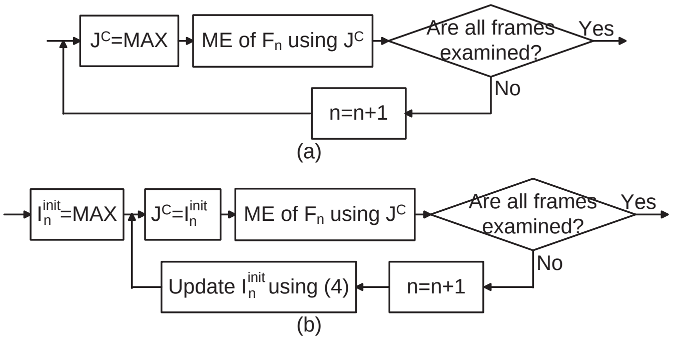


Fig. 1. ME of (a) the conventional and (b) the proposed methods.

where $m_j \notin \Psi_k$. When this inequality is satisfied at a search position, the cost function is evaluated at that particular position and if its value is less than J^c, J^c is updated.

From (3), we can see that the complexity reduction is highly dependent on the local minimum RD cost J^c. The lower J^c is, the less number of pixels the encoder needs to examine in the ME process. In the current implementation, the spiral search is applied to each reference frame separately. J^c is initialized with the highest possible value when starting the search process of a new reference frame (see Fig. 1(a)). This results in that, as shown in Fig. 2, the computational complexity of the ME process increases drastically proportional to the number of reference frames.

Suppose that there are N reference frames $\{F_1, F_2, ..., F_N\}$ in the reference list. Let I_n^{init} be an initial value of J^c, which will be used for ME of F_n, $0 < n \le N$. In order to reduce the number of pixels to be examined, the proposed method increases the tightness of the inequality constraint (3) by using the pre-calculated RD costs of the previously searched frames. Before starting the search process of F_n, the proposed method calculates I_n^{init} as follows

$$I_n^{init} = \min\{I_1^{\min}, ..., I_{n-1}^{\min}\} \tag{4}$$

where $I_u^{\min}$, $0 < u < n$, indicates the minimum RD cost obtained by the ME process of F_u. Then, the proposed method sets J^c to I_n^{init} and performs ME of F_n using (3). Fig. 1(b) shows the ME process adopting the proposed method. Note that, after finishing the ME process of F_u, the encoder can obtains the minimum RD cost $I_u^{\min}$ without additional computation. Therefore, the proposed method does not cause any computational overhead.

III. Simulation Results

In our simulation, we used JM 18.3 reference software and evaluated the performance by using several video sequences with CIF@30fps format. Only the first frame was encoded in intra mode and ME is performed with integer-pixel accuracy. The MV search range and the quantization parameter were set to 16 and 22, respectively. We measured the processing time of ME for the video sequences of 100 frames.

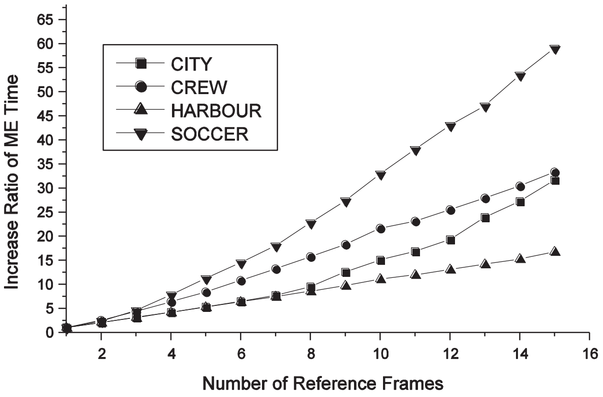


Fig. 2. Increase ratio of ME time to the number of reference frames for QCIF sequences with QP = 22.

Table I shows the measured encoding time (T) and the performance improvement (ΔT) of the conventional methods by adopting the proposed strategy, where N indicates the number of reference frames. In Table I, PDE and SP indicate, respectively, partial distortion elimination [2] and the spiral search of the reference software [3]. In our simulations, the reductions of average ME time of PDE and SP are 23.52% and 24.28%, respectively. And, Table I shows that, as N increases, the amount of performance improvement becomes larger. For example, when N is equal to 4, ΔT's of PDE and SP are 13.79% and 13.50%, respectively. And, for $N = 16$, they increase to 31.37% and 31.96%, respectively. The proposed method can reduce the average ME time up to 39.48% for the SOCCER sequence. Note that, in all simulations, the bitrate and PSNR of the encoder adopting the proposed method are exactly the same as those of the original encoder.

TABLE I
Performance Improvement of PDE and SP

Sequences	N=4				N=8			
	PDE		SP		PDE		SP	
	T(sec)	ΔT(%)	T(sec)	ΔT(%)	T(sec)	ΔT(%)	T(sec)	ΔT(%)
CITY	275	11.76	247	10.51	697	17.66	658	21.83
CREW	567	15.92	503	16.04	1280	24.06	1148	25.29
HARBOUR	352	10.75	343	10.33	937	18.85	873	18.22
SOCCER	504	16.72	455	17.11	1278	26.65	1151	30.00
	N=12				N=16			
CITY	1231	26.77	1155	27.04	1842	33.35	1755	33.68
CREW	1860	28.09	1727	29.22	2515	31.39	2354	30.97
HARBOUR	1436	19.89	1414	20.22	2084	23.36	2069	23.71
SOCCER	1925	33.69	1830	34.89	2647	37.36	2579	39.48

References

[1] C. Choi and J. Jeong, "New sorting-based partial distortion elimination algorithm for fast optimal motion estimation," *IEEE Trans. Consumer Electronics*, vol. 55, no. 4, pp. 2335-2340, Nov. 2009.

[2] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Boston, MA, Kluwer, 1991.

[3] K. P. Lim, G. Sullivan, and T. Wiegand, "Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods," *Joint Video Team*, Doc. JVT-N046, Jan. 2005.

[4] M. Z. Coban and R. M. Mersereau, "A Fast Exhaustive Search Algorithm for Rate-Constrained Motion Estimation," *IEEE Trans. Image Processing*, vol. 7, no. 5, pp. 769-773, May. 1998.

Complexity Scalable H.264/AVC-to-SVC Transcoding

Sebastiaan Van Leuven*, Jan De Cock*, Glenn Van Wallendael*, Rosario Garrido-Cantos†, and Rik Van de Walle*
*Multimedia Lab, Dept. of Electronics and Information Systems, Ghent University-IBBT, B-9050 Ledeberg-Ghent, Belgium
Email: {sebastiaan.vanleuven; jan.decock; glenn.vanwallendael; rik.vandewalle}@ugent.be
†Albacete Research Institute of Informatics, University of Castilla-La Mancha, Albacete, Spain, Email: charo@dsi.uclm.es

***Abstract*—Transcoding techniques require a high complexity and reduce the video quality with every transcoding step. When transcoding an input bitstream to scalable video coding (SVC), only one adaptation step is required and a scaled bitstream is extractable afterwards. To reduce the H.264/AVC-to-SVC transcoding complexity, we propose a transcoding architecture which combines optimized closed- and open-loop transcoding techniques. This transcoder scales the complexity depending on the available resources. Relative to a cascaded decoder-encoder the complexity can be reduced by 99.28%, while a high rate distortion efficiency is maintained and a high degree of scalability guaranteed.**

I. Introduction

Scalable video coding (SVC) is a powerful tool to handle varying network conditions or devices with different capabilities [1]. Video sequences encoded with SVC, the scalable extension of H.264/AVC [2], can easily be adapted to changing requirements by reducing resolution, quality, frame rate or a combination thereof. Since its standardization, an increasing amount of content has been encoded using H.264/AVC. in environments where scalability is required, an H.264/AVC-to-SVC transcoder can be used so only one transcoding step is required. After conversion, video streams can be scaled instantly using low complexity extraction, so energy consumption in the network is reduced. However, also the transcoding complexity should be as low as possible to reduce this energy cost.

Temporal transcoding has been proposed previously [3], [4]. Spatial transcoding [5] results in 60% complexity reduction due to fast mode decision. However, since only extracting lower resolution layers is possible, the bit rate can not be scaled with fine granularity. To overcome this scalability issue, [6] applies open-loop transcoding to coarse grain quality scalability (CGS), which reduces the computational complexity, while the resulting enhancement layer has no quality reduction. However, the extracted base layer suffers from drift because the quality reduction is performed in the transform domain, which results in propagation of faulty predictions.

II. Proposed System

To overcome bit rate, scalability and error drift issues in existing systems, we present a hybrid H.264/AVC-to-SVC transcoder for CGS based on [7]. This hybrid architecture combines an open-loop transcoder [6] and an optimized closed-loop transcoder [8]. The mode decision complexity of the closed-loop transcoding is reduced based on previous analysis [9], [10]. Furthermore, the complexity of the list prediction is reduced, as suggested in [11]. In previous work, the hybrid transcoder limits the open-loop transcoding to frames in the highest temporal layer (i.e., highest temporal ID Tid) such that drift effects are eliminated. A comparable RD is achieved and only 4.27% of the complexity of a cascaded decoder-encoder architecture is required. However, in a system with varying processing load, the transcoder might need to reduce the complexity even further.Therefore, we improve the hybrid transcoder so consecutive frames can be open-loop transcoded. To reduce the drift artifacts, the highest temporal layers should be open-loop transcoded first. Consequently, a small drift might slightly propagate through the open-loop transcoded frames, but is neutralized with every closed-loop transcoded frame. This results in a complexity-scalable transcoder, which depending on the available complexity adjusts the number of open-loop transcoded frames.

TABLE I
Complexity compared to a cascaded decoder-encoder. Hybrid $Tid \geq x$ means that these frames are open-loop transcoded.

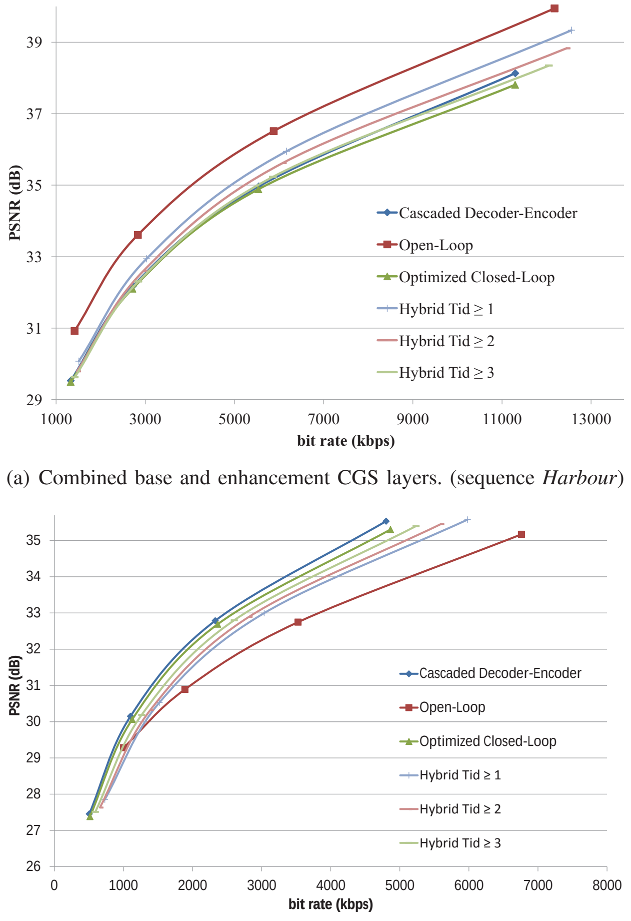
Level	Type	Complexity Reduction	Frames/GOP open-loop transcoded
1	Open-loop	∼100%	8
2	Hybrid $Tid \geq 1$	99.28%	7
3	Hybrid $Tid \geq 2$	98.10%	6
4	Hybrid $Tid \geq 3$	95.73%	4
5	Closed-loop	91.52%	0

With an 8 frames GOP, the required complexity is almost halved when open-loop transcoding frames with $Tid \geq 2$ compared to $Tid \geq 3$ (Table I), while only a small drift error is introduced. If even less complexity is available, the system can further shift towards an open-loop transcoding design, by increasing the number of open-loop transcoded frames, ultimately reaching the open-loop scenario. So, the proposed hybrid transcoding system is able to scale the complexity on a per frame basis ranging from optimized closed-loop transcoding to open-loop transcoding.

III. Results

The proposed system is evaluated against a cascaded decoder-encoder with six commonly used sequences (*Harbour, Ice, Rushhour, Soccer, Station, Tractor*). Each sequence was

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a) Combined base and enhancement CGS layers. (sequence *Harbour*)

(b) Extracted base layer for $\Delta QP = 5$. (sequence *Harbour*)

Fig. 1. RD for the original and proposed transcoding schemes.

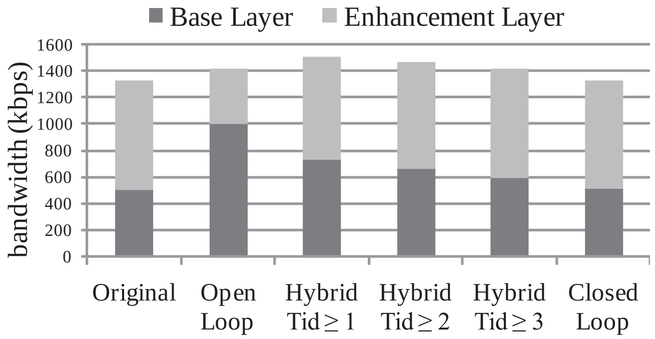


Fig. 2. Degree of scalability comparison for closed-loop, open-loop and hybrid transcoding (sequence *Harbour* with QP_{BL}= 37 and QP_{EL}= 32).

encoded as an H.264/AVC bitstream where different quantisation parameters were applied: $QP_{AVC} \in \{22, 27, 32, 37\}$. These bitstreams were transcoded to SVC with 2 CGS quality layers. To show opportunities for bit allocation per layer, multiple ΔQP values have been applied: $\Delta QP \in \{5, 6, 8\}$ ($QP_{BL} = QP_{AVC} + \Delta QP$). All bitstreams were generated using (JSVM_9_19_9) [12]. An intra period of 16 frames and a GOP size of 8 frames with a hierarchical prediction structure is applied, resulting in a maximal $Tid = 3$ which corresponds to five levels of complexity scalability, as enumerated in Table I.

The rate distortion of the SVC bit stream for *Harbour* is shown in Fig. 1(a). The open-loop RD-curve outperforms all other designs because the same quality as the input H.264/AVC bitstream is achieved. Closed-loop transcoded frames have a lower RD because the distorted decoded image is used as encoding input. As can be seen in Fig. 1(b), open-loop drift artifacts for the base layer are reduced in the hybrid scenarios. We also noticed that an increasing ΔQP will only slightly reduce the performance of the proposed system. To allow as much as possible devices to receive the base layer, the ratio of the base layer bit rate to the overall bit rate should be low. Fig. 2 shows the base layer bit rate in relation to the full bit rate. It can be seen that the base layer for the open-loop transcoded scenario requires a significantly higher bit rate.

IV. Conclusion

An H.264/AVC input bitstream can efficiently be transcoded to an SVC bitstream with CGS. By combining an optimized closed-loop transcoder with an open-loop transcoder, drifting artifacts of the base layer are reduced. Meanwhile the bit rate of both the base layer and the full bitstream are reduced, such that the scalability of the SVC stream is increased. For each temporal level, the complexity can be adapted, resulting in a complexity scalable transcoder.

References

[1] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” *IEEE Trans. CSVT*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.

[2] Joint Video Team of ISO/IEC MPEG & ITU-T VCEG, “Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264 and ISO/IEC 14496-10 Advanced Video Coding, Ed. 5.0,” Tech. Rep., 2010.

[3] A. Dziri, A. Diallo, M. Kieffer, and P. Duhamel, “P-Picture Based H.264 AVC to H.264 SVC Temporal Transcoding,” in *Int. Wireless Comm. and Mobile Computing Conf., 2008.*, aug. 2008, pp. 425 –430.

[4] R. Garrido-Cantos, J. De Cock, J.L. Martinez, S. Van Leuven, et al., “Motion-Based Temporal Transcoding from H.264/AVC-to-SVC in Baseline Profile,” *IEEE T-CE*, vol. 57, no. 1, pp. 239–246, Feb. 2011.

[5] R. Sachdeva, S. Johar, and E. M. Piccinelli, “Adding SVC Spatial Scalability to Existing H.264/AVC Video,” in *ACIS-ICIS*, H. Miao and G. Hu, Eds. IEEE Computer Society, 2009, pp. 1090–1095.

[6] J. De Cock, S. Notebaert, P. Lambert, and R. Van de Walle, “Architectures for Fast Transcoding of H.264/AVC to Quality-Scalable SVC Streams,” *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1209–1224, 2009.

[7] S. Van Leuven, J. De Cock, G. Van Wallendael, R. Van de Walle, R. Garrido-Cantos, J. Martinez, and P. Cuenca, “Combining open- and closed-loop architectures for H.264/AVC-TO-SVC transcoding,” in *18th IEEE Int. Conf. on Image Processing (ICIP)*, sept. 2011, pp. 1661 –1664.

[8] S. Van Leuven, J. De Cock, G. Van Wallendael, R. Van de Walle, R. Garrido-Cantos, J. Martinez, and P. Cuenca, “A low-complexity closed-loop H.264/AVC to quality-scalable SVC transcoder,” in *17th Int. Conf. on Digital Signal Processing (DSP)*, july 2011.

[9] G. Van Wallendael, S. Van Leuven, R. Garrido-Cantos, J. De Cock, J.-L. Martinez, P. Lambert, P. Cuenca, and R. Van de Walle, “Fast H.264/AVC-to-SVC Transcoding in a Mobile Television Environment,” in *6th Int. Mobile Multimedia Communications Conf.*, Sept. 2010.

[10] S. Van Leuven, K. De Wolf, P. Lambert, and R. Van de Walle, “Probability analysis for macroblock types in spatial enhancement layers for SVC,” in *Proceedings of the 11th IASTED International Conference on Signal and Image Processing*, Aug. 2009.

[11] S. Van Leuven, J. De Cock, R. Garrido-Cantos, et al., “Generic Techniques to Reduce SVC Enhancement Layer Encoding Complexity,” *IEEE Trans. CE*, vol. 57, no. 2, pp. 827–832, May 2011.

[12] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, “Joint Scalable Video Model,” MPEG / ITU-T, Tech. Rep., Jan. 2010.

Image Compression with Meanshift Based Inverse Colorization

Taekyung Ryu[1], Ping Wang[1], and Suk-Ho Lee[2]
Dept. Visual Contents[1] & Dept. Software Engineering[1], Dongseo University

Abstract-- **We propose a meanshift segmentation based inverse colorization method for image compression. The encoder makes use of the meanshift segmentation algorithm in automatically selecting the representative pixels from the original image from which the colored image is reconstructed by the decoder. Using the modes of the clustered regions as the representative pixels, the compression rate becomes high and the reconstructed image has good visual quality.**

I. INTRODUCTION

Colorization based coding refers to the color compression technique based on the use of colorization methods [1]-[4]. In the colorization approach, the color values of an image are obtained from a few pixels having color information [5]. The color information of these pixels is propagated to neighboring pixels by colorization methods making the whole image colorized.

Colorization based coding utilizes the fact that the required number of pixels having color information is small. The encoder chooses the pixels required for the colorization process, which are called RP (representative pixels) in [5], and maintains the color information only for the RP. The position information and the color values are sent to the decoder only for the RP. Then, the decoder restores the color information for the remaining pixels using colorization methods. The main issue in colorization based coding is how to extract the RP such that the compression rate and the quality of the restored image become good.

In this paper, we propose a meanshift segmentation [6] based RP selection method. When using the modes as the RP, the RP have a main effect in the colorization process on the regions which approximately correspond to the regions clustered by the meanshift segmentation. Therefore, it becomes possible to reconstruct all the colors to a sufficient level in the decoder. The meanshift procedure has to be repeated several times, until an almost optimal set of required RP are obtained.

II. PROPOSED SCHEME

The main steps of the iterative meanshift approach are as follows: we first perform a meanshift procedure on the image with a kernel having a rather large bandwidth. Then we evaluate the segmented regions by comparing the values of the reconstructed and the original color components. If the color components are similar enough, the modes are assigned as the RP for this region, where the modes refer to the local maxima of the probability density of the feature space used in the meanshift [6]. These modes are used as the RP in the decoder and have an effect over regions that approximately correspond to the regions which contain the same mode in the meanshift segmentation algorithm. If the difference in the color values is large in this region, we perform an additional meanshift segmentation inside this region with a kernel with smaller bandwidth. Then, each sub-region is evaluated again to decide if any further segmentation is needed. Due to the evaluation after each meanshift step, additional RP are added to the cluster. As a result, the set of RP grows gradually until it becomes an optimal set having good reconstructing results in the decoder. This is due to the fact that the modes, which are chosen as the RP, are the most representative pixels inside the segmented regions, and therefore, have the largest effect inside the segmented regions. As a result, assigning these modes as the RP, we can produce a minimal set of RP from which the color components can be reconstructed with high fidelity.

In the decoder, the meanshift segmentation can be performed again on the decoded luminance channel. The segmented result will be the same as in the encoder since the same illuminance channel is used. The segmentation result will then provide the decoder with information on the segmentation regions. Then, the colorization effect of the RP can be restricted within the segmented region. The whole system diagram of the proposed scheme is shown in Fig. 2 and the detailed system flow in Fig. 3.

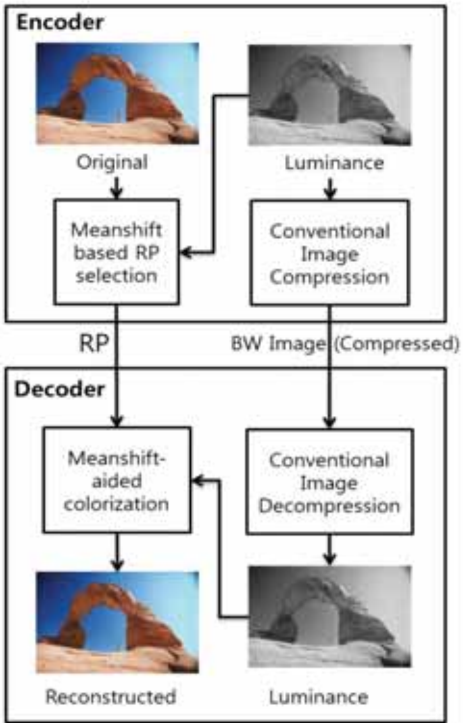

Encoder
Original
Luminance
Meanshift based RP selection
Conventional Image Compression
RP
BW Image (Compressed)
Decoder
Meanshift-aided colorization
Conventional Image Decompression
Reconstructed
Luminance

Fig. 2 Diagram of the proposed meanshift based encoder and decoder.

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No.2012-0003905).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

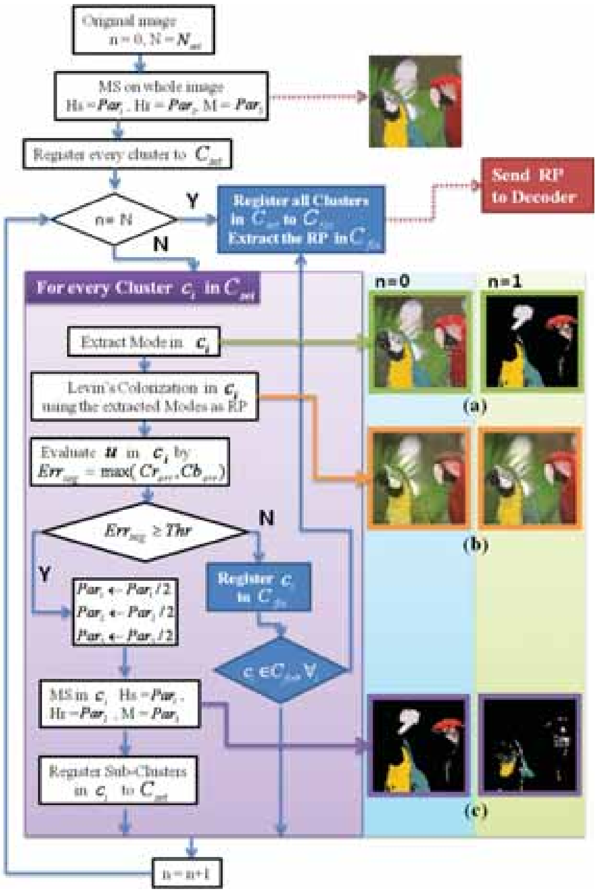

Original image
n = 0, N = N_{set}
MS on whole image
Hs = Par_1 , Hr = Par_2 , M = Par_3
Register every cluster to C_{set}
n= N
Y
N
Register all Clusters in C_{set} to C_{fix}
Extract the RP in C_{fix}
Send RP to Decoder
For every Cluster c_i in C_{set}
n=0
n=1
Extract Mode in c_i
Levin's Colorization in c_i using the extracted Modes as RP
(a)
Evaluate u in c_i by
Err_{avg} = max(Cr_{err} , Cb_{err})
$Err_{avg} \geq Thr$
N
(b)
Y
$Par_1 \leftarrow Par_1/2$
$Par_2 \leftarrow Par_2/2$
$Par_3 \leftarrow Par_3/2$
Register c_i to C_{fix}
$c_i \in C_{fix}, \forall_i$
MS in c_i Hs = Par_1 ,
Hr = Par_2 , M = Par_3
Register Sub-Clusters in c_i to C_{set}
(c)
n = n+1

Fig. 3 System flow chart of the encoder in the proposed scheme

III. EXPERIMENTAL RESULTS

We performed the proposed scheme with $N = 2$, and let the initial parameters be $Hs = 14, Hr = 8, M = 40$. We compared the reconstructed results between the JPEG encoded, the randomly selected RP encoded, and the proposed scheme encoded images. The sampling format used in the schemes is 4:2:0, which means that for the test image which has a 256×256 size, we recover the color components for a 128×128size. The number of RP obtained with the proposed scheme is 220, and we used the same number of RP in the random RP using scheme. Therefore, the size of data to store the color information is 880 bytes, using 2bytes for the position vector and 2bytes for the *Cb* and *Cr* components. Figure 4 compares the decoded images. The file size of the encoded image in the proposed and the random RP using scheme is 3954 bytes, using 3074 bytes for the *Y* component and 880 bytes for the color components. The file size of the encoded image using the JPEG scheme is 4257 bytes with QP (quantization parameter) = 19, where the same number of bytes (3074 bytes) is used for the *Y* component and 1183 bytes are used for the color components. Therefore, we have an additional 25.6% ($(1183-880)/1183 \times 100)$) compression gain over the JPEG scheme for the color components. The SSIM (Structural Similarity) values for the (*Cb*, *Cr*) color components are (0.8497, 0.8602) for the JPEG scheme, (0.8427, 0.8729) for the proposed scheme, and (0.7114, 0.7293) for the random RP scheme. We have similar SSIM values while having higher compression rate than the JPEG compression algorithm. Furthermore, we can see from Fig. 4 that the decoded image using the proposed scheme has less false colors and blocky artifacts than that using the JPEG algorithm.

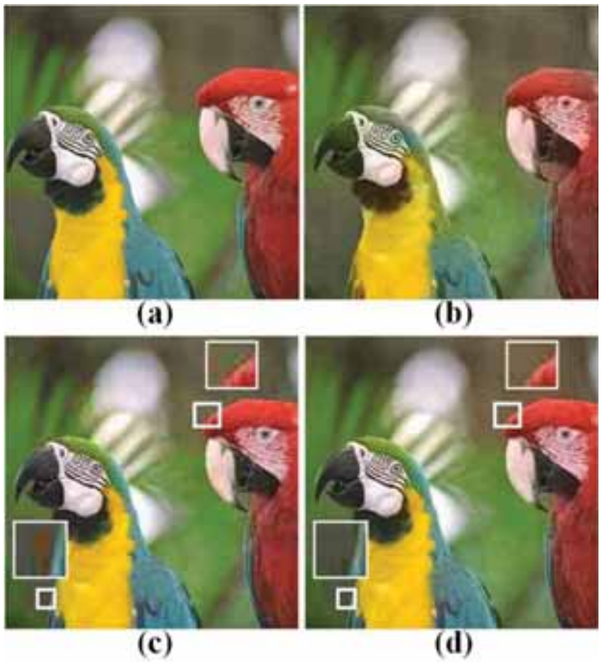


Fig. 4 Comparison between different compression schemes (a) Original image (b) Randgsview32om RP scheme (SSIM = (0.7114, 0.7293)) (c) JPEG (SSIM=(0.8497,0.8602) (d) Proposed (SSIM = (0.8427, 0.8729))

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of L. Cheng and S.V.N. Vishwanathan, "Learning to Compress Images and Videos," *Proc. ICML*, vol. 227, 2007, pp. 161-168.

[2] X. He, M. Ji, and H. Bao, "A Unified Active and Semi-supervised Learning Framework for Image Compression ," *IEEE CVPR,* 2009, pp. 65-72.

[3] T. Miyata, Y. Komiyama, Y. Inazumi, and Y. Sakai, "Novel Inverse Colorization for Image Compression," *Proc. Picture Coding Symposium,* 2009, pp. 1-7.

[4] O. Shunsuke, M. Takamich, and S. Yoshinori, "Colorization-based Coding by focusing on Characteristics of Colorization Bases," *Proc. Picture Coding Symposium,* 2010, pp. 11-17.

[5] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using Optimization," *ACM Transactions on Graphics,* vol. 23, Aug. 2004, pp. 689-694.

[6] D. Comaniciu, and P. Meer, "Mean Shift : A Robust Approach toward Feature Space Analysis," *IEEE Trans. PAMI,* vol. 24, May 2002, pp. 603-619.

Resolution-Based Compressed Sensing for Catadioptric Omni-Directional Imaging Method

Jingtao Lou, Yu Liu, Yongle Li, and Maojun Zhang
College of Information System and Management
National University of Defense Technology, Changsha, China

Abstract-- **This paper presents a compressed sensing method to solve the problem of low and non-uniform resolution in catadioptric omni-directional imaging system. A non-uniform measurement matrix is designed according to the distribution of the system's resolution. Numerical simulation shows that the proposed method is feasible and effective.**

I. INTRODUCTION

Catadioptric omni-directional imaging is widely used in many applications, such as video surveillance, robot navigation, three-dimensional (3-D) reconstruction, etc, owing to its advantage of one-shot seamless panoramic imaging with a 360 degree field of view. However, as the application research of omni-directional imaging is further investigated, the problem of low and non-uniform resolution has obviously become the main obstacle of its generalization. Some resolution enhancement methods depending on post-processing are published in [1, 2]. But, due to the low and non-uniform resolution of original sampled omni-images, the improvement is very limited. Using the SVAVISCA sensor which has a non-uniform pixel density that decreases with radius, Stefan [3] designed an omni-directional camera. This approach is simple and intelligible, but has high cost and alignment problems. Based on multiple reflecting mirrors, Chen [4] designed a complementary-structure omni-sensor for resolution enhancement. This method has low cost, but decreases the vertical field of view. This paper combines compressed sensing [5, 6] and omni-directional imaging to solve the resolution problem. Next we describe the method in details, give the experimental results, and draw the conclusion of the proposed work.

II. ALGORITHM

Compressed sensing (CS) suggests that one can reconstruct signals from significantly fewer samples or measurements than Nyquist/Shannon sampling theory uses, since CS relies on the fact that many natural signals are sparse or compressible in proper basis. It provides a new approach for improving imaging resolution. By designing the distribution of measurement matrix, the image sensor obtains the compressed samples of the observed scene. Through the reconstruction algorithm, we can get the high resolution image without reducing pixel-pitch. The proposed method applies CS to omni-directional imaging. We design a non-uniform measurement matrix according to the distribution of the system's resolution. Using the unequal compressed omni-image, the high and uniform resolution image can be recovered from the reconstruction algorithm.

A. Non-uniform Distribution of Resolution

Since omni-directional sensors have a 360 degree field of view that is much wider than conventional cameras, the pixels of a same spatial object in omni-directional imaging must be relatively very few when the amount of pixels on the image plane is fixed. Simultaneously, because of the difference between the amounts of pixels in the inner and outer within a same radial span, the resolution of the inner part is poorer than the outer in omni-images. Based on the unifying model for catadioptric projective geometry (Fig. 1), which proposed by Geyer [7], we analyze the distribution of system's resolution (defined as dA/dv), and obtain the relationship of resolution between omni-sensor and projective camera.

$$\frac{dA}{dv}=\left[\frac{(1-\xi^2)(1-\xi^2+r^2+(\xi-z)^2)}{2\times((\xi-z)^2+r^2\times(1-\xi^2))^{3/2}}\right]\frac{dA}{d\omega'} \tag{1}$$

Where $dA/d\omega'$ is the resolution of projective camera, and (r, z) is the coordinate on the unit sphere.

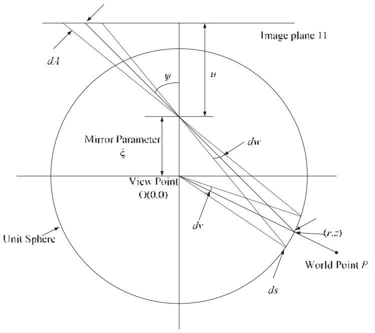


Fig. 1 The unifying model for catadioptric projective model

B. Design of Non-uniform Measurement Matrix

Fig. 2 describes the relationship between the recovery error and number of measurements.

Suppose that $\boldsymbol{x}$ is a K-sparse signal of length N, where N is 256 and K takes the values of 5, 10, 15, 20 and 80 in Fig.2. When K is invariable, the MSE grows against the number of measurements. If it aims to obtain the same MSE of different signal, larger M should be applied for dense signal. The main idea of our imaging method is applying the distribution of system's resolution to CS scheme.

This work was partially supported by National Natural Science Foundation of China with project No.61175006, No.61175015, No.61275016, No.61271438.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

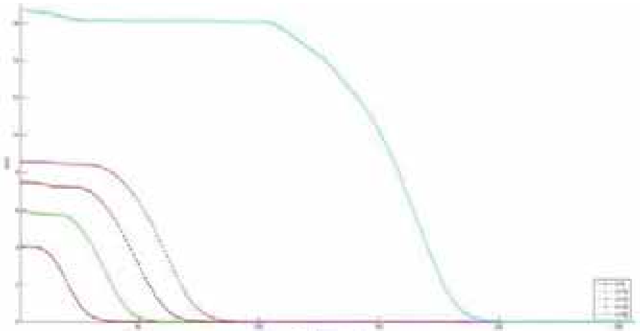


Fig. 2 Recovery error versus number of measurements for *K*-sparse signal. The error is measured using the standard definition of mean square error.

Since CS is performed in the manner of block-processing, we need to calculate the resolution of each image block. Assuming that the omni-image is split into non-overlapped blocks of fixed size, and let B_i denote the *i*-th block of the image. The resolution of *i*-th block can be defined as

$$\varsigma_i = \frac{1}{N}\sum_{j\in B_i}\rho_j \qquad (2)$$

Where *N* is the total number of pixels in B_i, and ρ_j is the resolution at position *j* (see (1) for its computation).

After the resolution of B_i is obtained, the number of CS measurements assigned to block B_i can be calculated as

$$M_i = round(\frac{1}{\varsigma_i}.(M_{\max} - M_{\min}) + M_{\min}) \qquad (3)$$

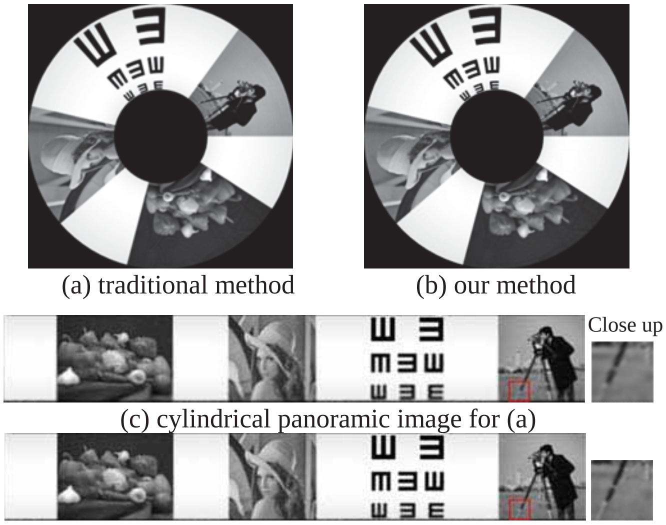
Where the function "*round*(*)" forces input quantity equal to the nearest integer. Note that $M_{\max}$ and $M_{\min}$ are the possible maximum and minimum of the number of random CS measurements that can be assigned to a block.

According to the non-uniform measurement matrix *M*, the linear Bregman iteration [8] is employed for the reconstruction in order to obtain the high and uniform resolution omni-image. According to human visual habits, we unwrap the omni-image to cylindrical panoramic image based on panoramic unwrapping method.

III. Numerical Experiments

To demonstrate the effectiveness of the above compressive sensing method, we simulate reconstruction of the "Cameraman", "lena", "peppers" and USAF resolution test images in a cylindrical panoramic space. Fig. 3 shows the results of original imaging method and our proposed method. Fig. 3(a) is the image captured by traditional catadioptric imaging method. Fig. 3(b) is the image captured by our method. Due to the serious annular distortion, the original captured omni-image usually needs to be unwrapped to cylindrical panoramic image, which is more proper for human visual perception. The cylindrical panoramic images correspond to Fig. 3(a) and (b) are shown in Fig. 3(c) and (d).

The experimental results show that the proposed imaging method can obtain high and uniform resolution cylindrical panoramic image without changing on image sensor or reflecting mirror.



(a) traditional method

(b) our method

Close up

(c) cylindrical panoramic image for (a)

(d) cylindrical panoramic image for (b)

Fig. 3 The simulation results. (a)omni-image captured by traditional catadioptric imaging method. (b) omni-image obtainend by proposed method. (c) cylindrical panoramic image unwrapped from (a). (c) cylindrical panoramic image unwrapped from (b).

IV. CONCLUSION

We discussed a novel algorithm to solve the resolution problem of catadioptric imaging system. Firstly, we analyzed the distribution of the omni-sensor's resolution. Then we applied the distribution to get the non-uniform measurement matrix. Experimental results show that our algorithm is feasible and effective. And the proposed method can also work well in fisy-eye camera. The omni-imaging device can be used in remotely-piloted Unmanned Surface Vehicle (USV) application [9]. As a future work, we will conduct further study to optimize the measurement matrix and reconstruction algorithm in order to develop a real-world system for catadioptric omni-directional compressive imaging.

REFERENCE

[1] Jeng S.W. and Cai W.H.. Improving quality of unwarped omni-images with irregularly-distributed unfilled pixels by a new edge-preserving interpolation technique[J]. Pattern Recognition Letters, 2007, 28: 1926-1936.

[2] Nagahara H., Yagi Y., Yachida M.. Resolution improving method from multi-focal omnidirectional images[C]. Proceedings of International Conference on Image Processing, 2001, 654-657.

[3] Stefan G.. Mirror Design for an omni-directional camera with a uniform cylindrical projection when using the SVAVISCA sensor. Research Reports of CMP, Czech Technical University in Prague, 2001.

[4] Chen L.D., Wang W., Zhang M.J.. Complementary-structure catadioptric omnidirectional sensor design for resolution enhancement[J]. Optical Engineering, 2011, 50(3): 033201.

[5] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.

[6] Ma J. Single-pixel remote sensing[J]. IEEE Geoscience and Remote Sensing Letters, 2009, 6(2): 199-203.

[7] Geyer C, Daniilidis K. Catadioptric projective geometry[J]. International Journal of Computer Vision, 2001, 45(N3): 223-243.

[8] Yin W, Osher S, Goldfarb D, et al. Bregrnan iterative algorithms for l1 minimization with applications to compressed sensing[J]. SIAM Journal on Imaging Sciences, 2008:143-168.

[9] http://www.remotereality.com/

Multi-histogram Based Scene Change Detection for Frame Rate Up-Conversion

Suk-Ju Kang[1], Sung In Cho[2], Sungjoo Yoo[2] and Young Hwan Kim[2]

[1]Department of Electrical Engineering, Dong-A University, Busan, Republic of Korea

[2]Department of Electronics and Electrical Engineering, POSTECH, Pohang, Republic of Korea

***Abstract*-- In this paper, we propose a new scene change detection method based on multi-histogram for frame rate up-conversion. The proposed method manages multiple per-block histograms to extract the locality of image change. Thus, it can detect local scene change as well as global scene change between frames. Experiments show that the proposed method improves by 14.05dB the image quality of the interpolated frame with the local scene change.**

I. INTRODUCTION

Motion compensated frame rate up-conversion (MC-FRUC) is a technique that changes an original frame rate (e.g., 30fps) to a higher frame rate of moving images using the motion vector, which is the displacement of an object between consecutive frames [1]. MC-FRUC assumes that consecutive frames can be reconstructed using motion vectors, previous frame, and current frame. In reality, it is often that the assumption does not hold when scene changes. In such a case, if MC-FRUC is still applied, the estimated motion vectors will be erroneous thereby yielding poor images as the result of MC-FRUC. In order to resolve this problem, MC-FRUC needs to be applied selectively to consecutive frames without intervening scene change. To do that, it requires scene change detection. If a scene change is detected, MC-FRUC is not applied to the first frame of a new scene although it is applied to subsequent frames. Several scene change detection methods have been proposed: methods based on motion vector [2], histogram difference [3], and edge matching [4]. These methods, however, consider only global scene change (GSC) detection between consecutive frames. However, it is usual that scene changes occur locally, e.g., captions in News, sports, etc. In these cases of local scene change (LSC), MC-FRUC estimates motion vectors incorrectly, and the incorrect motion vector results in poor quality in the interpolated frames, e.g., block artifacts.

The proposed method enables the detection of both LSC and GSC. In order to detect LSC, it splits a frame into several blocks and performs histogramming for each block. Note that conventional histogram-based methods perform histogramming for the entire frame [3]. Thus, the LSC can be easily detected by the histogram difference on the corresponding blocks. MC-FRUC is applied to blocks without LSC. Thus, the artifacts on the blocks with LSC, which could occur if MC-FRUC were applied, are avoided. The proposed method detects GSC in the same manner as the conventional methods [2]-[4].

II. PROPOSED SCENE CHANGE DETECTION METHOD

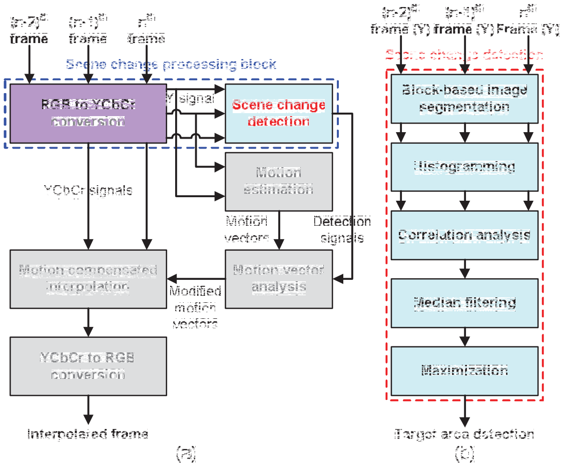


Figure 1. (a) MC-FRUC architecture using the proposed scene change detection, and (b) the architecture of the proposed scene change detection

Fig. 1 (a) shows the block diagram of the proposed multi-histogram based scene change detection. The scene change detection block gives the LSC information to motion vector analysis block to filter out motion vectors for blocks with LSC. Fig. 1 (b) shows the internal operation of the scene change detection block. It takes, as the input, three consecutive frames, $(n\text{-}2)^{th}$, $(n\text{-}1)^{th}$, and n^{th} frames in Y (luminance) values. The correlation analysis step calculates two metrics. For each block in the frame, it calculates the sum of histogram absolute difference (SHAD) between the corresponding blocks in two consecutive frames. For each frame, it calculates the mean value of SHADs for all blocks, M_n (for n^{th} frame). Equation (1) shows how to calculate the two metrics.

$$\mathrm{SHAD} = \sum_{k=0}^{255} \left| \mathrm{BH}_{\mathrm{n}}(k) - \mathrm{BH}_{\mathrm{n\text{-}1}}(k) \right|, \qquad (1)$$
$$\mathrm{M}_{\mathrm{n}} = \frac{1}{\mathrm{pq}} \sum_{i=1}^{\mathrm{p}} \sum_{j=1}^{\mathrm{q}} \mathrm{SHAD}(i,\, j),$$

where $\mathrm{BH_n}$ and $\mathrm{BH_{n\text{-}1}}$ denote the histogram of each block in the current (n^{th}) and previous ($n\text{-}1^{th}$) frames. p and q denote the number of blocks in a row and a column, respectively. Using SHADs and mean values, we perform the correlation analysis as shown in (2).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$GSC_n = \begin{cases} 1 & \text{if } M_{n-1}/M_n \le T \\ 0 & \text{otherwise}, \end{cases} \qquad (2)$$

$$LSCC_n = \begin{cases} 1 & \text{if } SHAD_{n-1}/SHAD_n \le T \\ 0 & \text{otherwise.} \end{cases}$$

If Mn-1/Mn is below T, the correlation threshold, we decide that GSC occurs. Otherwise, we check to see if the LSC occurs by comparing SHADn-1/SHADn to T for each block. If there is any block b whose SHADbn-1/SHADbn is below T, then the block b is classified to be LSC candidate (LSCCbn becomes 1 as shown in (2)), which has the possibility of LSC. To reduce the possibility of wrong LSC detection due to pixel errors, we apply median filtering to the histogram differences of LSCC blocks. Then, in order to identify the LSC area as large as possible, we perform maximization which selects the minimum and maximum values in x and y axes on the LSCC blocks.

III. Experimental results

We evaluated the performance of the proposed method for both LSC and GSC detections. For test sequences, we used News sequence (320x240), which is a user-defined sequence with the LSC (a caption disappears during News broadcast), and Table tennis sequence (320x240), which is a general test sequence with the GSC. In the experiments, the frame was split into 40x40 pixel blocks. Fig. 2 shows the LSC experiments: the input sequences ((a), (b) and (c)), correlation results ((d), (e), and (f)), and two result frames obtained from the conventional method (g) and the proposed method (h). Fig. 2 (d) shows that the histogram differences between corresponding blocks in the (n-2)th and (n-1)th frames. Fig. 2 (e) shows the same data for (n-1)th and nth frames. The comparison of the two figures shows that there is a local scene change at the bottom-right corner. The proposed method detected the LSC area accurately as shown in Fig. 2 (f) which was obtained from the median filtering and the maximization (Fig. 1 (b)). We compared the image quality of two interpolated frames (one from the conventional method [3] and the other from the proposed method) with respect to the original image as in [1]. In the objective evaluation, the comparison shows that the proposed method improved by 14.05 dB the PSNR (peak signal to noise ratio): 22.94 dB (conventional method) and 36.99 dB (proposed method). It is because the proposed method removed the block artifacts in the interpolated frame while the conventional method generated block artifacts in the interpolated image. Fig. 3(c) shows the GSC experiment with the Table tennis sequence, which has two GSCs in the total sequence as shown in Fig. 3 (a) and (b). Fig. 3 (c) shows that the proposed method detects the two GSCs using the mean value of multi-histogram differences.

IV. Conclusion

In this paper, we proposed the multi-histogram based scene change detection. The proposed method detects LSC as well as GSC by utilizing per-block histograms. Our experiments show that the proposed method detects both the GSC and the LSC. The MC-FRUC enhanced with the proposed scene detection method gives a quality improvement of 14.05dB in the PSNR of the interpolated frames with LSC.

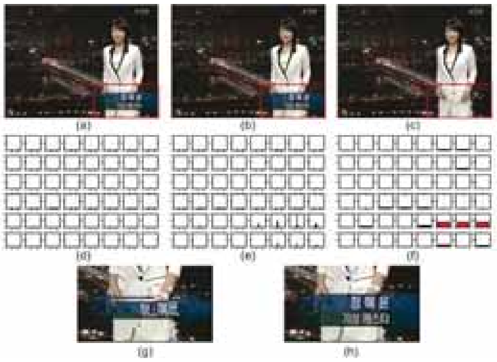


Figure 2. (a) (n-2)th frame, (b) (n-1)th frame, (c) nth frame, (d) multi-histogram difference between the (n-2)th and (n-1)th frames, (e) multi-histogram difference between the (n-1)th and the (n)th frames, (f) final detected area, (g) interpolated frame for the conventional methods, and (h) interpolated frame for the proposed method

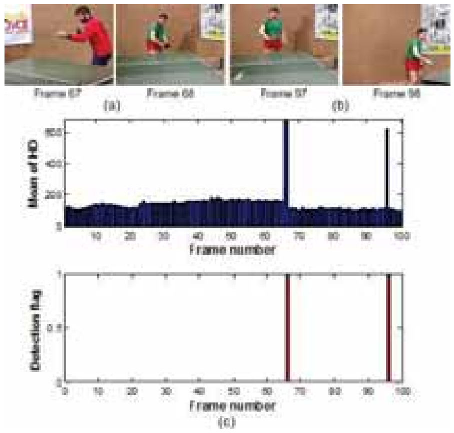


Figure 3. (a) (b) Previous and current frames, and (c) the mean value of multi-histogram differences and the detection flag

References

[1] S. J. Kang, K. R. Cho, and Y. H. Kim, "Motion compensated frame rate up-conversion using extended bilateral motion estimation," *IEEE Trans. Consumer Electronics*, vol. 53, no. 4, pp. 1759-1767, Nov. 2007.

[2] P. Bouthemy,M. Gelgon, and F. Ganansia, “A unified approach to shot change detection and camera motion characterization,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 7, pp. 1030-1044, Oct. 1999.

[3] J. R. Kim, S. Suh, and S. Sull, "Fast Scene Change Detection for Personal Video Recorder," *IEEE Trans. Consumer Electronics*, vol. 49, no. 3, pp.683-688, Aug. 2003.

[4] U. Gargi, R. Kasturi, and S. H. Strayer, “Performance characterization of video-shot-change detection methods,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 1-13, Feb. 2000.

Active Shutter Glasses for 3D HDTV with Flexible Liquid Crystal Lens

Jeong In Han
Department of Chemical and Biochemical Engineering,
Dongguk University-Seoul, Seoul, 100-715, Korea
hanji@dongguk.edu

Abstract - The active shutter glasses for 3D HDTV is developed using flexible liquid crystal (FLC) lens. LC lens is made on polycarbonate (PC) substrates through conventional LCD processes. It shows the excellent viewing properties for 3D image.

Keywords-component; active shutter goggle; 3D HDTV, flexible liquid crystal lens; fast response time; liquid crystal lens driver

I. INTRODUCTION

3D movies such as Avatar have attracted lots of interests. The active shutter type 3D HDTV displays stereoscopic 3D images, presenting the image through left eye while blocking the right eye's view. Subsequently, the vice versa procedure performs with fast repetition rate which allows no interference with the perceived fusion of the two images into a single 3D image [1]. The twisted nematic (TN) liquid crystal lens work as the optical shutter. Each eye's glass of the active shutter glasses contains a TN liquid crystal layer which becomes dark when the voltage is applied and transparent in case of no applied voltage. Controlling of the glasses can be achieved by an alternative trigger signal that allows sequent darkening of one eye and then the other in accordance with the refresh rate of the screen. [1]-[2]

So far, TN liquid crystal lens have been fabricated with rigid and brittle glass substrates. Therefore, it is easily broken and degraded when external forces or impacts are applied, threatening fatal wounds to the human's eyes. Moreover, glass materials are too brittle or hard to control and thus the applications to the various shapes or designs of the lenses have been limited.

In this paper, to overcome these problems, flexible lens system has been made on the flexible polymeric substrate. The prototype of active shutter glasses with the flexible liquid crystal lens is designed and demonstrated.

II. FABRICATION OF FLEXIBLE LIQUID CRYSTAL LENS AND THE ACTIVE SHUTTER GLASSES

A. *Making a flexible liquid crysal lens and measurements of the electro-optical properties*

Flexible liquid crystal lens is the flexible LCD with one huge pixel. They were made by the conventional fabricating procedures of the flexible LCD [3]-[5] except the pixel patterning process. The flexible liquid crystal lens is fabricated on the substrate of the polycarbonate (PC) film of 130 □m on which the transparent conducting layer of ITO film are deposited by the sputtering. The rubbing process of polyimide (PI) is performed on the ITO films of PC substrates for the liquid crystal alignment between ITO coated PC films. The sheet resistance of ITO film is 16.81Ω/□. PI rubbing direction of the upper and lower PC films have to be 90^{O} to twist the liquid crystal layer. Then, the spacers are splayed on one PC film and the sealant is dispensed on the other PC substrate. The upper and lower PC films are assembled together and the liquid crystal is injected to the flexible twisted nematic liquid crystal device. Finally polarizers are attached on it to make very large one pixel flexible liquid crystal lens

Fig. 1 shows the vertical view of the flexible liquid crystal lens fabricated. The cell gap between the upper and lower PC film is 5 □m.

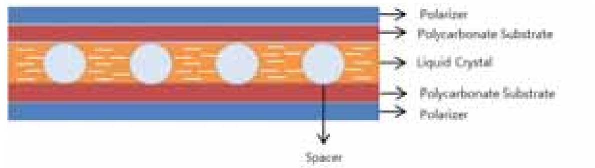


Fig. 1 Vertical view of the flexible liquid crystal lens

The light transmittance with the applied voltage and the response time of the flexible liquid crystal lens are measured. The light transmittance with the applied voltage is shown in Fig. 2. The maximum transmittance of the flexible liquid crystal lens is 32% and that of the glass liquid crystal lens is less than 35 %. From Fig. 2, the contrast ration (C/R) of 177:1 can be calculated.

The response time of the flexible liquid crystal lens are measured and illustrated in Fig. 3. The rising and falling response time is 160 □s and 2.4 ms, respectively. These response times are relatively fast than that of the TFT-LCD.

Fig. 4 shows the flexible liquid crystal lens fabricated on PC substrate in this works. Tts shape is not rectangular, which is one of the advantages of the flexible liquid crystal lens.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

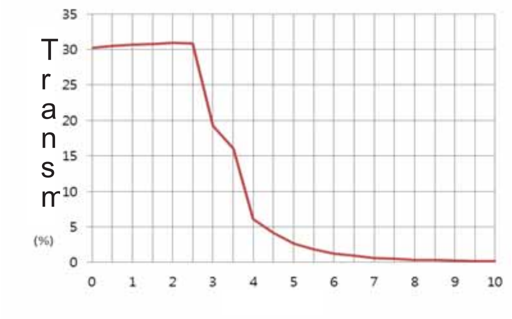


Fig.2 The light transmittance with the applied voltage of the flexible liquid crystal lens fabricated on PC substrate.

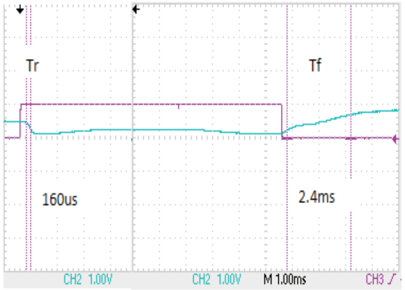


Fig. 3 The rising and falling response time of the flexible liquid crystal lens.

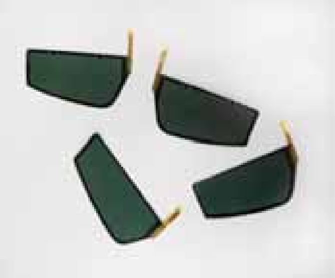


Fig. 4 The flexible liquid crystal lens fabricated in this works.

B. Making a prototype of active shutter glasses with the flexible liquid crysal lens and measurements of the electro-optical properties

Block diagram of the reception module for the active shutter glasses with the flexible liquid crystal lens is shown in Fig. 5. As a power supply part, the charger (XC6801), the regulator (XC6221) and the DC-DC converter (XC9119) are added for the charging of Li-polymer battery, the output of the constant voltage of 3 VDC and converting of 3 VDC to 10 VDC for the driving of the flexible liquid crystal lens, respectively. In order to control the right and left flexible liquid crystal lens with the synchronization with IR 3D signal, MCU of ATtiny24 is used. AD1334 is adopted to drive the flexible liquid crystal lens. Also, as a power source, Li-polymer battery is adopted. In upper right corner, power line of 3.8 VDC, 3 VDC and 10 VDC and signal line of control and check are illustrated in Fig. 5.

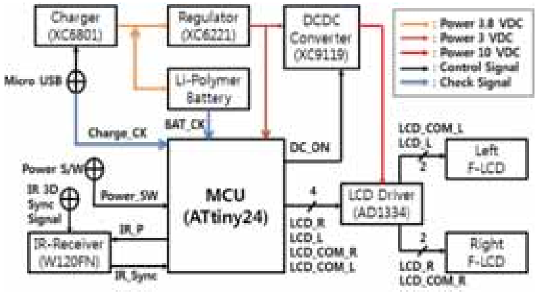


Fig. 5 The Block diagrams for the active shutter glasses with the flexible liquid crystal lens

The prototype of the active shutter glasses with the flexible liquid crystal lens is fabricated and demonstrated. The active shutter glasses for 3D HDTV with flexible liquid crystal lens shows the excellent viewing characteristics for 3D images qualities**.**

III. SUMMARY

The flexible liquid crystal lens for the active shutter glasses for 3D HDTV is developed. The flexible liquid crystal lens of twisted nematic (TN) mode is made on the polycarbonate (PC) substrate through the conventional processes. The flexible liquid crystal lens shows the maximum transmittance of 32% and the total response time of 2.56 ms. C/R is 177:1. For the prototype of the active shutter glasses with the flexible liquid crystal lens, the reception module with the power supply part, MCU and the drivers are developed, also. The active shutter glasses for 3D HDTV with flexible liquid crystal lens shows the excellent viewing properties for 3D images of 3D HDTV.

REFERENCES

[1] Active shutter 3D system-Wikipedia, www.wikipedia.org

[2] Lin Edwards, “Active Shutter 3D Technolgy for HDTV”, Phys Org, Sept 25, 2009

[3] S. K. Park, J. I. Han, W. K. Kim and M. G. Kwak, “Development of 2-Inch Plastic Film STN LCD”, J. of Information Display, 1 (1), 2000, pp 14 - 19.

[4] S. J. Hong C. J. Lee, J. I. Han, W. K. Kim, D. G. Moon, M. G. Kwak, S. K. Park and Y. H. Kim, “Flexible Metal-Insulator-Metal Devices for Plastic Film AM-LCD”, Current Applied Physics, Vol 2, 2002, pp 245 - 248.

[5] Y. H. Kim, S. K. Park, D. G. Moon, W. K. Kim and J. I. Han, “Organic Thin Film Transistor-Driven Liquid Crystal Displays on Flexible Polymer Substrate”, Japanese Journal of Applied Physics, Vol 43, No 6A, 2004, pp 3605 – 3608.

Encryption for High Efficiency Video Coding with Video Adaptation Capabilities

Glenn Van Wallendael[1], *Student Member, IEEE*, Andras Boho[2], Jan De Cock[1], *Member, IEEE*,
Adrian Munteanu[3], *Member, IEEE*, Rik Van de Walle[1], *Member, IEEE*
[1]Department of Electronics and Information Systems – Multimedia Lab, Ghent University – IBBT, Ghent, Belgium
[2]Department of Electrical Engineering, KU Leuven, Leuven, Belgium
[3]Department of Electronics and Informatics, Vrije Universiteit Brussel – IBBT, Brussels, Belgium

***Abstract*—In this paper, we describe encryption possibilities for the High Efficiency Video Coding (HEVC) standard under development. Bitstream elements which maintain HEVC compatibility after encryption are listed and their impact on video adaptation is described. From this list, three bitstream elements are selected, namely intra prediction mode difference, motion vector difference sign, and residual sign. These elements provide good protection of the video information and result in 0.0% Bjøntegaard delta bitrate increase because of their equal probability entropy encoding property.**

I. Introduction

Video encryption is an attractive technique enabling video service providers to prevent unauthorized devices from playing back their content. The most straightforward and most secure solution would be to encrypt the entire compressed video stream. This would obfuscate all the information in the video stream although this might not be necessary or desirable. In particular, video adaptation performed at network level requires access to specific syntax elements in the video stream. In this case, the adaptation node should be able to decrypt in real-time parts of the video stream needed during the adaptation process. Consequently, the adaptation node should be trusted with decryption keys or certain parts of the video stream should be left unencrypted. We aim at jointly offering encryption and video adaptation capabilities, while avoiding the deployment of decryption keys at network level. Our solution is then to degrade the visual quality as much as possible for untrusted decoders by encrypting a minimal set of bitstream elements.

For the widely adopted H.264/AVC standard, encryption strategies taking into account adaptation possibilities are investigated thoroughly in [1]. To efficiently compress higher resolutions, a successor of H.264/AVC is being developed, called High Efficiency Video Coding (HEVC) [2]. In this paper, strategies for encrypting HEVC video streams are described together with their impact on transcoding algorithms. In terms of structure, first a description of HEVC is given in Section II. Then, adaptation algorithms that can be mapped from H.264/AVC to HEVC and their restrictions on encryption strategies are described in Section III. Next, Section IV looks further into the encrypted bitstream elements and their impact on adaptation. Finally, measurements on a subset of the proposed encryption strategies and a conclusion are provided in Sections V and VI respectively.

II. High Efficiency Video Coding

In HEVC, a picture can be divided in slices, which can be further divided in Coded Tree Blocks (CTB) of size 64x64. These CTBs are divided in square Coding Units (CU) ranging from 8x8 up to 64x64 in a quadtree structure. CUs are divided in Prediction Units (PU) which form the basic unit on which intra or inter prediction is applied. For the inter prediction, reference picture lists are created and the selected reference pictures must be signaled in the video stream.

By subtracting the predicted pixels from the original ones, residual information is obtained. On this residual data, Transform Units (TU) are divided. After transformation, quantization is applied using a Quantization Parameter (QP) and a quantization matrix which can both be signaled in the video stream.

After the reconstruction process, three loop filters can be applied, including the deblocking filter, sample adaptive offset and adaptive loop filter. The entire encoding loop as described here is visualized in Fig. 1.

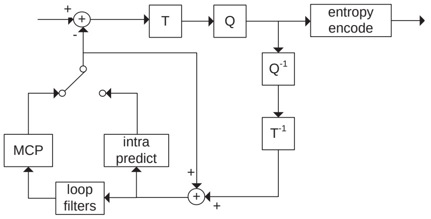


Fig. 1. HEVC encoding loop indicating quantization (Q), transformation (T), and motion compensated prediction (MCP). Although the individual tools are different from H.264/AVC, on an abstract level the closed-loop predictive coding paradigm of HEVC corresponds to that of H.264/AVC.

III. Video Adaptation

The most flexible, but at the same time most complex video adaptation technique is to decode the video stream and apply the adaptation in the pixel domain. For this transcoder to work, it must be trusted with the decryption key because a full reconstruction of the video stream is made. With this technique, for example, resolution, frame rate, bit depth, or random access period adaptations can be provided.

At the other extreme, scalable coding is another possibility offering adaptation capabilities. Because only temporal scalability can be applied to the current HEVC specification,

The research activities that have been described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), Ph.D. and post-doctoral fellow grants of the Agency for Innovation by Science and Technology (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union. Furthermore, this work was carried out using the Stevin Supercomputer Infrastructure at Ghent University.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

spatial or quality scalability will not be considered. Temporal scalability can be obtained by hierarchically coding bidirectionally predicted pictures. Video stream adaptation can then be applied by low complex picture dropping operations.

An efficient alternative solution is given by compressed domain quality transcoding algorithms [3]. With these techniques, the video stream is decoded until the residual information is accessed, and no full reconstruction of the video being necessary. The residual information is requantized at a lower quality and is then merged back in the video stream.

IV. Encryption Strategies for HEVC

When encrypting bitstream elements from an encoded video stream, compatibility with the video standard should be maintained. This is important because devices handling the video stream on the network should not be aware of the applied encryption mechanism. This imposes the restriction that encryption can only be applied on bitstream elements that do not alter the entropy decoding process of other bitstream elements. For example, encrypting the prediction mode may change it from inter to intra prediction. The entropy decoder would expect intra information instead of inter information. Consequently, it will get out of sync with the real encoded elements and unpredictable behavior will follow.

TABLE I
Independent Bitstream Elements in HEVC

- Short-term reference picture set
- Scaling list coefficients
- QP information (initial QP, chroma delta QP, slice delta QP, CU delta QP)
- Intra information (intra luma/chroma prediction flag)
- Inter information (reference picture indices, motion vector prediction indices, motion vector differences)
- Residual information
- Deblocking filter parameters
- Sample adaptive offset parameters
- Adaptive loop filter parameters

Within the HEVC specification, nine potential sets of independent information are identified in Table I. None of these elements influence adaptation processes as described in Section III, except for the QP information and the residual information. With compressed-domain quality transcoding, residual information gets requantized at a lower quality. Therefore, residual coefficients should not be encrypted. In general, requantization transcoders only need absolute residual coefficients, so the sign information of the residual can still be encrypted. The QP information about the residual is used during the requantization process and the new QP should be signaled in the bitstream. Therefore, QP information should be readable and adaptable by a quality transcoder.

In this paper, we propose to encrypt three bitstream elements from this list, namely intra prediction mode difference, motion vector difference sign, and residual sign. The intra prediction mode difference must be seen as the mode that gets signaled after the most probable mode prediction. Similarly, the motion vector difference is the value indicating the difference between the predicted motion vector and the real one. These elements are chosen because it is expected that they provide a large impact on visual quality when the decryption key is unknown by the decoder. Additionally, in HEVC these elements are encoded with equal probability assumption. Therefore, the bits are not entropy encoded, but directly inserted in the bitstream. Consequently, it is expected that by encrypting and therefore changing the values of these bitstream elements, there will be no impact on the final bitrate.

V. Results

The proposed encryption of intra prediction modes, motion vector difference signs, and residual signs is implemented in HEVC reference Model (HM) v6.1. The test is conducted on 22 test sequences (8-bit), as used during the HEVC standardization process. The GOP (Group Of Pictures) size is set to eight and a random access period of approximately one second is applied. QP values of 22, 27, 32, and 37 are used on every sequence. With these four test points, Bjøntegaard delta (BD) bitrate measurements are calculated, indicating the overall bitrate increase over the entire PSNR test range.

BD-bitrate measurements indicate a 0.0% BD-rate increase when encrypting the intra prediction mode difference, motion vector difference signs, and residual signs. This corresponds to our decision to encrypt these elements because of their equal probability property.

To illustrate the capabilities of the encryption algorithm, an example decoded picture by a decoder without decryption key is given in Fig. 2.

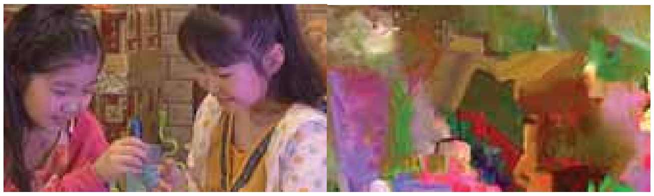


Fig. 2. Original BlowingBubbles sequence(left) and version decoded by an untrusted decoder (right).

VI. Conclusion

In this paper, we indicate potential bitstream elements that can be used for HEVC compatible encryption. A description is given about which elements can have an influence on video adaptation processes deployed in the network. Additionally, three elements were selected to be encrypted (intra prediction mode difference, motion vector difference sign, and residual sign) based on the fact that they have a significant visual impact when decoded without the decryption key and because they do not impact compression efficiency.

ReferenceS

[1] S. Lian, Z. Liu, Z. Ren, and H. Wang, "Secure advanced video coding based on selective encryption algorithms," *IEEE Transactions on Consumer Electronics*, vol.52, no.2, pp. 621- 629, May 2006.

[2] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, T. Wiegand, "High efficiency video coding (HEVC) text specification draft 7" *9th JCT-VC Meeting*, Geneva, CH, May 2012.

[3] J. De Cock, S. Notebaert, P. Lambert, R. Van de Walle, "Requantization transcoding for H.264/AVC video coding", *Signal Processing Image Communication,* vol.25, no. 4, pp. 235-254, Apr. 2010.

High-performance HOG Feature Extractor Circuit for Driver Assistance System

Seonyoung Lee, Haengseon Son, Jong-Chan Choi, and Kyungwon Min
SoC Platform Research Center, Korea Electronics Technology Institute, Korea

Abstract—**In this paper, we propose the architecture of high-performance HOG feature extractor circuit for the driver assistance system. We have proposed the simplified hardwired methods for the square root of the gradient calculation and the division operation of the block normalization. Our HOG feature extractor circuit can process 37 frames per seconds for 640x480 VGA images in real-time.**

I. INTRODUCTION

Video-based smart vehicle technology such as advanced driver assistance systems (ADAS) becomes an important method in the automobile industry. Typical applications using images in ADAS are pedestrian and vehicle recognition system for safety. These system detect pedestrian and vehicle in images, and measures its distance from the object. These detection results prevent an accident that may occur while driving. In general, object recognition uses a variety feature which can be obtained from images and the sliding window-based methods to obtain better recognition performance. Paper [1] proposed the histogram of orientation gradient (HOG) method to clearly distinguish pedestrians from images. This method can be obtained good recognition results through applying the robust feature set based on gradient and computes locally normalized gradient orientation histograms over blocks of size 16x16 to represent a detection window. When the block histograms within the window are concatenated, the resulting feature vector is powerful enough to classify humans with 88% detection rate at 10^{-4} false positives per window (FPPW) using a linear SVM (support vector machine) [2].

Although the HOG have a good recognition performance, the huge computational complexity remains the problem of the processing speed. Because HOG uses the object detection method based on the sliding window, an image has a different scale of tens of thousands windows. In addition, there are thousand dimensional features in each sliding window. Some papers show the good recognition performance and processing speed in a PC environment. However, these methods cannot be applied to the embedded processor-based platforms such as robots and vehicles required the real-time object recognition. There have been proposed various methods to solve this problem. One uses the graphics processing unit (GPU) which can be processed in parallel. These methods are much faster than the central processing unit (CPU) implementation (Maximum 40 times). Another uses FPGA+GPU or GPU+ CPU devices. The cyclic operation of large amount is processed on the GPU or FPGA and non-linear operation is processed on the CPU. However, these methods are high-performance PC-based platform environment with GPU or CPU, and are not suitable for embedded platforms.

In this paper, we propose the architecture of high-performance HOG feature extractor circuit for the driver assistance system. Our HOG feature extractor circuit extracts features from the image data for every sliding window. Our implemented HOG feature extractor circuit supports weighted gradient value, 2-dimensional (2D) histogram interpolation, block normalization, and variable sliding window sizes of 32x32, 48x48 and 64x64 for pedestrian recognition, and 32x64, 48x96 and 64x128 for vehicle recognition. Also, we have used the simplified hardwired circuit implementation methods of the square root for the gradient calculation and the division and square root operation for the block normalization. Our HOG feature extractor has been verified in the FPGA development board and can process 37 frames per second for 640x480 VGA images in real-time.

II. HARDWARE ARCHITECTURE

Our proposed circuit architecture for HOG feature extraction consists of some modules, as shown in Fig. 1. Each module runs in parallel. ‘Line Buffer’ inputs image data from external frame buffer memory and passes pixel data to calculate the gradient operation. ‘Line Buffer’ has two 64x8-bit line buffers with respect to the horizontal size of sliding window. ‘Gradient Calculator’ consists of ‘Difference Calculator’, ‘Magnitude Calculator’, and ‘Bin Calculator’. ‘Difference Calculator’ calculates the difference for *x* and *y* direction, respectively. ‘Magnitude Calculator’ obtains the gradient magnitude through calculating the square root operation for the difference values of the previous module. ‘Bin Calculator’ determines the bin index to calculate histogram values by using the difference of the ‘Gradient Calculator’. ‘Magnitude & Bin Buffer’ stores the gradient magnitude and the bin index into single-port SRAM buffers. It changes the line-based operation to the cell-based operation for the histogram calculation. This module uses sixteen 128x64-bit single-port SRAM buffers. ‘Weight Calculator’ calculates the gradient magnitude values weighted by a Gaussian function for one block. The value of Gaussian coefficients is provided by ‘Weight Coefficient ROM’. ‘Histogram Generator/Interpolator’ generates histogram and performs the 2D histogram interpolation in order to reduce the aliasing effect between cells within one block. ‘Normalization

This work was supported by the grant from the Industrial Strategic Technology Development Program (KI002162) and the Industrial Convergence Strategic Technology Development Program (10041656) of the Ministry of Knowledge Economy (MKE) of Korea.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

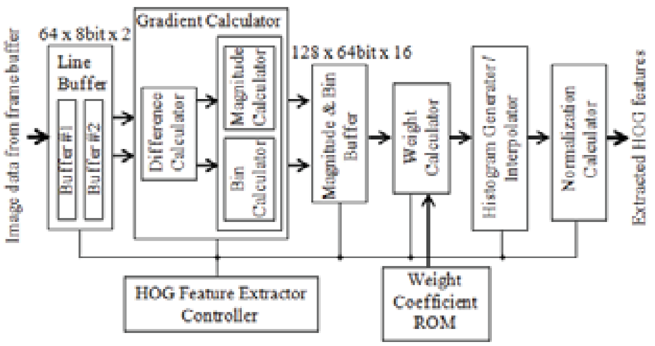


Fig. 1. Block diagram of our proposed HOG Feature Extractor.

Calculator' uses L2-*norm* to increase the robustness of histogram for four cells in one block. 'HOG Feature Extractor Controller' controls each block of 'HOG Feature Extractor' circuit. The proposed architecture supports six sliding window sizes. The cell size of 4, 6, and 8 pixels can be processed. In our circuit, it requires one block (4 cells) per one cycle to extract the HOG feature in real-time for 640x480 VGA image.

HOG feature extraction has some operation that is difficult to implement the hardwired logic. Many studies have tried to hardware simplification. However, they have many errors in its methods. We have proposed the simplification methods for the gradient magnitude and L2-*norm* which have small calculation errors. Gradient magnitude calculation uses the square root for the sum of squares. We implemented the hardwired square root calculator using the approximation of paper [3]. This method is eliminating a computational complexity for a square root operation at a maximum 4% error compared with the exact magnitude calculation. Because this is so big error values, we have applied the piecewise linear method. In this method, we have a maximum 0.7537% error compared with the exact calculation. We use the L2-*norm* to process the block normalization. Because this equation includes a square root and division operation, it is difficult to implement the hardware. To implement the hardware of square root and division, we have simplified these operations. Paper [4] assumes the denominator $\sqrt{\|V_{k,l}\|^2+\varepsilon^2}$ into $\|V_{k,l}\|$. Instead, we use the simplified square root and division method for the denominator calculation. We can reduce an error which is a maximum 0.4678% for the normalization operation.

III. IMPLEMENTATION RESULT

Our proposed HOG feature extractor circuit was designed at register transfer level using Verilog hardware description language. The synthesized circuit using 0.13um CMOS standard cell library consists of 1,009,729 gates and has the maximum operating frequency of 211MHz. Our implemented circuit requires 352 cycles to process one sliding window of 32x32 pixels. One cell has 9 bins and one sliding window has 576 or 1,152 HOG features according to the size of sliding window. For the 640x480 VGA image has 17,024 sliding windows, our circuit extracts the HOG features of 36.92 frames per second at 211MHz. We have used the weighted gradient, 2D interpolation and the L2-*norm* for the block normalization. Fig. 2 shows a difference for the square root with the sum of squares to calculate the gradient magnitude. Fig. 3 shows an error for the square root and division operation.

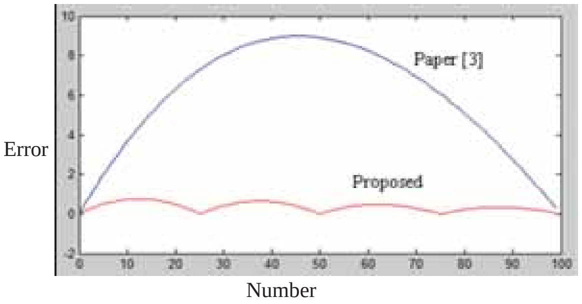


Fig. 2. Square root error for the sum of squares.

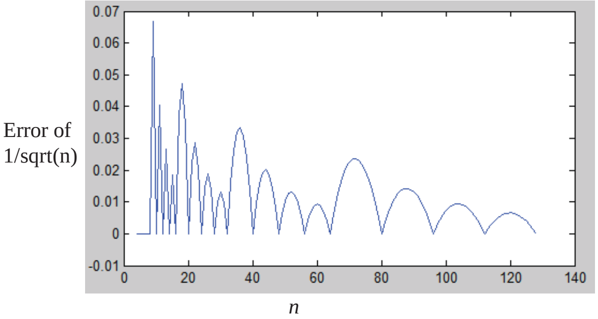


Fig. 3. Error for the square root and division.

IV. CONCLUSION

In this paper, the architecture of high-performance HOG feature extractor circuit for the driver assistance system. We have applied the simplified square root and division operation circuit which can minimize an error. Our circuit can support various window sizes and cell sizes to enable the hardware expansion. It can process 37 640x480 VGA frames per second to extract HOG features. Because our proposed circuit can operate in real-time, it can apply to the autonomous robot systems, the security surveillance systems, and the human and vehicle detection system required the real-time processing.

REFERENCES

[1] N. Dalal and B. Triggs, "Histogram of Oriented Gradients for Human Detection," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 886–893, Apr. 2005.

[2] B. Bilgic, B.K.P. Horn, and I. Masaki, "Fast human detection with cascaded ensembles on the GPU," *IEEE Intelligent Vehicles Symposium*, pp. 325–332, June 2010.

[3] T. Wilson, M. Glatz, and M. Hodlmoser, "Pedestrian detection implemented on a fixed-point parallel architecture," *International Symposium on Consumer Electronics*, pp. 47–51, May 2009.

[4] K. Negi, K. Dohi, Y. Shibata and K. Oguri, "Deep pipelined one-chip FPGA implementation of a real-time image-based human detection algorithm," *IEEE Conference on Field-Programmable Technology*, pp. 1–8, Dec. 2011.

Speech Enhancement by Kalman Filtering with a Particle Filter-Based Preprocessor

Yun-Kyung Lee, Gyeo-Woon Jung, and Oh-Wook Kwon, *Member, IEEE*

***Abstract*—To reduce nonstationary noise in real environments, we propose to use a particle filter as a preprocessor of Kalman filtering. From noisy input speech signals, the autoregressive (AR) model parameters are estimated by using a particle filter. Clean speech signal is estimated by a Kalman filter configured with the estimated parameters. Experimental results show that when speech signal is corrupted by babble noise, the proposed algorithm improves the output SNR by 1.5 dB.**

I. Introduction

A Kalman filter is an effective algorithm to enhance speech signals from a series of measurements observed over time, containing random noise and other inaccuracies. The Kalman filter achieves a faster convergence behavior than a normalized least-mean-square (NLMS)-based adaptive filter. There have been numerous studies on Kalman filtering for speech enhancement [1]. Even though noise observed in real situations has a nonstationary and dynamic feature, previous studies on Kalman filter were mostly applied by using the stationary white Gaussian noise assumption for simplicity.

We present a sequential nonstationary speech enhancement method using the Kalman filtering combined with a particle filter [2] to estimate the parameters of speech signal and the variance of nonstationary additive noise. The sequential importance sampling (SIS) is used to estimate the parameters of the particle filter and clean speech signal is estimated by the particle filter in a frame-wise manner and is applied to a Kalman filter. In this work, speech signal is modeled as an autoregressive (AR) process. The noise variance and the parameters of the AR process are estimated in the Kalman filter. Our experimental results shows that the proposed Kalman filtering with a particle filter leads to significant signal-to-noise ratio (SNR) gain and improves the speech quality remarkably.

II. System Description

A. Particle filter-based parameter estimation

In the general formulation of the state estimation problem, the objective is to track the time evolution of the filtering density. If we assume that the parameters of speech and noise signal are known, the optimal estimate of the original speech signal can be obtained from a Kalman filter. However, in realistic scenarios, the background noise signal is unknown and the noise sources are mostly nongaussian. We use a particle filter in order to estimate the parameters (AR coefficients and noise variance) of the Kalman filter in nongaussian noise environments. The particle filter sequentially updates the filtering density under a relaxed Gaussian assumption. We estimate the parameters by using the SIS algorithm, which is summarized as [2]:

1. Sample $x_i^{(m)} \propto p(x_i \mid x_{i-1}^{(m)}) \forall m$
2. Compute the weights $w_i^{(m)} \propto p(y_i \mid x_i^{(m)}) \forall m$
3. Normalize $\tilde{w}_i^{(m)} = w_i^{(m)} / \sum_{k=1}^{K} w_i^{(k)} \forall m = 1,\ldots,M$
4. Compute the new filtering density
   $p(x_i \mid y_{1:i}) \approx \sum_{m=1}^{M} \tilde{w}_i^{(m)} \delta(x_i - x_i^{(m)})$
5. Resample to obtain M new equally-weighted set of particles.

Resampling has the effect of removing particles with low weights and amplifying particles with high weights. Accordingly, the posteriori probability distribution of the resampled particles has a sharper distribution. The concept of the above particle filtering process is visualized in Fig. 1.

From the speech signal estimated in the particle filter, the speech and noise parameters are computed through linear predictive coding (LPC). The estimated speech signal from the particle filter can be described by the p-th order AR model and the state transform matrix $\tilde{F}$ is defined as:

$$\tilde{F} = \begin{bmatrix} \tilde{l}_1 & \tilde{l}_2 & \ldots & \tilde{l}_{p-1} & \tilde{l}_p \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

where $\tilde{l}_1,\ldots,\tilde{l}_p$ are LPC coefficients.

B. Estimation of clean speech by using the Kalman filter

After the speech signal is estimated in the first stage by using the particle filter, we compute the AR parameters and noise variance. Given these parameters, the final clean speech signal is extracted with a Kalman filtering process. Let $\tilde{v}_i$

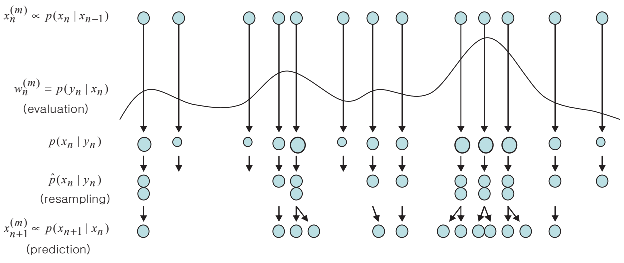

$x_n^{(m)} \propto p(x_n \mid x_{n-1})$
$w_n^{(m)} = p(y_n \mid x_n)$
(evaluation)
$p(x_n \mid y_n)$
$\hat{p}(x_n \mid y_n)$
(resampling)
$x_{n+1}^{(m)} \propto p(x_{n+1} \mid x_n)$
(prediction)

Fig. 1. Concept of the particle filtering process [2].

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0001730).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

denote the estimated observation (measurement) noise: $\tilde{v}_i = y_i - s_i$, where y_i is the observed speech signal and s_i is the original speech signal. Let $\tilde{V}$ denote the covariance matrix of estimated measurement noise $\tilde{v}(n)$ obtained from the result of the particle filtering process. Then the Kalman filter is applied as follows.

$$\begin{aligned}
&\hat{x}_{i|i-1} = \tilde{F}_{i|i-1}\hat{x}_{i-1|n-1} && \text{(Prediction)}\\
&K_{i|i-1} = \tilde{F}_{i|i-1}K_{i-1}\tilde{F}_{i-1|i-1}{}^{T} + U_{i+1} &&\\
&G_i = K_{i|i-1}H_i{}^{T}\left[H_i K_{i|i-1}H_i{}^{T} + \tilde{V}_i\right]^{-1} &&\\
&s_i = y_i - H_i\hat{x}_{i|i-1} &&\\
&\hat{x}_{i|i} = \hat{x}_{i|i-1} + G_i s_i && \text{(Correction)}\\
&K_i = (I - G_i H_i)K_{i|i-1} &&
\end{aligned}$$

In the above, where $\hat{x}_i$ is the predicted state estimate, $K_{i|i-1}$ is the predicted state-error covariance matrix, K_i is the filtering-error covariance matrix, U_i is the covariance matrix of process noise, H_i is the observation matrix and G_i is the Kalman gain.

III. Experimental Results

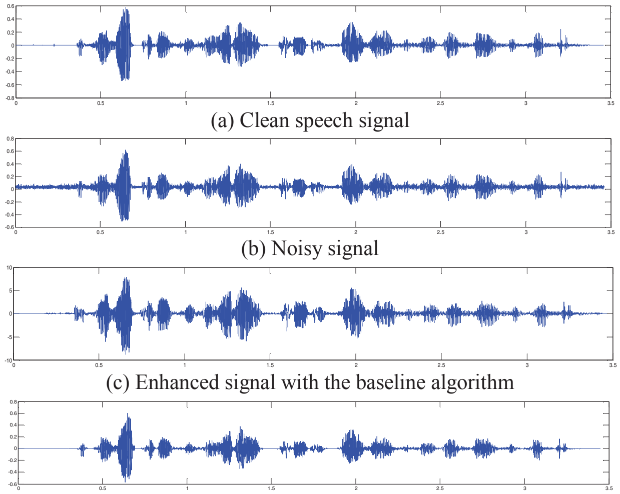
We performed computer experiments to evaluate the proposed algorithm in various noise environments, by using the database of the Speech Separation Challenge [3]. The added noise sources are three types: N1 (car noise), N2 (babble noise), and N3 (white Gaussian noise). The sampling rate of speech database was lowered from 25 kHz to 16 kHz. The noisy speech signal was generated by mixing clean speech with the noise sources at −10, −5, 0, 5, 10 dB SNRs. Note that N1 and N3 are stationary noise but N2 noise has a nonstationary nature.

Fig. 2 shows the waveforms of the clean, the noisy, and the enhanced speech signals, from top to bottom. In the figure, the noisy signal was corrupted with the N2 (babble) noise with input SNR=0 dB. We confirmed that noise was suppressed remarkably to yield enhanced speech signal.

We also computed the output SNR (dB) of enhanced speech signal. Table I compares the output SNR with respect to speech signals under the three noise conditions by using the Kalman filter with an NLMS adaptive filter-based preprocessor (Baseline) and the Kalman filter with a particle filter-based preprocessor (Proposed), respectively. The proposed algorithm provides the average SNR increase of 2.7 dB, 1.5 dB, and 0.5 dB under the N1, N2, and N3 noise conditions, respectively. From these results, it is justified that our algorithm significantly improves the objective quality measure in nonstationary environments as well as in stationary environments.

IV. Conclusions

We proposed a Kalman filter-based speech enhancement algorithm where a particle filter is used as a preprocessor to estimate the Kalman filter parameters in nonstationary noise conditions. In computer experiments with artificially-mixed noisy speech signals, the proposed algorithm achieved the improvement of the output SNR by 2.7 dB, 1.5 dB, and 0.5 dB for the car noise, babble noise, and white Gaussian noise conditions, respectively.


(a) Clean speech signal

(b) Noisy signal

(c) Enhanced signal with the baseline algorithm

(d) Enhanced signal with the proposed algorithm

Fig. 2. Sample waveforms from out computer experiments.

TABLE I
Output SNR (dB) under different noise conditions

Noise Type	Algorithm	Input SNR(dB)					Average
		-10	-5	0	5	10	
N1	Baseline	1.7	2.9	4.1	5.5	6.3	4.1
	Proposed	3.8	5.2	7.5	8.4	9.1	6.8
N2	Baseline	1.5	3.8	5.2	6.2	6.9	4.7
	Proposed	2.9	4.7	7.1	7.9	8.6	6.2
N3	Baseline	3.0	2.1	2.6	4.1	5.8	6.0
	Proposed	3.4	4.6	7.6	7.8	8.8	6.4

References

[1] D.C. Popescu and I. Zeljkovic, “Kalman filtering of colored noise for speech enhancement,” *Proc. ICASSP*, pp3 997-1000, 1998.

[2] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” *IEEE Trans. Signal Processing*, vol. 50, no. 2, Feb 2002.

[3] M.P. Cooke, J. Barker, S.P. Cunningham, and X. Shao, “An audio-visual corpus for speech perception and automatic speech recognition,” *J. Acoust. Soc. Am.*, vol. 120, issue 5, pp. 2421-2424, Nev., 2006.

Graph-Based Parallelization Algorithm for Deblocking Filter in H.264/AVC

Seongmin Jo and Yong Ho Song, *Member, IEEE*
Department of Electronics and Communication Engineering, Hanyang University, Seoul, Korea

***Abstract*—Wavefront parallelization speeds up the deblocking filter in multi-core platforms, but it has a limitation in achieving sufficient parallelism owing to data dependency among macroblocks. To overcome this limitation, we propose a novel approach of parallelizing the deblocking filter by exploiting the conditional data dependency affected by the boundary strength [1]for each macroblock edge.**

I. INTRODUCTION

H.264/AVC standard [1] introduced the in-loop deblocking filter (DF) to reduce blocking artifacts caused by quantization, as well as to increase coding efficiency. However, this in-loop filter also increases the computational complexity of the H.264/AVC decoder [2]. The recent increase in video resolution up to full high-definition (FHD) or ultra-definition (UD) requires even more computation efficiency for DF, motivating the utilization of parallelization of multi- or many-core platforms.

The macroblock (MB) reference sequence and inter-MB dependency in DF operation hinder the achievement of the maximum parallelism with multi-core processors: the DF works on MBs in a raster scan order to modify pixels near vertical and horizontal block edges, which creates inter-MB dependency between adjacent MBs. Therefore, when DF starts to process an MB, it must have completed processing the upper and left neighbor MBs. Wavefront parallelization [3] takes into consideration all the MB dependencies in such a way that each processor core filters a row of MBs after the dependencies have been resolved (Figure 1(a)). However, the wavefront approach has two limitations: low core utilization and insufficient parallelism.

This work has been motivated by the observation that DF conditionally filters block edges based on the boundary strength. In some cases, the data dependency that often resides at MB edges does not exist, which allows processor cores to handle these MBs independently. In this paper, we present a graph-based parallel DF (GPDF), which exploits this independency to achieve high DF performance of processor cores. The proposed technique achieves a considerably higher speedup than the wavefront parallelization scheme for B frames in which about 50% of MB edges are not filtered owing to the boundary strength condition. Although it fails to explore the independency for I and P frames, GPDF exhibits no performance penalty for these frames when compared to the wavefront scheme.

II. GRAPH-BASED PARALLEL DEBLOCKING FILTER

For each MB, the DF in H.264/AVC applies finite impulse response filters to the vertical and then horizontal edges. This filtering order creates data dependencies between adjacent MBs, as shown in Figure 1(a), which limits the degree of parallelization of DF on an MB basis. The data dependency later triggers wavefront parallelization [3], which has been implemented in the recent reference software [5]. The wavefront approach concurrently performs DF for MBs in a skewed diagonal in order to not violate the data dependencies, as shown in Figure 1(a). However, there are not sufficient MBs to fully utilize the available processing cores during the execution [3], as shown in Figure 1(b). In addition, the maximum parallelism of the wavefront method is limited to video resolution; in the case of an FHD video, at most 60 MBs can be filtered concurrently in the wavefront scheme [4].

To overcome the low core utilization and limited parallelism in the wavefront method, our proposed GPDF exploits the feature that the DF conditionally filters edges depending on their boundary strength (BS) values [1]. A BS value of a block edge is derived from the coding information of two adjacent blocks, such as a prediction mode. When the BS value is 4, a strong filter is applied to the edge; when the BS value is 2, 3, or 4, a normal filter modifies pixels near the edge. However, if the BS value is 0, DF does not filter the edge; this occurs when no data dependency exists between adjacent MBs.

In our investigation, approximately 50% of MB edges in the B frames have a BS of 0, whereas in P frames approximately 10% have a BS of 0. The data independency enabled by a BS of 0 allows GPDF to exploit more parallelism in the performing DF. Unfortunately, all the edges in I frames have a BS value of 3 or 4, implying that GPDF is not effective for I frames. Even in this case, the parallelism of GPDF is equal to that of the wavefront scheme.

The utilization of such a dependency condition for parallelization requires the management of data dependencies of MBs in a frame. To achieve this, GPDF builds a MacroBlock Dependency Graph (MBDG) for a frame by checking all the BS values of MB edges before parallelization. An MBDG is a chained collection of four-bit values for each MB; each bit indicates the existence of data dependency on the left-upper, upper, right-upper, and left MB. Figure 1(c) is an example where only the bold MB edges are filtered. In this example, MBDG can be built by GPDF, as shown in Figure 1(e). The arrow in the figure denotes the required dependency

"This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2012-H0301-12-1011)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

order in filtering between MBs. For example, MB 14 should be filtered after MB 6. After building the MBDG, GPDF allocates independent MBs to each core that has no incoming arrow in MBDG. For MBs that have been filtered, outgoing arrows are removed in MBDG by reversing corresponding bits to make the MBs subsequent to the completed MBs ready to be scheduled on processing cores. Figure 1(d) shows the overall allocation of MBs on a four-core processor for the example frame, which demonstrates that our GPDF scheme completely utilizes the cores.

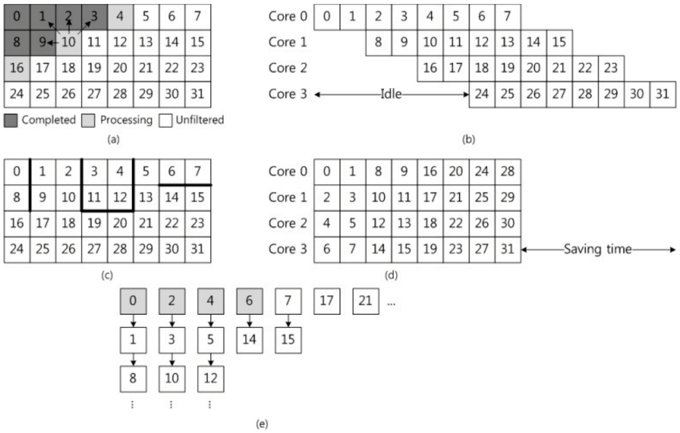

0 1 2 3 4 5 6 7
8 9 10 11 12 13 14 15
16 17 18 19 20 21 22 23
24 25 26 27 28 29 30 31
■ Completed □ Processing □ Unfiltered
(a)
Core 0 0 1 2 3 4 5 6 7
Core 1 8 9 10 11 12 13 14 15
Core 2 16 17 18 19 20 21 22 23
Core 3 Idle 24 25 26 27 28 29 30 31
(b)
0 1 2 3 4 5 6 7
8 9 10 11 12 13 14 15
16 17 18 19 20 21 22 23
24 25 26 27 28 29 30 31
(c)
Core 0 0 1 8 9 16 20 24 28
Core 1 2 3 10 11 17 21 25 29
Core 2 4 5 12 13 18 22 26 30
Core 3 6 7 14 15 19 23 27 31 Saving time
(d)
0 2 4 6 7 17 21 ...
1 3 5 14 15
8 10 12
(e)

Fig. 1. (a) Wavefront parallelization, (b) execution sequence of wavefront method, (c) example frame with filtering of several MB edges, (d) execution sequence of GPDF, and (e) MBDG for the example frame

III. Experiment Results

To evaluate the GPDF method, we developed a simulator and used the BS values extracted from reference software as input [5]. We assumed that the operation time of DF is the same for all MBs. Owing to space constraints, we present the results for only the *rush hour* video sequence [6] compressed with a hierarchical B-picture structure [7]. However, a similar trend has been observed for other video sequences.

Figure 2 shows the speedups of GPDF and wavefront schemes when the number of processing cores is varied from 1 to 128. The wavefront parallelization shows almost the same performance for all frame types because it considers all the data dependencies regardless of the frame type. On the other hand, GPDF shows different speedups owing to different BS distributions depending on frame types.

As aforementioned, the speedup of wavefront schemes does not scale well due to low core utilization. In addition, the theoretical limit of the utilizable number of cores limits the speedup of the wavefront method over 60 cores, whereas GPDF scales well beyond 60 cores in the B frames. However, as expected, the speedups of GPDF for I/P frames are close to that of the wavefront method, because I/P frames contain no or few MB edges with a BS of 0. Although the GPDF algorithm does not greatly improve the speedup for I/P frames, it is still effective in an encoded video with many B frames [7].

Figure 3 shows the speedup of GPDF for each frame in the *rush hour* video when 32 processing cores are used. We draw two lines on the figure as a guideline for speedups achievable from the wavefront and ideal parallelization. As shown in the figure, GPDF outperforms the wavefront scheme for all frame types. In addition, GPDF yields a speedup that is close to the case of perfect parallelization for many frames.

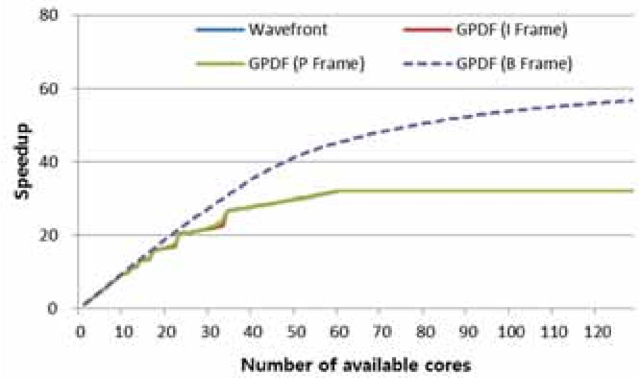


Fig. 2. Speedups of wavefront method and GPDF

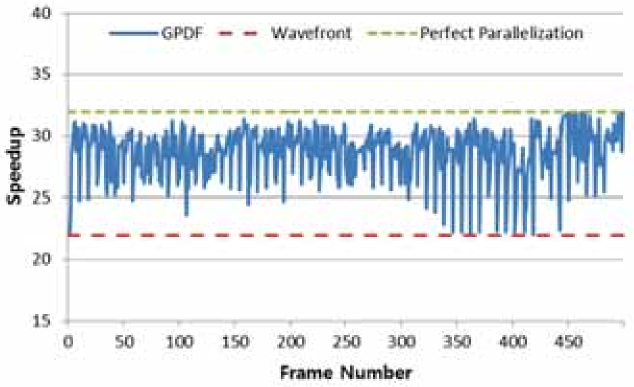


Fig. 3. Speedups of wavefront method and GPDF for each frame in *rush hour* video sequence

IV. Conclusion

This paper proposed a novel parallelization scheme applicable for the deblocking filter operation in H.264/AVC, and it exploits the conditional data dependencies determined by boundary strength. This scheme effectively overcomes the limitations of low core utilization and maximum parallelism compared to wavefront parallelization, especially in B frames.

References

[1] Joint Video Team (JVT) of ITU-T and ISO/IEC JTC 1, “Advanced video coding for generic audiovisual services,” ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), version 5, Jul. 2007.

[2] Sung-Wen Wang, et al., “A multi-core architecture based parallel framework for H.264/AVC deblocking filters,” *J. Signal Process. Syst.*, vol. 57, pp. 195, 2009.

[3] C. Meenderinck, A. Azevedo, M. Alvarez, B. Juurlink, and A. Ramirez, “Parallel scalability of H.264,” In MULTIPROG Workshop, Jan. 2008.

[4] Seongmin Jo and Yong Ho Song, “Exploring parallelization techniques based on OpenMP in H.264/AVC encoder for embedded multi-core processor,” *J. of Syst. Arch.*, submitted for publication.

[5] JM 18.3, http://iphome.hhi.de/suehring/tml/download.

[6] *Rush hour* video sequence, http://nsl.cs.sfu.ca/video/library/tu-muenchen.de/1080p_rush_hour.yuv.

[7] Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, “Analysis of hierarchical B pictures and MCTF,” in *Proc. IEEE Intl. Conf. Multimedia and Expo*, pp. 1929-1932, 2006.

Energy Efficient Video Decoding for the Android Operating System

Wen-Yew Liang[1], Ming-Feng Chang[1], Yen-Lin Chen*[1] and Chin-Feng Lai[2]

[1]Department of Computer Science and Information Engineering National Taipei University of Technology
Taipei 106, Taiwan
[2]Institute of Computer Science and Information National Ilan University, Ilan 260, Taiwan
Email: wyliang@mail.ntut.edu.tw, t8599009@ntut.edu.tw, *ylchen@csie.ntut.edu.tw, cinfon@ieee.org

Abstract-- **Dynamic voltage and frequency scaling (DVFS) is an effective technique for reducing power consumption. Due to the increasing popularity of multimedia applications for portable consumer electronic devices, the importance on reducing their power consumption becomes significant. This paper proposed a table-based DVFS mechanism for frame decoding that reduces the power consumption of a processor by exploiting the frame decoding complexity. A table-based DVFS predictor is used in the frame decoding prediction. This study was implemented in the Android operating system on an Intel PXA27x embedded platform. Experiment results showed that the energy consumption of decoding videos can be reduced from 9% to 17%, whereas the frame drop-rate is less than 3%.**

I. INTRODUCTION

Portable devices have been widely used in people's daily life nowadays. Among the design issues, energy consumption is the most important one. Dynamic voltage and frequency scaling (DVFS) is one of the techniques which can be used to extend battery durability in portable devices by dynamically scaling the frequency and voltage of the processors according to the computing complexity of the running applications.

Since multimedia applications, which are typically associated with higher computing complexity and thus consume more power, are getting more and more popular in the portable devices, conserving the battery energy becomes an important issue. Some previous studies [3, 4] have focused on analyzing the frame decoding complexity and predicting the suitable frequencies and voltages. However, these approaches focus on single codec and require knowledge of video decoding. Instead, this paper proposes a table-based DVFS mechanism, called Frame Table-based DVFS (FT-DVFS), for frame decoding which adapts to various types of video decoders. Fig. 1 illustrates the advantage of DVFS in frame decoding. Fig. 1(a) shows the processor activity when DVFS is not used. Once a frame has been decoded into a buffer, the processor is switched to the waiting mode until the next frame needs to be decoded. Fig. 1(b) shows the ideal DVFS where the processor is scaled to a lower frequency and voltage under the video playback timing constraint.

The goal of this paper is to predict the frame decoding time so that a proper frequency and voltage can be determined to reduce the energy consumption. The complexity of the frame decoding process are evaluated and recorded. The table-based predictions [1,2] are then used to predict the decoding time for the next frame, based on the complexity information. An appropriate frequency and voltage can thus be evaluated based on the predicted decoding time. This method has been realized in the Android operating system on an Intel PXA27x embedded platform. The experiment used VLC media player to decode the MPEG video contents.

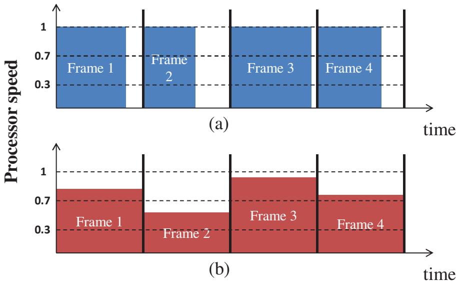


Figure 1. (a) Frame decoding without DVFS. (b) Frame decoding with DVFS.

II. THE PROPOSED TABLE-BASED DVFS FOR FRAME DECODING

To evaluate the complexity of the decoding process, run-time information of the hardware performance counters is adopted. The advantage of using hardware performance counters is that the decoding behavior is no longer required to be determined beforehand. Therefore, this mechanism can be adapted to various types of video decoders through approximating the best-fitted decoding complexity without knowing the details of the decoding methods. Two hardware performance counters, including the number of instructions executed and the number of cache misses, and one Linux kernel information "task_io_accounting" for the I/O accesses time are used for evaluating the decoding complexity.

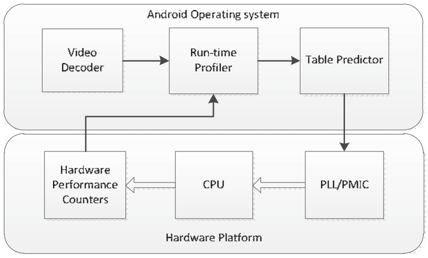


Figure 2. Structure of the table-based DVFS for video decoding prediction.

Fig. 2 illustrates our Frame Table-based DVFS for video decoding prediction. The Run-time Profiler and the Video Decoder provides the following online information: CPU frequency, hardware performance statistics and video decoding time. The Table Predictor records the online information and

Acknowledgments: This work is supported by the National Science Council of the Republic of China under Contract No. NSC-101-2219-E-027-006 and NSC-101-2219-E-027-004

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

predicts a suitable CPU speed for decoding the next frame. The PLL and PMIC were used to support dynamic frequency and voltage adjustment.

Our Table Predictor records the frame decoding complexity related information, CPU frequency and frame decoding time once the Video Decoder has finished decoding a frame. The predictor is divided into two parts, as depicted in Fig. 3: (1) The Frame Decoding Complexity History Table (FDC-HT) collects the decoding complexity information of video frames; (2) Each Decoding Time Tables (DTT) of the FDC-HT entry contains the CPU frequency and the corresponding decoding time. After a frame has been decoded, the Table Predictor searches the decoding complexity at FDC-HT and writes the CPU frequency and frame decoding time to the DTT. If the decoding complexity does not exist, then the FDC-HT will create a new entry and the corresponding DTT for it.

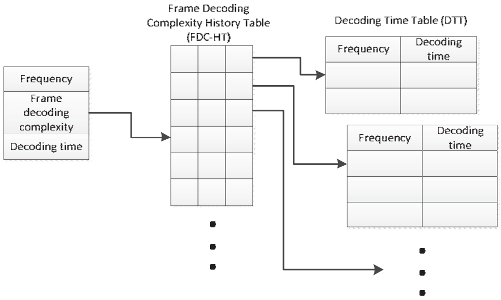

Frame Decoding Complexity History Table (FDC-HT)
Decoding Time Table (DTT)
Frequency
Frame decoding complexity
Decoding time
Frequency
Decoding time
Frequency
Decoding time

Figure 3. The frame decoding table predictor structure.

To predict the next frame decoding complexity W_{t+1}, the GPHT [2] predictor is used. The GPHT predictor uses the pattern from recent history (W_t ~$W_{t\text{-}depth}$) and predicts the next frame decoding complexity W_{t+1} from the global history table. After W_{t+1} has been determined, the Table Predictor searches the it in the FDC-HT and finds the next frequency f_{t+1} with the decoding time (D_{t+1}) which is closest but smaller to the deadline of the frame, from the DTT. The deadline of the frame is determined based on the video's frame-per-second value. Finally, the Table Predictor sends the next frequency f_{t+1} to the PLL/PMIC for adjusting the CPU frequency and voltage.

III. Experiments

The experiments were performed on a real platform, the Creator PXA270 development board, on which we have ported the Android with Linux kernel 2.6.25. Actual energy data were also collected by a high performance data acquisition instrument (DAQ). Other DVFS mechanisms, including Linux on-demand governor [5] and Performance governor were also measured for comparative performance evaluation.

The experiment used the open source VLC media player to decode MPEG files. Six MPEG clips were used. The Picture clip is a low-motion type of video; Shrek is an animation movie; and, others are famous high-motion action movies. The resolution of these videos is 320*240 pixels and the frame rate is set to 24 fps. The initial data of the tables were constructed before the experiment by running several test videos with different frequencies. Fig.4 shows the energy consumption of the videos for the experiment. While the Performance governor always runs at the highest frequency, the Linux On-demand governor reduced the energy consumption by 1% to 3%. For our frame table-based DVFS (FT-DVFS), it further reduced the energy consumption by 9% to 17%. The frame drop-rate is also measured and compared in Fig. 5. From the figure, we can see that the frame drop-rate of FT-DVFS is less than 3%, much better than the result of 4% to 8% drop-rate for the Linux On-demand governor.

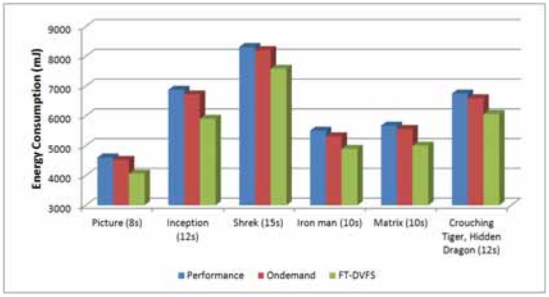


Figure 4. Energy consumption of the tested videos.

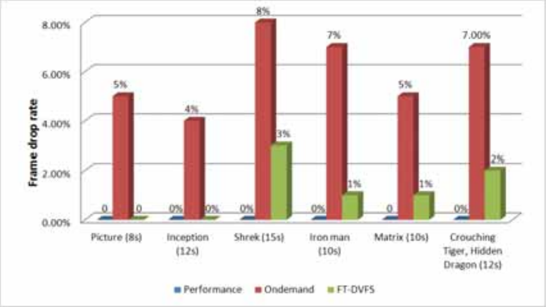


Figure 5. Frame drop rate of the tested videos

IV. Conclusion

This paper provides a practical DVFS method in video decoding for portable consumer electronic devices. However, the FT-DVFS mechanism is currently only specific for video decoding. Our future work will be to combine the mechanism within the power management subsystem in the commodity products. We will realize FT-DVFS in other Android-enabled portable consumer electronic devices, so as to further evaluate the efficiency of the energy-saving mechanism.

Reference

[1] T. Sherwood, S. Sair and B. Calder, "Phase tracking and prediction," *in Proceedings 30th Annual International Symposium on Computer Architecture*, pp. 336- 347, June 2003

[2] C. Isci, G. Contreras, M. Martonosi, "Live, Runtime Phase Monitoring and Prediction on Real Systems with Application to Dynamic Power Management," *in proc. 39th Annual IEEE/ACM International Symposium on Microarchitecture*, Dec. 2006, pp. 359-370.

[3] Z. Ma, H. Hu and Y. Wang, "On Complexity Modeling of H.264/AVC Video Decoding and Its Application for Energy Efficient Decoding," *IEEE Transactions on Multimedia*, vol.13, no.6, pp.1240-1255, Dec. 2011

[4] B. Lee, E. Nurvitadhi, R. Dixit, C. Yu and M. Kim, "Dynamic voltage scaling techniques for power efficient video decoding," *Journal of Systems Architecture*, Vol. 51, Issues 10–11, October–November 2005, pp. 633-652

[5] V. Pallipadi and A. Starikovskiy, "The Ondemand Governor," *Proc. of the Linux Symp.*, vol. 2, July 2006.

A Fast Low-Light Multi-Image Fusion with Online Image Restoration

Young-Su Moon, Shi-Hwa Lee, Yong-Min Tai, and Junguk Cho
Samsung Advanced Institute of Technology, Samsung Electronics, Korea

***Abstract*—This paper presents a new low-light multi-frame fusion algorithm to get a bright and clear shot even under dark conditions. To this end, using multiple short-exposure images and one proper-exposure blurry image as an input, a new hierarchical block-wise temporal noise filtering is done. Finally, an online image restoration of the denoising result is conducted along with the blurry image input. Test results on real low-light scene show its effectiveness like fast processing speed and satisfactory visual quality.**

I. INTRODUCTION

Digital camera photos taken under a low-light condition reveal significant image artifacts such as motion blur by long-exposure shooting or strong noise corruption by High-ISO setting. Furthermore, as camera sensor's resolution increases, such artifacts are getting worse due to lack of incoming lights on each sensor cell.

To solve it, many research works have been studied. In a new camera imaging system [1], both color and (white+Near) IR images are fused to provide a high sensitivity color image. In [2], two different exposure-time (short-exposure and long-exposure) images are combined to reconstruct an enhanced image via the luminance and chrominance fusion procedures, but it asks the two images to be geometrically and photo-metrically aligned quite well. In the V-BM4D [3], a concept of 4-D nonlocal spatiotemporal transforms is suggested to recover the limits of its previous work, V-BM3D. These works require either a huge amount of computations or some specially designed HW components. To enable its practical in-camera SW processing, it is very important to develop a cost-effective and quality-competitive algorithm.

In this work, firstly, a short-exposure of multiple noisy input shots are fused on a new hierarchical block-wise temporal denoising framework to get a good denoised (dark) image. Then, a proper-exposure of blurry (bright) input shot is referred into online image restoration procedure without any calibration model to transform the denoised (dark) intermediate image into a visually good final shot.

II. PROPOSED LOW-LIGHT MULTI-FRAME FUSION

Our proposed low-light fusion algorithm consists of two processing parts with an input including a set of short-exposure noisy images and one proper-exposure blurry image, as in Fig.1.

A. Multi-Resolution Temporal Image Denoising

Multiple still shots captured with hand-held continuous shooting mode need to be geometrically aligned. To achieve this effectively, global image motion between a reference short-exposure input image and other short-exposure input images is estimated with a fast and effective method using a translation model [4]. For convenience, the first short-exposure input image is selected as the reference. Since actual image motion between the input frames is complicated, subsequent block-based local motion estimation is required.

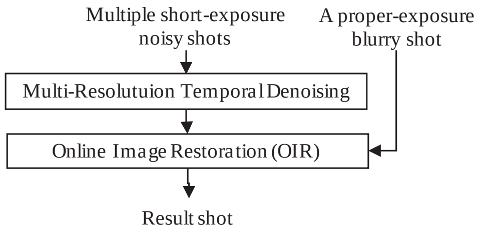
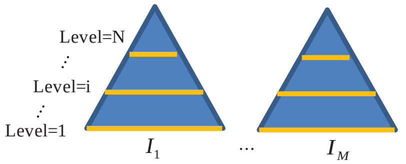

Multiple short-exposure noisy shots
A proper-exposure blurry shot
Multi-Resolutuion Temporal Denoising
Online Image Restoration (OIR)
Result shot

Fig. 1. Block diagram of the proposed low-light fusion algorithm.

More specifically, for every image block in the reference input, matching image blocks over other input images are searched by using a similarity measure like SAD, SSD, or NCC. Of course, its searching can be accelerated by using the global motion info. After determining correspondence of each block on the reference image, block-wise temporal denoising is conducted by simply averaging the set of matching blocks. This is very important, because, as mentioned in [3], a simple temporal denoising method gives better performance than spatial denoising methods with even higher complexity.



```
For level, i = N:1
(1) If top level, {Just use LPF band images}
     else {construct a current level (i)'s LPF band images
      from both upper level (i+1)'s LPF band images &
      current level's HPF band images }
(2)  Estimate/refine global image motions between the
     LPF band images using upper level's global motion
(3)  Find/refine local motion between the LPF band images
     using current level's global motion and upper level's
     local motion
     Do a block-wise temporal filtering for all image
     blocks in a reference image and redistribute the
     filtered block into its corresponding blocks in other
     noisy input images
end
```

Fig. 2. Pseudo-code of the proposed multi-resolution block-wise temporal denoising algorithm.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

To accelerate the above process and simultaneously improve its performance, a new multi-resolution framework based on Laplacian pyramid is combined with the above process, as shown by the pseudo-code in Fig.2. Especially, at every pyramid level from top level (N) to bottom level (1), all the input images must be denoised, and then their results must be reprojected into next fine level to accomplish even better denoising performance. In reference, the number of short-exposure input images (M) influences on its denoising quality and the amount of the data to be processed. Also, the number of pyramid levels (N) leads to the speed-up of the denoising process as well as the improvement of the denoising quality SNR. Of course, the number (N) should be also determined depending on the size of the input image.

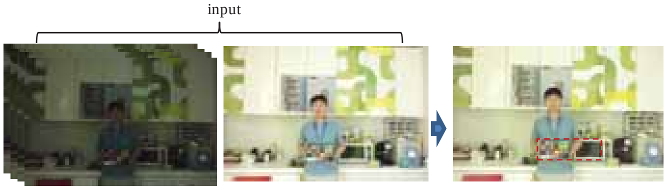
B. Online Image Restoration (OIR)

The denoised (dark) image is transformed by three channels of brightness mapping curves from a well-known histogram matching scheme to get a bright and clear image using the blurry (bright) input image. This histogram matching allows any blurry input image, even significantly shaked in reference to the denoised image, to be used without any image alignment between them. As a result it enables to achieve satisfactory image restoration anytime, anywhere without a complicated offline-calibrated mapping model (showing inherently limited performance).

III. Experiments and Conclusion

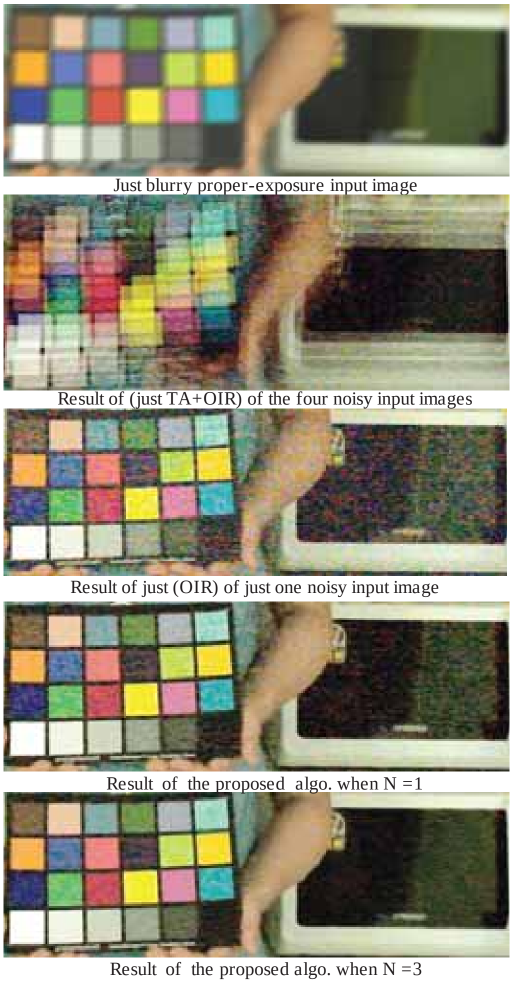
The proposed algorithm was tested on real low-light scenes with image size (5472x3648) by using the matlab tool on PC with Intel Core2 Quad 2.4GHz, 4GB RAM. Fig. 3 shows one sample of the test results, in which the result image looks bright and clear. For detailed analysis, various image cropped regions from intermediate result, final result, and input noisy image are compared in Fig. 3 (b). The last one shows the effectiveness of the proposed one in terms of brightness (/color) enhancement and salient color noise suppression. The proposed one with N=3 takes about 727.0 sec and yields the best denoising performance, whereas the case of N=1 takes about 2928.0 sec and produces worse result. In reference, one matlab program of the method [5] takes about 1550.0 sec to denoise only one of the input images.

We showed that the proposed multi-resolution low-light fusion algorithm could give a visually satisfactory image shot under low light condition and also could be processed more fast compared to any other low-light enhancement fusion methods.



Four short-exposure noisy shots One proper-exposure blurry shot Result of the proposed algorithm

(a)



Just blurry proper-exposure input image

Result of (just TA+OIR) of the four noisy input images

Result of just (OIR) of just one noisy input image

Result of the proposed algo. when N =1

Result of the proposed algo. when N =3

(b)

Fig. 3. Test results. (a) input captured in low light condition, e.g. 20 lux (left: four shots with ISO 6400, exposure time 1/100 sec, right: one blurry shot with ISO 3200, exposure time 1/6 sec), and result of the proposed algorithm, (b) comparison of some intermediate results (**first**: just the blurry input image, **second**: result of both the temporal averaging (**TA**) of four input images without image alignment and its OIR, **third**: result of just OIR of one noisy input image, **fourth** & **last** : result of the proposed algorithm when N=1 and N=3, respectively) over a specific image region by red-dotted rectangle in Fig 3. (a).

References

[1] B.K. Park, S.W. Han, and et al, "Low light imaging system with Expanding spectrum band for digital camera", *IEEE International Conference on Consumer Electronics (ICCE)*, 2012

[2] M. Tico and K. Pulli, "Image enhancement method via blur and noisy image fusion", *IEEE International Conference on Image Processing (ICIP)*, Nov., 2009

[3] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, “Video denoising using separable 4D nonlocal spatiotemporal transforms”, *Proc. SPIE Electronic Imaging*, January 2011

[4] A. Sibiryakov and M. Bober, “Low Complexity Motion Analysis for Mobile Video Devices”, *IEEE Conference on Consumer Electronics*, 2007

[5] Ming Zhang and Bahadir K. Gunturk, “Multiresolution Bilateral Filtering for Image Denoising”, *IEEE Transactions on Image Processing*, December, 2008

Fast Generating Thumbnail in MBAFF Mode of H.264/AVC Intra-Coded

Huy Tran
College of Electronics and Information
Kyung Hee University
Gyeonggi-do, Republic of Korea 446-701
Email: huytn@khu.ac.kr

Wonha Kim
College of Electronics and Information
Kyung Hee University
Gyeonggi-do, Republic of Korea 446-701
Email: wonha@khu.ac.kr

***Abstract*— In this paper, we propose a method for generating thumbnail images from the intra-sliced macroblock-adaptive frame/field coding mode in H.264/AVC bit stream. The pixel value of thumbnail image is combined by two elements: the average value of residual pixels, which is extracted in the transform domain, and the average value of estimated pixels, which is calculated in the pixel domain. Also, the pixels only required for the intra prediction in the pixel domain is reconstructed by using the reduced form of inverse integer discrete cosine transform. The thumbnail images produced by the proposed method are indistinguishable to the ones by the method that decodes the H.264/AVC-I slice bit streams and then scales them down. For most of images, the proposed method also executes at least 2 times faster than the full decoding method at frequently used bandwidths.**

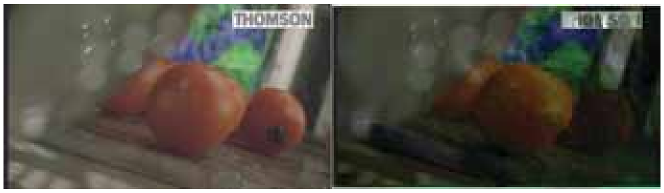
***Index Terms*—Thumbnail Image, H.264/AVC, MBAFF**

I. Introduction

H.264/AVC adopts macroblock-adaptive frame/field coding mode (MBAFF) [1] to encode mixed regions in a picture. Cause, tt is more efficient to compress each field separately the region of motion existences in sequence. Whereas, the frame coding method is suitable for the non-moving regions in a picture.

For producing thumbnail images in H.264/AVC only in frame mode, a representative study was conducted by Kim *et al.* [2]. Kim used a matrix operation for performing the intra prediction in the transform domain.Thus, he adopted lookup tables for compensating the errors, and directly extracted DC values in the transform domain. However, the integer approximation errors cannot be exactly compensated in transform domain and so it is difficult for their method to control the errors, and then induces mismatch errors between the encoder and decoder [3]. As a result, it is urgent need to develop a method for efficient generation of thumbnail images from H.264/AVC compressed video adopting MBAFF mode.

In this work, we propose new algorithms to generate thumbnail images from H.264/AVC intra-coded bit streams in MBAFF mode. It not only radically eliminates the source of the integer approximation errors, but also performs integer operations. The developed method executes at least 2 times faster than the full decoding method at frequently used bandwidths.



(a) (b)

Fig. 1. Thumbnail image of HD (1920 × 1080) sequence. (a) Full decoding method. (b) Kim's method in the transform domain [2].

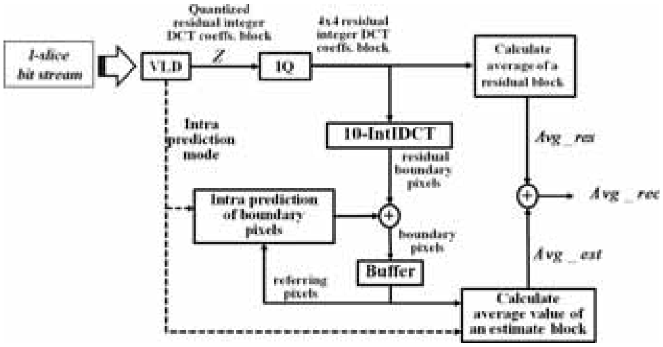

I-slice bit stream
VLD
Quantized residual integer DCT coeffs. block
Z
IQ
4x4 residual integer DCT coeffs. block
Calculate average of a residual block
Intra prediction mode
10-IntIDCT
residual boundary pixels
Avg _res
+
Avg _rec
Intra prediction of boundary pixels
+
boundary pixels
Avg _est
Buffer
referring pixels
Calculate average value of an estimate block

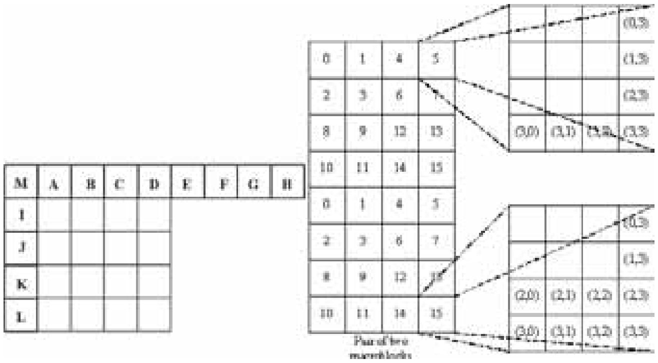
Fig. 2. The scheme generates a thumbnail pixel of a block in frame coding from the H.264/AVC bit stream.

II. Proposed Method

This section describes how to generate thumbnail image from the data bit stream of an I-slice 4×4 block, and output is an average value of that 4×4 block in the reconstructed image Avg_rec. Fig. 1(a) shows the result of extracting the thumbnail image by the full decoding method, that is error-free. While, the quality of thumbnail image in Fig. 1(b) degrades rapidly towards the bottom-right corner image. It causes by the integer approximation error accumulates as the intra prediction proceeds in the HD resolution sequences.

Fig. 2 illustrates the scheme of the proposed method for a block in frame coding. The average of residual blocks Avg_res is calculated by directly extracting the integer DCT coefficients of each block. Simultaneously, the residual integer DCT coefficients are used to reconstruct the essential residual pixels in the step 10-IntIDCT. These values also combine with estimated pixels to reconstruct the boundary pixels, which are

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a) (b)

Fig. 3. Referencing pixels of H.264/AVC intra prediction. (a) The 13 reference pixels for the intra predictions of a 4×4 block. (b) Reconstruction of the required pixels to make the reference for the intra prediction of a MB pairs in MBAAF mode.

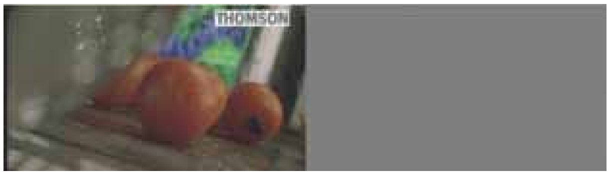
stored in buffer to make the reference for intra prediction of the next block. H.264/AVC standard decoder uses the value pixel in the spatial domain as referencing samples to make prediction of intra coding. Therefore, by adopting multiple domains which exploit both the transform and spatial domain, we can guarantee that the proposed method radically eliminates the source of the integer approximation errors. Because there is no any mismatch errors in the referring pixels comparing to the H.264/AVC standard decoder.

If the MB pairs is frame coding type, the process in figure 2 can be used. Each value Avg_rec is a sum of Avg_res and Avg_est. However, if the pair of two MBs are in field mode, the process in figure 2 has a little modification. Instead of having one output value in step calculation of residual and estimated, now each step has two output values. The step of calculation of residual has two results: Avg_res_A and Avg_res_B which is the residual value from the field of block A and B. As same as for output values of step calculation of average of estimated filed from block A Avg_est_A and block B Avg_est_B, respectively. Based on the intra prediction of H.264/AVC, the average value of estimated block A Avg_est_A and block B Avg_est_B can be calculated from those referring pixels A-M in Fig. 3(a).

Fig. 3(b) shows the position of 32 $4{\times}4$ blocks in a MB pairs, the number on each block is the decoding order of each 4×4 block in the whole MB. If mode of MB is field coding, all $4{\times}4$ blocks need to reconstruct only 7 boundary pixels like the 5^{th} block. When MB is a frame coding, most of the 4×4 block reconstructs only 7 boundary pixels. Except for the 10^{th}, 11^{th}, 14^{th}, 15^{th} of the lower MB must reconstruct 10 pixels. Since the proposed method generates only 7 pixels of a 4×4 block, excepted 4 blocks need 10 pixels, instead of reconstructing the entire 4×4 block, therefore the proposed method can reduce 60% complexity compared to the full decoding method.

III. Experiments

The effect of the proposed method is assessed through the quality of thumbnail image and actual computation time in comparison to the full decoding method. The proposed method was implemented on the H.264/AVC-JM17.2. Resolution of the thumbnail images (480×270) is generated from various HD sequences.



(a) (b)

Fig. 4. Thumbnail images from ‘Rolling Tomatoes’. (a) Thumbnail image generated by the proposed method. (b) Differenence image between propose method and average method.

TABLE I

Comparison of the average execution time for generating a thumbnail image $(msec)$

Sequences	Method	Quantization Parameter (QP)				
		10	16	25	33	45
Table Setting	Full decoding	485	352	208	152	122
	Proposed	411	262	99	47	24
Playing Cards	Full decoding	456	327	201	153	123
	Proposed	384	235	98	50	24
Rolling Tomatoes	Full decoding	361	252	149	128	116
	Proposed	288	157	45	27	18

For an objective performance comparison, we use Structural SIMilarity index (SSIM) which models error as perceived by a human observer [4]. SSIM score between results of the proposed method and the full decoding method is 0.99, which indicates that the result images are indistinguishable. Subjective quality test is also demonstrated in Fig. 4.

Table I compares the processing times for generating a thumbnail image. At the QP values of around 25 that produces frequently used bandwidth for HD size, the proposed method executes at least 2 times faster than the full decoding method.

IV. Conclusion

We presented a novel method for generating thumbnail images from the MBAFF mode of H.264/AVC coded bit streams. The proposed method radically eliminates the source of the integer approximation errors. Also, the experimental results show that our proposal demonstrates excellent performance for generating the thumbnail images.

Acknowledgment

This work was supported by grant No. 2010-0012331 from National Research Foundation of Korea.

References

[1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. on CSVT*, vol. 13, no. 7, pp. 560 –576, july 2003.

[2] E.-S. Kim, T.-W. Um, and S.-J. Oh, “A fast thumbnail extraction method in h.264/avc video streams,” *IEEE Trans. on CE*, vol. 55, no. 3, pp. 1424 –1430, august 2009.

[3] P.-H. Wu, C. Chen, and H. Chen, “Rounding mismatch between spatial-domain and transform-domain video codecs,” *IEEE Trans. on CSVT*, vol. 16, no. 10, pp. 1286 –1293, oct. 2006.

[4] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600 –612, april 2004.

Constrained Two-bit Transform for Low Complexity Motion Estimation

Changryoul Choi and Jechang Jeong, *Member, IEEE*

***Abstract*—In transforming the original image frames into two-bit representations, the typical two-bit transform (2BT) needs to calculate the variances of the local blocks. This calculation of variances inevitably involves multiplication operations and renders the computational complexity of typical 2BT somewhat high. In this paper, we propose a constrained two-bit transform (C2BT) for low complexity motion estimation (ME). By exploiting the advantages of the typical constrained one-bit transform (C1BT) and the typical 2BT, the proposed algorithm significantly reduces the computational complexity of transformation of image frames into two-bit representations. Experimental results show that the proposed algorithm enhances the ME accuracy by 0.35dB and 0.28dB compared with the 2BT-based ME and the C1BT-based ME, respectively.**

I. Introduction

In video compression, motion estimation (ME) and motion compensation (MC) is the key technique to reduce the inherent temporal redundancy between the neighboring frames [1]. The most popular technique for ME is the block matching algorithm (BMA), which is adopted in many video coding standards [2] due to its simplicity and effectiveness. In BMA, video frames are partitioned into small rectangular blocks and a motion vector for that rectangular block is determined by finding the closest rectangular block according to a certain matching criterion such as the sum of absolute differences (SAD) or the sum of squared differences (SSD). The full search BMA (FSBMA) can give optimal estimation of motion in terms of minimal matching error by checking all the candidates within the search range, but due to the prohibitively huge computational complexity, it is impractical for the real-time video applications. Therefore, many techniques have been proposed to reduce the high computational complexity of the FSBMA. Among those techniques, ME algorithms that use different matching criteria instead of the typical SAD or SSD are proposed to make the faster computation of the matching criteria using bit-wise operations [3]-[11]. Fast computation of the matching criterion and reduced memory bandwidth in the interim of ME process are two benefits which are expected from the use of bit-wise operations. These techniques include one-bit transform (1BT), multiplication-free 1BT, constrained 1BT (C1BT), two-bit transform (2BT), weighted 2BT, truncated gray-coded bit plane matching (TGCBPM), weightless TGCBPM, etc.

In 1BT-based ME, video frames are transformed into one-bit representations by comparing the original pixel value against a filtered output [3]. Each frame is filtered with the 17×17 kernel to create a one-bit frame. After this transform, the bit-wise matching criterion which is called the number of non-matching points of 1BT ($NNMP_{1BT}$), is given by

$$NNMP_{1BT}(m,n) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\{B^{t}(i,j) \oplus B^{t-1}(i+m, j+n)\} \qquad (1)$$

where $B^{t}(i, j)$ and $B^{t-1}(i, j)$ are the one-bit representations of the current and the reference frames, respectively and $\oplus$ denotes the Boolean XOR operation.

A 2BT-based ME was proposed to increase the ME accuracy of the 1BT-based ME algorithms [5]. The values of local mean μ, variance σ^2, and the approximate standard deviation σ_a are used to transform frames into two-bit representations. To exploit the full advantage of the increased bit-depth from 1-bit to 2-bit, variations of the $NNMP_{2BT}$, Extended $NNMP_{2BT}$ ($ENNMP_{2BT}$) and weighted $NNMP_{2BT}$s ($WNNMP_{2BT}$) were proposed in [7].

A constraint mask bit-plane was introduced to improve the performance of 1BT, which is called the C1BT in [6]. Although C1BT-based ME uses two bit-planes in matching criterion similar to 2BT, it is very simple to create another bit-plane in C1BT. In general, C1BT-based ME provides slightly better ME performance compared to the 2BT based ME.

TABLE I
Number of Operations for 1BT, 2BT and C1BT (per pixel)

Operations	1BT	2BT	C1BT
Addition	25	2.8125	16
multiplication	1	1.0625	-
Subtraction	-	0.03125	1
Shift	-	-	1
Comparison	-	3	3

Table I shows the number of operations for transformation using 1BT, 2BT and C1BT [6]. From the table, we can see that the computational complexity of transforming frames into two-bit representations is relatively high because it needs multiplication operations in calculating variances. However, except multiplication operations, the total operations needed for 2BT is very low compared with other transforms. Therefore, if the multiplication operations in 2BT can be replaced with other operations, a low complexity transform can be attained. Note that for the C1BT, the computational complexity for attaining the first bit-plane is relatively high, but for attaining the second bit-plane (which is so called "constraint mask") is simple. To exploit the advantages of the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

C1BT-based ME and the 2BT-based ME, we propose a C2BT as follows:

$$
\begin{aligned}
&\mu = E[I_{tw}] \\
&C_1(i,j) = \begin{cases} 1, & I(i,j) \geq \mu \\ 0, & otherwise \end{cases} \qquad (2) \\
&C_2(i,j) = \begin{cases} 1, & if\ |I(i,j) - \mu| \geq D \\ 0, & otherwise \end{cases}
\end{aligned}
$$

where I_{tw} are the pixel values in the local threshold window around the transforming block, $I(i, j)$ are the pixel values of the transforming block and $C_1(i, j)$ and $C_2(i, j)$ are the two-bit representations of C2BT. Unlike the 2BT, we set the local threshold window as 32×32 to avoid multiplication operations. Then the number of operations (per pixel) for the proposed algorithm can be summarized as Table II.

TABLE II
NUMBER OF OPERATIONS FOR THE PROPOSED ALGORITHM (PER PIXEL)

Addition	multiplication	Subtraction	Shift	Comparison
1.1133	-	1.0642	0.015625	3

Note that when calculating the mean around the transforming block, we adopted the method in successive elimination algorithm [12] to reduce total operations. Compared with other transforms, the proposed algorithm is multiplication-free and the total computational complexity is very low.

For the corresponding matching criterion for C2BT, we adopted the methods in [7][10]. However, there are two problems regarding the matching criterion in [10]. The first one is that it needs multiplication operation in matching, which should be avoided in low complexity ME. The second one is that the matching criterion is not symmetric. To remedy these problems, we propose a matching criterion for C2BT as follows:

$$
\begin{aligned}
&NNMP_{C2BT,1}(m,n) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\{C_1^t(i,j) \oplus C_1^{t-1}(i+m, j+n)\} \\
&NNMP_{C2BT,2}(m,n) = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\{C_2^t(i,j) \oplus C_2^{t-1}(i+m, j+n)\} \\
&NNMP_{C2BT,3}(m,n) \\
&= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}[C_2^t(i,j) \bullet \{C_1^t(i,j) \oplus C_1^{t-1}(i+m, j+n)\}] \qquad (3) \\
&NNMP_{C2BT,4}(m,n) \\
&= \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}[C_2^{t-1}(i+m, j+n) \bullet \{C_1^t(i,j) \oplus C_1^{t-1}(i+m, j+n)\}] \\
&NNMP_{C2BT}(m,n) \\
&= \sum_{i=1}^{2} NNMP_{C2BT,i}(m,n) + 2 \times \sum_{i=3}^{4} NNMP_{C2BT,i}(m,n)
\end{aligned}
$$

where ⊕ and • denote the Boolean XOR and AND operation, respectively.

Table III shows the average PSNR comparison results when the motion block size is 16×16 and the search range is ±16. In this case, the threshold for C1BT was set 10 and that of the C2BT was set 30. From the table, we can see that the proposed algorithm outperforms other algorithms.

TABLE III
AVERAGE PSNR RESULTS OF ALGORITHMS
WHEN THE MOTION BLOCK SIZE IS 16×16 (SEARCH RANGE =±16)

Sequences	2BT [5]	Weighted 2BT [7]	C1BT [6]	Weighted C1BT [11]	Full 2BT [10]	SAD	Proposed
stefan	25.26	25.35	25.23	25.42	25.52	25.75	25.53
football	23.05	23.38	23.03	23.36	23.59	24.00	23.67
akiyo	42.39	42.42	42.54	42.66	42.48	42.84	42.49
mobile	23.57	23.72	23.64	23.77	23.77	23.92	23.77
tempete	27.26	27.40	27.28	27.38	27.43	27.70	27.5
table	27.86	28.19	28.07	28.27	28.36	28.87	28.47
flower	25.83	25.89	25.78	25.88	25.92	26.03	25.94
children	28.32	28.72	28.48	28.72	28.90	29.24	28.92
average	**27.94**	**28.13**	**28.01**	**28.18**	**28.25**	**28.54**	**28.29**

REFERENCE

[1] Z. He, C. Tsui, K. Chan, and M. Liou, "Low-power VLSI design for motion estimation using adaptive pixel truncation," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 10, no. 5, pp. 669-678, Aug. 2000.

[2] Advanced Video Coding for Generic Audiovisual Services, *ITU-T Recommendation H.264*, May 2005.

[3] B. Natarajan, V. Bhaskaran, and K. Konstantinides, "Low-Complexity Block-based Motion Estimation via One-Bit Transforms," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 7, no. 5, pp. 702-706, Aug. 1997.

[4] Sarp Erturk, "Multiplication-free one-bit transform for low-complexity block-based motion estimation," *IEEE Signal Processing Letters*, vol. 14, no. 2, 109-112, Feb. 2007.

[5] A. Erturk and S. Erturk, "Two-bit transform for binary block motion estimation," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 15, no. 7, pp. 938-946, July 2005.

[6] O. Urhan and S. Erturk, "Constrained one-bit transform for low complexity block motion estimation," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 17, no. 4, pp. 478-482, Apr. 2007.

[7] Changryoul Choi and Jechang Jeong, "Enhanced two-bit transform based motion estimation via extension of matching criterion," *IEEE Trans. Consumer Electron.,* vol. 56, no. 3, pp. 1883-1889, Aug. 2010.

[8] Anil Celebi, O. Akbulut, O. Urhan, and S. Erturk, "Truncated Gray-coded bit-plane matching based motion estimation method and its hardware architecture," *IEEE Trans. Consumer Electron.*, vol. 55, no. 3, pp. 1530-1536, Aug. 2009.

[9] Anil Celebi, Hyuk-Jae Lee, and Sarp Erturk, "Bit plane matching based variable block size motion estimation method and its hardware architecture," *IEEE Trans. Consumer Electron.*, vol. 56, no. 3, pp. 1625-1633, Aug. 2010.

[10] S. Erturk, "Exploiting the full potential of two-bit transform based motion estimation," *International Conference on Consumer Electronics, ICCE 2011*, Las Vegas, USA, Jan. 2011.

[11] M. K. Gullu, "Weighted constrained one-bit transform based fast block motion estimation," *IEEE Trans. Consumer Electron.*, vol. 57, no. 2, pp. 751-755, May 2011.

[12] W. Li and E. Salari, "Successive elimination algorithm for motion estimation," *IEEE Trans. Image Processing*, vol. 8, pp. 105-107, Jan. 1995.

A High Performance Hearing Aid System with Fully Programmable Ultra Low Power DSP

Yunseo Ku[(1)], Junil Sohn[(1)], Jonghee Han[(1)], Yonghyun Baek[(2)] and Dongwook Kim[(1)]
(1) Samsung Advanced Institute of Technology, Samsung Electronics, Yongin, Korea
(2) Division of Computer & Telecomm. Eng, Yonsei University, Wonju, Korea

Abstract-- **This paper presents the design of fully programmable digital signal processor for the hearing aid system and optimized implementation of digital hearing aid algorithms. The average performance of the developed processor is measured by performing Fast Fourier Transform (FFT) algorithm. The proposed applications such as variable multi-band loudness compensation algorithm with wide dynamic range compression and sub-band based adaptive feedback cancellation algorithm are optimized for real-time processing by exploiting the data and instruction parallelism of the developed processor. The results show that the computational ability and power consumption of the developed processor are suitable to use for the hearing aid system. The optimization achieves that the computational time to perform the full applications on the processor takes only under 40% of the constraint for real-time processing.**

I. INTRODUCTION

Digital hearing aids require low power consumption and high computation performance at the same time due to limited battery capacity and signal processing complexity. As modern hearing aids are getting smaller for cosmetic matter, it becomes hard to use the bigger sized battery which has enough capacity to process complex hearing aid algorithms. [1] Therefore, the performance of digital signal processor (DSP) in the hearing aids and the algorithm optimization play a major role to fully exploit the benefits from sophisticated hearing aid algorithms. Flexibility and programmability of the processor are also significant because the applied algorithms in the hearing aids vary based on different types of the hearing impaired and their ambient environment. [2] In order to achieve the low power consumption and small size of chip area, the fixed-point DSP is generally used in digital hearing aids since the fixed-point hardware is much simpler than floating-point hardware. [3] When implementing signal processing algorithms on the fixed-point DSP, however, the careful optimization process is needed to minimize quantization error and to avoid overflow/underflow. [4] This paper concerns not only the development of low power fixed-point DSP for the hearing aids but also the hearing aid algorithm optimization.

II. DIGITAL SIGNAL PROCESSOR FOR HEARING AID

A 16/32-bit customized low power programmable DSP is designed to perform the hearing aid algorithms and to meet low power consumption requirement. Memory data path size and instruction-set are customized by the simulation when performing signal processing algorithms such as Fast Fourier Transform (FFT), which is mostly applied to hearing aid algorithms. The simulation result shows that maximum 128-bit data vector processing with twelve 16-bit multipliers is most suitable for the hearing aid system in the view of the computational ability and the power consumption. Fig.1 shows the overall architecture of developed processor which includes the low power SRAM and the peripherals such as SPI and I2S.

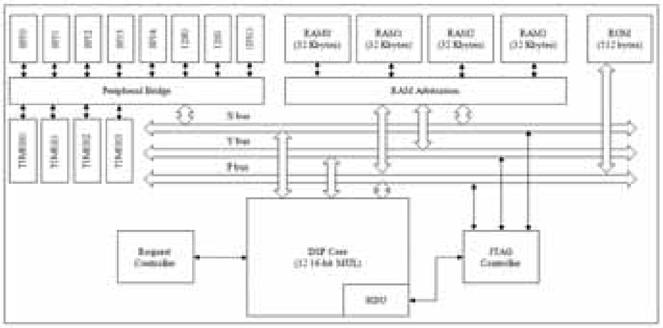


Fig.1. Proposed Digital Signal Processor for Hearing Aids

In order to evaluate the performance of developed DSP in the view of the computational ability and the power consumption, the required cycles for 256-point radix-4 FFT is profiled by the simulator and the power consumption for DSP core when computing FFT infinitely is measured on the test board which has the developed DSP chips on it. The average computational ability of the DSP core can be measured by calculating total operations to perform FFT and the number of required cycles to execute FFT. The number of required cycles of proposed processor is 900 while other low power processors need more than 3K cycles as shown in Table I. [5]

TABLE I
COMPARISON OF NUMBER OF CLOCK CYCLES FOR 256 FFT

Processor	Required Cycles
Blackfin BF531 (Analog Devices)	3.2K
CoolFlux DSP (NXP)	5.5K
LSI403LP (LSI Logic)	5K
Proposed DSP	**0.9K**

The number of operations of proposed processor is 15,000, which leads that the average number of operations is 16.7 per cycle. The measured power consumption for DSP core is 935uA at 0.9V supply voltage and 8.25MHz operating clock frequency. These lead that the number of million operations

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

per one second (MOPS) per one milliwatt (mW) is 163, which can be considered as the average performance of developed DSP in this work.

III. HEARING AID ALGORITHM OPTIMIZATION

Due to smaller hearing dynamic range of the hearing impaired, the proper gains should be applied to the input signal according to input signal intensity and hearing loss level of each frequency band. The FFT filter bank is used to split frequency bands of the input signal according to the critical band frequency scale of the human auditory system. Thanks to the programmability of the developed DSP, variable band number processing (up to 64) can be performed according to the different hearing characteristics. Fig.2 shows that the variable band architecture compensates the hearing loss more accurately than the fixed band architecture.

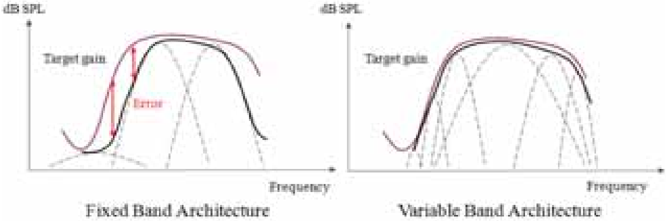


Fig.2. Compensation Error Difference between the Fixed & Variable Band Architecture

In each band, the band energy of input signal is calculated and then the proper gain is determined based on the loudness scaling function and the calculated current band energy. The determined gains are applied the analysis output and the compensated output are synthesized through inverse FFT filter bank. Due to very close distance between the microphone and the loudspeaker in the hearing aids, the acoustic feedback is prone to occur when the sound from the loudspeaker loops back through the microphone. The acoustic feedback limits the maximum gain and degrades the sound quality of the hearing aids. In this work, the sub-band based adaptive feedback cancellation algorithm is adopted in order to integrate with the multi-band loudness compensation algorithm in the frequency domain. Fig.3 shows the block diagram of multi-band loudness compensation algorithm with sub-band based adaptive feedback cancellation algorithm.

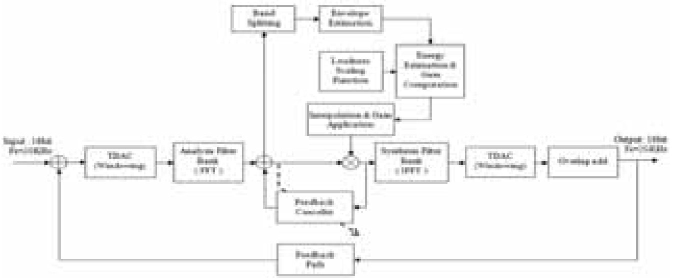


Fig.3. Multi-Band Wide Dynamic Range Compression & Sub-Band based Adaptive Feedback Cancellation Algorithm

In order to optimize critical functions down to the assembly language level, the candidate functions are identified to perform the parallelism in the view of data level and instruction level. To reduce the number of loop iterations, single instruction-multi data instructions are applied to perform 64-bit to 128-bit vector data processing. Instruction level parallelism for reducing the code lines is also performed by avoiding or tolerating the instruction latency. Moreover, reconfigurable addressing instructions are used for specialized addressing such as the bit-reversed addressing for FFT processing. For the real time implementation, maximum 51,200 cycles are allowed for the algorithm computation under the condition of 8MHz operating clock frequency, 256-point FFT, half overlapping and 20KHz sampling frequency. After the optimization procedure, the total computational load of both algorithms is 37.5% as shown in Table II. Even for the maximum band number case (64 bands), the computational load is only 46.1%.

TABLE II
THE NUMBER OF CYCLES FOR HEARING AID ALGORITHMS

Algorithm Type	Required Cycles	Computational Load (%)
Multi-Band Wide Dynamic Range Compression (8 bands) & Sub-Band based Adaptive Feedback Cancellation	19,199	37.5
Multi-Band Wide Dynamic Range Compression (64 bands)	23,621	46.1

IV. CONCLUSION

Fully programmable ultra low power DSP for the hearing aid system is developed and two main hearing aid algorithms are optimized for real-time processing. The results show that the computational efficiency of the developed processor is suitable to use for digital hearing aids. The power consumption may be further decreased by modifying the operating clock frequency. Moreover, further software optimizing also would allow further decreasing the power consumption.

ACKNOWLEDGMENT

This work was supported by grant No.10031741 from the Strategic Technology Development Program of the Ministry of Knowledge Economy.

REFERENCES

[1] Henning Puder, “Hearing Aids: An Overview of the State-of-the-Art, Challenges, and Future Trends of an Interesting Audio Signal Processing Application,” Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis, 2009.

[2] Peng Qiao, “A 0.964mW digital hearing aid system,” Design, Automation & Test in Europe Conference & Exhibition (DATE), 2011.

[3] Dake Liu, “Embedded DSP Processor Design, Application Specific Instruction Set Processor,” Morgan Kaufmann Publishers, 2008.

[4] Brian Delaney, “A Low-power, fixed-point, front-end feature extraction for a distributed speech recognition system,” IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2002.

[5] Vojin G. Oklobdzija, “High-Performance Energy Efficient Micro processor Design,” Springer, 2006.

Implementation and Verification of a Platform for Bluetooth Linked Hearing Aids System with Smart Phone and Multimedia Devices

Dong-Wook Kim[1], Eui-Sung Jung[1], Ki-Woong Seong[2], Jyung-Hyun Lee[2] and Jin-Ho Cho[1], *Member, IEEE*
[1]Kyungpook National University, Daegu, Republic of Korea
[2]Kyungpook National University Hospital, Daegu, Republic of Korea

***Abstract*--Binaural hearing aids manufacturers have been developed control and communication technology by utilizing the smart-phone and multimedia devices. However, those systems need to use a wireless repeater system or remote control. In this paper, we've designed and implemented a hearing aids open platform system that can combine smart phone and multimedia devices through a small sized bluetooth adaptor for the development of binaural hearing aids. And also, we implemented android based control GUI (graphical user interface) for hearing aids volume and fitting parameters control. From the experiment results, we verified open platform board and bluetooth adaptor that was successfully controlled the hearing aids, thus, it doesn't need to a wireless repeater system and remote control.**

I. INTRODUCTION

Recently, the major manufacturers as Starkey and Siemens are commercializing wireless control product using the multimedia unit and smart phone for its hearing aids control [1]-[5]. However, these products are necessary wireless repeater system or remote control for wireless communication between the hearing aids and the multimedia unit. Accordingly, hearing aid user should have a wireless system for communication with other multimedia devices and that reason can cause the inconvenience to use. Therefore, more study is required about mini-size wireless repeater system that combines with hearing aids .

In this paper, we designed and implemented a hearing aid system that is open platform board and bluetooth adaptor for wireless controls of volume and fitting parameter. And also, we implemented a graphical user interface (GUI) based on Android system for wireless controlling the hearing aids system. From the experimental results, we confirmed performance of open platform board, bluetooth adaptor and GUI that successfully controlled the hearing aids system.

II. IMPLEMENTED OPEN PLATFORM BOARD

In this paper, we implemented open platform board which is appropriate for study of wireless control hearing aids system. The implemented open platform board shows fig. 1 and it composed two parts that are signal processing and wireless communication control [6]. For hearing aids algorithm, the signal processing part used general-purpose DSP processor (TMS320C6713, Texas Instruments Inc.) and it can be programmed by general programming language such as ANSI C, MATLAB and Simulink using Code Composer Studio. The DSP processor performs 32-bit arithmetic and logical operations at 200 MHz clock frequency and it can be interface to analog audio signals through an AIC23 code in serial port interface format.

The wireless communication control part used the FPGA (XC3S400, Xilinx Inc.). The XC3S400 offers 400 million system gates and operating voltage is 3.3 ~ 1.2 V. Also, designed FPGA can be received variable data for volume control and fitting parameters using the external input port. The data communication between DSP processor and FPGA designed using the general purpose input output port. Fig. 2 shows the structure of open platform board.

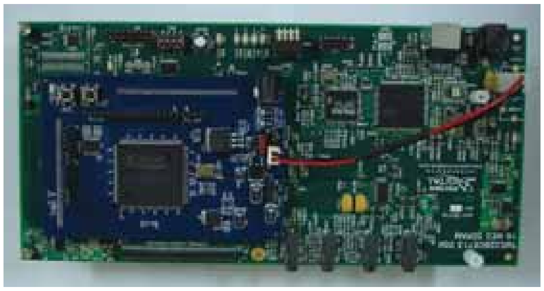


Fig. 1. The implemented open platform board.

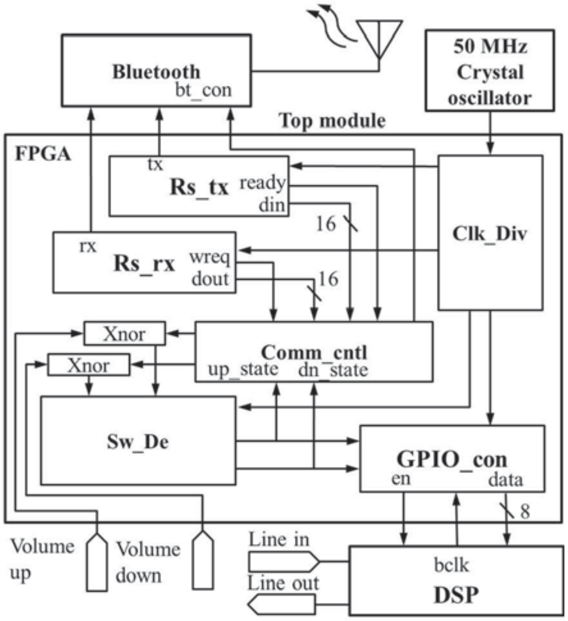


Fig. 2. Block diagram of implemented open platform board.

Acknowledgment: This study was supported by a grant of the Korea Healthcare Technology R&D Project, Ministry of Health & Welfare (A092106). And this work was supported by grant No. 10031779 from the Strategic Technology Development Program of Ministry of Knowledge Economy.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

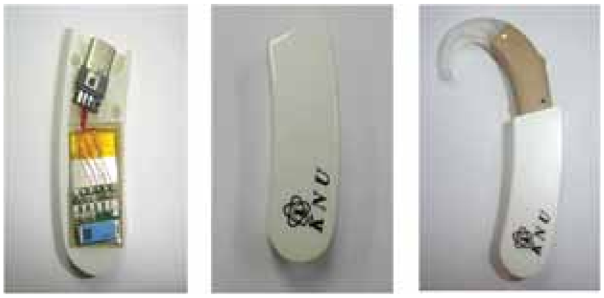


Fig. 3. Implemented Bluetooth adaptor for wireless communication.

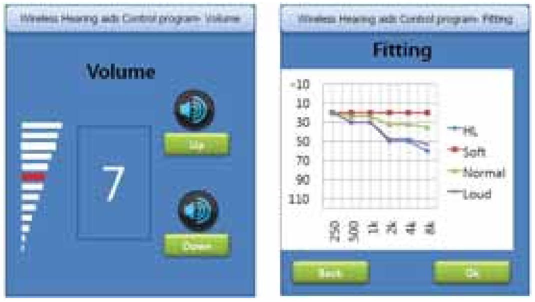

Wireless Hearing aids Control program- Volume
Volume
7
Up
Down
Wireless Hearing aids Control program- Fitting
Fitting
-10
10
30
50
70
90
110
250
500
1k
2k
4k
8k
HL
Soft
Normal
Loud
Back
Ok

Fig. 4. Smart-phone GUI for wireless volume control.

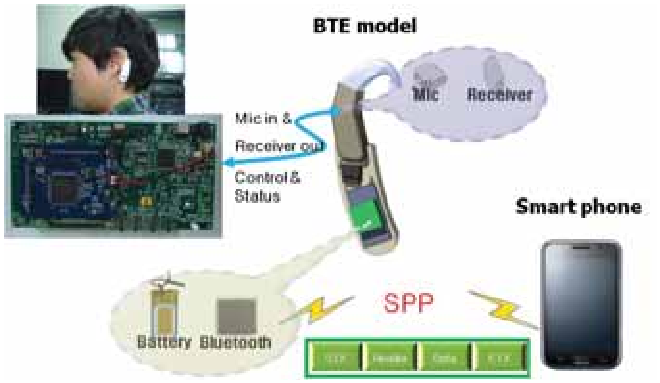


Fig. 5. Experiment set up for wireless volume control using hearing aids system and smart-phone.

III. BLUETOOTH ADAPTOR

The bluetooth adaptor provides wireless communication with smart-phone based on Android and that adaptor was implemented using the BlueCore3 (BC352239A, CSR), lithium polymer battery (3.7 V, 85 mAh) and micro-USB.

The BlueCore3 supports serial port profile (SPP), advanced audio distribution profile (A2DP) and hands free profile (HFP). In experiment of this paper, we used SPP for reception of volume control signal and commonly used baud rates to 460800 bps. Also, the bluetooth adaptor was conveniently designed to use that is the width of 15 mm, the length of 46 mm and the height is 12mm. The implemented Bluetooth adaptor is smaller than wireless repeater system and remote control.

IV. EXPERIMENT METHOD AND RESULT

To verify the open platform board and bluetooth adaptor, we fabricated the hearing aids mock-up. The mock-up has microphone, receiver and micro-USB connector and it is combined bluetooth adaptor. Also, we designed a graphical user interface (GUI) based on Android for wireless controlling the hearing aids system and fig. 4 shows it. The GUI was designed using the eclipse and android-sdk revision 11 based on windows and GUI was installed in smart-phone.

Fig. 5 shows that the experiment set-up for the wireless control using hearing aids system and smart-phone. The input-output ports of microphone, receiver and bluetooth adaptor are connected open platform board. The microphone signal is inputted to the open platform board and then signal is processed currently volume level. The volume stage of hearing aid system is divided from 0 to 12 in 13 steps. When the volume is changed, the sine-tone that is equal to currently volume level is generated.

For the verification of wireless communication and volume control, while controlling the GUI, we measured the communication data using the oscilloscope and confirmed the volume change. Also, we identified that volume was not changed when volume stage is reached to maximum and minimum stage. From the experimental results, we confirmed performance of open platform board, bluetooth adaptor and GUI that successfully controlled the hearing aids system volume.

V. CONCLUSION

In this paper, we designed and implemented an open platform board and bluetooth adaptor for wireless control for binaural hearing aids that doesn't need to a wireless repeater system and remote control. To verify the open platform board and bluetooth adaptor, we fabricated the hearing aids mock-up and smart-phone GUI based on android system. Through the experimental results, we confirmed performance of open platform board, bluetooth adaptor and GUI that successfully controlled the hearing aids system.

Therefore, it is expected that the implemented open platform board and bluetooth adaptor for the wireless control of hearing aids can be applied to developing the hearing aids system and to improve the performance of hearing aids.

REFERENCES

[1] Starkey Inc., U. S. A., http://www.starkey.com.

[2] Simens Inc., Germany, http://www.hearing.siemens.com.

[3] H. Stephen Berger, "Hearing Aid Compatibility with Wireless Communications Devides," Siemens Business Communications Systems, pp. 123-128.

[4] George S. A. Shaker, Mohammad-Reza Nezhad-Ahmadi, S. Safavi-Naeini, Gareth Weale, "On Design of a Low Power Wireless Hearing Aid Communication System," IEEE, pp. 903-906, 2008.

[5] Marlene Skopec, "Hearing Aid Electromagnetic Interference from Digital Wireless Telephones," IEEE Transaction on Rehabilitation Engineering, vol. 6, No. 2, pp.235-239, 1998.

[6] D. W. Kim, J. M. Park, Q. Wei, H. G. Lim, H. J. Park, K. W. Seong, J. H. Lee, M. N. Kim and J. H. Cho, "Implementation of Binaural Communication Open Platfrom for Binaural Hearing Aids Developing," Journal of Sensor Science and Technology, vol. 20, No. 4, pp. 272-278, Republic of Korea, 2011.

Smartphone-based Self Hearing Assessment Using Phonemes

Jong Min Choi, Junil Sohn, Yunseo Ku, and Dongwook Kim
Samsung Advanced Institute of Technology, Republic of Korea

***Abstract*--Phonemes provide an interesting alternative to pure tones in hearing tests. We propose a new smartphone-based method for self hearing assessment using the four Korean phonemes which are similar to the English phonemes /a/, /i/, /sh/, and /s/, respectively. We conducted tests on 15 subjects diagnosed with mild to severe hearing loss and estimated their conventional pure tone hearing thresholds from their phoneme hearing thresholds using regression analysis. The phoneme-based self hearing assessment (PhoSHA) was found to be sufficiently reliable in estimating the hearing thresholds of hearing-impaired subjects. The difference between the hearing thresholds obtained through conventional pure tone audiometry and those obtained using our method was 5.6 dB HL on average. The proposed hearing assessment was able to significantly reduce the mean test time compared to conventional pure tone audiometry.**

I. INTRODUCTION

Smartphones are becoming increasingly powerful, with modern audio codec chips providing smartphone audio of high quality. This makes it possible to conduct hearing tests like pure tone audiometry (PTA) on a smartphone. Although commercial audiometers provide a dynamic range of more than 100 dB in air-conduction audiometry, which cannot be matched by current audio codec chips, the possibility of implementing hearing tests on a smartphone is already widely accepted. Existing literature proposes PTA on personal computers [1, 2] and mobile phones [3].

We proposed a phoneme-based self hearing assessment (PhoSHA) in our previous paper [4] and now implement it on a smartphone running the Android operating system. The PhoSHA method provides an alternative way to test hearing thresholds with sufficient accuracy and reduced test time. Like the results of conventional audiograms, the results of this method can be used to fit hearing aids.

II. METHODS AND RESULTS

Pure tones were replaced by four Korean phonemes in our proposed self hearing assessment. We feel that it is reasonable to use phonemes because they are more representative of human speech. Phonemes have multiple formants and, compared to tones, a wider spectral energy distribution and much more information content. We chose the four Korean phonemes which are similar to the English phonemes /a/, /i/, /sh/ and /s/, respectively. The phonemes were recorded by a female announcer. The vowel /a/ has three formants at 710 Hz, 1100 Hz, and 2640 Hz, /i/ has three formants at 400 Hz, 1900 Hz, and 2550 Hz, and the consonants of /s/ and /sh/ have distinguishable spectral characteristics from 3000 Hz to 4000 Hz and from 2500 Hz to 4000 Hz, respectively.

PhoSHA was implemented on a Samsung Galaxy S II smartphone (Samsung Electronics, Republic of Korea) running the Android operating system (Google Inc, USA); it can also be calibrated for other Android smartphones. The earphones bundled with the S II were used for our experiments. Calibration was done for both left and right earpieces. Figure 1 shows the PhoSHA application on the Galaxy S II smartphone.

The application plays the phonemes within a 30 dB SPL to 85 dB SPL range. The level is automatically increased or decreased by 5 dB SPL according to the response of the subject. To begin with, /a/ is played at 50 dB SPL. If the subject responds to this sound, it is played at a lower level. If the subject does not respond, it is played at a higher level. After determining the threshold for /a/, the application moves on to /i/, /sh/, and /s/. We recorded the time taken to perform the test for each subject in addition to their hearing threshold. To reduce the test time, we used the threshold for /a/ (previous phoneme) as an initial sound level of /i/ (current phoneme) and the same procedures were repeated for the next phonemes. Practically, the initial sound level for /a/ was 35 dB SPL and the initial sound level for the next phoneme is set to 5dB SPL lower than the threshold level of the previous phoneme. For example, if a subject could hear /i/ at 50dB SPL, the initial sound level of /sh/ is set to 45 dB SPL to reduce test time.

We recruited 15 subjects diagnosed with mild to severe hearing loss and tested both ears of the subject. Three of these ears were normal hearing; 27 ears were tested. Two of these ears were excluded because of experimental problems; we finally tested 25. The number of left ears was 13 and the number of right ears was 12. The age range was 20 to 78 years, with a mean age of 69.9 years and a standard deviation of 12.2 years. We also recruited 17 normal subjects and analyzed 31 ears in this group. The age range of these subjects was 24 to 41 years, with a mean age of 28.9 years and a standard deviation of 5.0 years.

The phoneme-based hearing thresholds were converted into audiogram hearing thresholds using the regression formulae in Table 1. Tha, Thi, Thsh, and Ths denote the measured hearing threshold levels (in dB SPL) of the test phonemes /a/, /i/, /sh/, and /s/, respectively. eTh250 to eTh8000 denote the estimated hearing thresholds at various frequencies. The regression coefficients in the formulae shown in Table 1 were derived using the measurements for all 25 ears and were calculated via least-squares. In accordance with convention, we rounded the values to the nearest 5 dB.

A. Comparison between PTA and PhoSHA

Figure 1 shows the mean differences between the hearing thresholds estimated by PTA and those calculated using our method and the formulae in Table 1. The mean difference is higher than 5 dB HL at low frequencies (250 Hz and 500 Hz), but close to 5 dB HL at the other frequencies. The overall mean value across the six frequencies is 5.6 dB HL. The standard deviations for the six tested frequencies were 5.4 dB

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

HL, 6.7 dB HL, 5.3 dB HL, 4.1 dB HL, 5.8 dB HL, and 4.1 dB HL, respectively.

TABLE I
REGRESSION FORMULAE AND STATISTICAL PARAMETERS AT SIX FREQUENCIES

Freq.	Regression Formula	p-value
250	$eTh_{250} = 0.6Th_a+0.14Th_i-0.38Th_{sh}+0.39Th_s-4.81$	$p<0.01$
500	$eTh_{500} = 0.89Th_a-0.11Th_i-0.59Th_{sh}+0.66Th_s-4.33$	$p<0.01$
1k	$eTh_{1k} = 1.38Th_a-0.27Th_i-0.03Th_{sh}+0.08Th_s-18.9$	$p<0.01$
2k	$eTh_{2k} = 0.45Th_a-0.01Th_i+0.84Th_{sh}-0.15Th_s-17.26$	$p<0.01$
4k	$eTh_{4k} = 0.05Th_a-0.26Th_i+0.83Th_{sh}+0.25Th_s-0.8$	$p<0.01$
8k	$eTh_{8k} = -0.22Th_a-0.06Th_i+0.27Th_{sh}+0.92Th_s+4.19$	$p<0.01$

eThx : estimated hearing threshold for frequency x Hz

Of the 150 measurements, 29 measurements differed by 10 dB HL, 14 by 15 dB HL, and 4 by 20 dB HL. The maximum difference was 20 dB HL. The remaining 103 measurements differed by 5 dB HL or less (53 differed by 5 dB HL, and 50 were exactly correct).

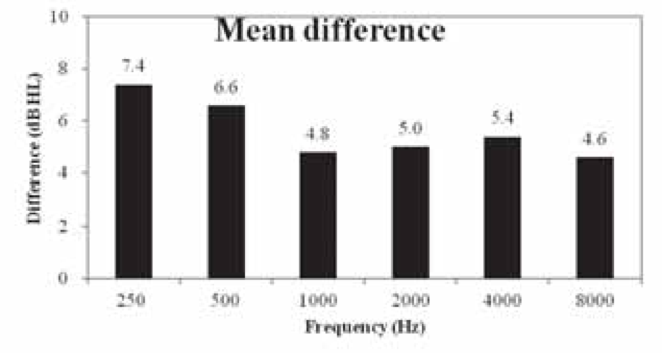


Fig. 1. Mean difference between hearing thresholds estimated by conventional pure-tone audiometry and phoneme-based self hearing assessment

B. Test time for PhoSHA

Table 2 shows the mean and standard deviation of the time taken by the hearing-impaired group and the normal hearing group to conduct a PhoSHA test. The mean test time for the hearing-impaired group was 95.4 s, but the standard deviation was 44.1 s. The mean test time for the normal hearing group was 28.9 s and the standard deviation was only 3.0 s.

TABLE II
MEAN TEST TIME

Group	N	Age	Mean test time
Hearing Impaired	25	69.9 (12.1)	95.4 (44.1)
Normal Hearing	31	28.9 (5.0)	24.7 (3.0)

The numbers in brackets showed standard deviations.

III. DISCUSSION & CONCLUSIONS

Subjects under 50 years of age understood our experimental procedures and were able to perform the test themselves. On the other hand, most subjects over 50 years of age found it difficult to perform the test using a smartphone. As seen in Table 2, there was a difference in age between normal hearing and hearing-impaired groups. The normal hearing subjects were generally younger than the hearing-impaired subjects.

PhoSHA was conducted in a sound-proof booth. In the real world, it may be conducted without a sound-proof booth in the presence of ambient noise. While this could pose a problem, the canal-type earphones with ear protection muff we use do reduce background noise to a certain extent.

Our method to determine hearing thresholds was a simplified version of the conventional rule of testing at least three times at a candidate level. This was one of main reasons for the dramatic reduction in mean test time for normal hearing subjects. If we implement the conventional rule such as two responses of three trials, the mean test time might increase to approximately twice its current duration.

One set of regression formulae was calculated from all 25 hearing-impaired measurements. To verify how well these formulae work, it is advisable to test them on new data. Therefore, we chose 24 measurements as training data to calculate new sets of regression formulae and used these to predict the measurement for the one remaining ear. This is a method of cross-validation. We obtained 25 predictions using this strategy and compared them with PTA results. The estimates were not as good as those obtained from the first set of regression formulae, but the mean error was only 7.3 dB HL as shown in Figure 5. This value was higher than that in Figure 3, which was 5.6 dB HL.

We proposed a new method of hearing assessment using phonemes and implemented it on a smartphone to allow hearing-impaired people to perform self hearing tests. Regression formulae to estimate the PTA audiogram for six frequencies using PhoSHA with four Korean phonemes were successfully applied. When we did a cross-validation test, the difference was not much higher. In addition, the mean test time was reduced significantly because we used four phonemes instead of six pure tones. We conclude that PhoSHA is suitable for the estimation of audiograms with sufficient accuracy and reduced testing time.

REFERENCE

[1] J. M. Choi, H. B. Lee, C. S. Park, S. H. Oh, and K. S. Park, “PC-based tele-audiometry,” Telemed J E Health, vol. 13, pp. 501-508, Oct. 2007.

[2] L. Honeth, C. Bexelius, M. Eriksson, S. Sandin, J. E. Litton, U. Rosenhall, and O. Nyrén, D. Bagger-Sjöbäck, “An internet-based hearing test for simple audiometry in nonclinical settings: preliminary validation and proof of principle,” Otol Neurotol, vol. 31, pp. 708-714, Jul. 2010.

[3] N. Nakamura, “Development of "MobileAudiometer" for screening using mobile phones,” Conf Proc IEEE Eng Med Biol Soc, vol. 5, pp. 3369-3372, 2004.

[4] J. Sohn, D. Kim, Y. Ku, K. Lee, and J. Lee, “Study on self hearing assessment using speech sounds,” *Conf Proc IEEE Eng Med Biol Soc*, pp. 2384-2387, 2011.

Improved Kanade-Lucas-Tomasi tracker for images with scale changes

Hyoung-Ki Lee*, Ki-Whan Choi, Donggeon Kong, and Jonghwa Won
Samsung Advanced Institute of Technology, Samsung Electronics, Korea

***Abstract*— To match two images with a large scale difference, a scale parameter can be introduced into the warp parameters of the KLT tracker. For some images, the KLT tracker with the scale warp parameter fails to converge. We assume that this result is caused by the singularity of the Hessian matrix. An improved KLT tracker is proposed to avoid this tracking failure. The proposed method introduces an index (scale invariance index) to determine how close the Hessian matrix is to a singular one with the scale warp parameter. According to the index, either of two different sets of warp parameters is selected: one is with the scale warp parameter and the other is without it.**

I. INTRODUCTION

The feature extraction and matching is a key technology to visual SLAM to register visual features as landmarks. Many feature extractors such as the Harris corner detector [1], the SIFT [2], and the KLT tracker [3] have been proposed. They are required to meet the performance metric of maximizing the correct match rate while minimizing the incorrect match rate over a wide range of view direction and scale changes [4]. Especially, robust feature tracking through large scale changes is very critical for the localization for mobile robots because their forward movement with a front view camera causes a lot of scale change of captured images.

SIFT has been shown to have significant invariance to a scale change, but is computationally demanding. To obtain the real time performance, Chekhlov proposed a scheme to avoid the computationally expensive task of constructing scale space representation for each frame and scale invariance is achieved by constructing descriptors over multiple resolutions [5].

In spite that the KLT tracker is not robust to scale changes, the KLT tracker is still a good candidate as a feature extractor because it outperforms the SIFT and the Harris corner detector with a small change in viewpoint [4,6].

In this paper, we propose an improved KLT tracker to deal with scale changes. We claim that there is a group of images for which the conventional KLT tracker with a scale warp parameter diverges. We define a scale invariance index to classify these images. If this index is less than a threshold value, a scale warp parameter is more likely to cause the instability of the tracker and only translation parameters are included in the warp parameters. Otherwise, translation parameters plus a scale warp parameter are used for a warp function.

II. SCALE INVARIANCE INDEX

In general, the KLT tracker can deal with any parametric motion models between two images to align. For example, the dimension of the warp parameter should be 8 to model the projective translation in 2D planar motions.

Here we assume the 3 DOF warp parameters of two translations and a scale change. 3 DOF is good and simple enough to model the camera motion of our robot because it moves forward straightly. The warp of the KLT tracker is given by the following:

$$\mathbf{W}(\mathbf{x};\mathbf{p}) = \begin{pmatrix} s \cdot x + t_x \\ s \cdot y + t_y \end{pmatrix} \tag{1}$$

where $\mathbf{x} = (x, y)^T$ is pixel coordinates and the vector of parameters $\mathbf{p} = (t_x, t_y, s)^T$ is the translation and scale change.

In experiments, we observed that there are some images which make the KLT tracker diverges and fail to track. Fig. 1 (a) shows the typical examples of the tracking failure. It should be noted that the shape of these image patches is almost the same even though they are scaled up. This leads to the result that the determinant of the Hessian matrix [3] is much smaller than that of Fig. 1 (b), which doesn't diverge. We think that the singularity of the Hessian matrix causes the instability of the tracker because the parameter update of the KLT tracker includes the inverse of the Hessian matrix.

We propose an index to determine whether an image is likely to diverge with a scale warp parameter or not. The index, SII (Scale Invariance Index) is given by

$$\mathrm{SII} = \min_{x_0, y_0} \frac{1}{n} \sum_{i=1}^{n} \left((x_i - x_o) \frac{\partial T}{\partial x}(\mathbf{x}_i) + (y_i - y_o) \frac{\partial T}{\partial y}(\mathbf{x}_i) \right)^2 \tag{2}$$

where T is a template image, $\mathbf{x_0} = (x_0, y_0)^T$ is the center of scale change and n is the pixel number of the image. This index is inspired by the mean of the squared intensity change due to scale changes, that is, $\frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial T}{\partial s}(\mathbf{x}_i) \right)^2$.

This value indicates how invariant to scale changes the shape of the image is. Through some calculations, $\mathbf{x_0} = (x_0, y_0)^T$ is given by $\mathbf{x}_0 = (\mathbf{J}_t^T \mathbf{J}_t)^{-1} \mathbf{J}_t^T \mathbf{J}_s$, where $\mathbf{J}_t$ and $\mathbf{J}_s$ are the Jacobian of translation and scale, respectively and are given by

$$\mathbf{J}_t = \begin{bmatrix} \frac{\partial T}{\partial x}(\mathbf{x}_1) & \frac{\partial T}{\partial y}(\mathbf{x}_1) \\ \frac{\partial T}{\partial x}(\mathbf{x}_2) & \frac{\partial T}{\partial y}(\mathbf{x}_2) \\ \vdots & \vdots \\ \frac{\partial T}{\partial x}(\mathbf{x}_n) & \frac{\partial T}{\partial y}(\mathbf{x}_n) \end{bmatrix}, \quad \mathbf{J}_s = \begin{bmatrix} x_1 \frac{\partial T}{\partial x}(\mathbf{x}_1) + y_1 \frac{\partial T}{\partial y}(\mathbf{x}_1) \\ x_2 \frac{\partial T}{\partial x}(\mathbf{x}_2) + y_2 \frac{\partial T}{\partial y}(\mathbf{x}_2) \\ \vdots \\ x_n \frac{\partial T}{\partial x}(\mathbf{x}_n) + y_n \frac{\partial T}{\partial y}(\mathbf{x}_n) \end{bmatrix}$$

Then, inserting $\mathbf{x}_0 = (\mathbf{J}_t^T \mathbf{J}_t)^{-1} \mathbf{J}_t^T \mathbf{J}_s$ into equation (2), we get

$$\mathrm{SII} = \frac{1}{n} (\mathbf{J_t x_0} - \mathbf{J_s})^T (\mathbf{J_t x_0} - \mathbf{J_s}) \tag{3}$$

If SII is smaller than a threshold value, the shape of the image is more invariant to scale changes and likely to diverge with a scale warp parameter.

* Corresponding Author

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

The meaning of the index SII can be explained in a different way. The Hessian matrix of the 3 DOF KLT tracker with the warp parameter of equation (1) is given by

$$H = \begin{bmatrix} \mathbf{J}_t^T \mathbf{J}_t & \mathbf{J}_t^T \mathbf{J}_s \\ \mathbf{J}_s^T \mathbf{J}_t & \mathbf{J}_s^T \mathbf{J}_s \end{bmatrix} \quad (4)$$

The determinant of the equation (4) is given by

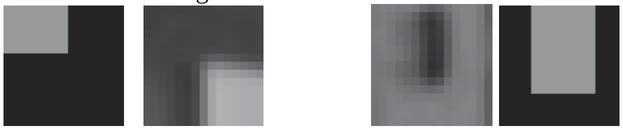
$$\det(H) = \det(\mathbf{J}_t^T \mathbf{J}_t) \cdot \det(\mathbf{J}_s^T \mathbf{J}_s - \mathbf{J}_s^T \mathbf{J}_t (\mathbf{J}_t^T \mathbf{J}_t)^{-1} \mathbf{J}_t^T \mathbf{J}_s) \quad (5)$$

If det(H) is zero, the Hessian matrix is singular and the KLT tracker diverges. Because $\det(\mathbf{J}_t^T \mathbf{J}_t)$ is not zero for corner like features, it is sufficient to examine the value of $\det(\mathbf{J}_s^T \mathbf{J}_s - \mathbf{J}_s^T \mathbf{J}_t (\mathbf{J}_t^T \mathbf{J}_t)^{-1} \mathbf{J}_t^T \mathbf{J}_s)$ to check the singularity of the Hessian matrix. With some calculation, it can be shown that SII is equal to $\frac{1}{n} \times \det(\mathbf{J}_s^T \mathbf{J}_s - \mathbf{J}_s^T \mathbf{J}_t (\mathbf{J}_t^T \mathbf{J}_t)^{-1} \mathbf{J}_t^T \mathbf{J}_s)$. Thus, the smaller SII is, the closer the Hessian matrix is to a singular one.

The proposed algorithm to deal with the scale change in the frame of KLT tracker is given as follows. If the index SII is less than a given threshold value, the warps are modeled by the translations:

$$\mathbf{W(x;p)} = \begin{pmatrix} x + t_x \\ y + t_y \end{pmatrix} \quad (6)$$

where the vector of parameters $\mathbf{p} = (t_x, t_y)^T$ is the optical flow. In this case, two images are similar to each other in spite of the scale change, and the optical flow works well to align them. And if the index, SII, is larger than a given threshold value, the warps are given by the eq. (1). The KLT tracker doesn't diverge with the scale parameter because the determinant of the Hessian matrix is large.



det(H) = 8.1e4 , det(H) = 8.1e4 det(H) = 7.2e5, det(H) = 2.5e6

(a) diverging patches (b) converging patches

Fig.1. KLT tracker with a scale warp parameter

III. EXPERIMENTAL RESULTS

We implemented our algorithm to cleaning robot VC-RE70V. Experiments were performed using the images captured by a cleaning robot, which has a 604×480 resolution front view camera and a dead reckoning system [7]. Fig. 2 shows the captured images and the typical examples of a scale variant patch (above) and a scale invariant patch (below). The threshold of SII for determining the scale invariant patch is 900. SII for the scale invariant patch and scale variant patch are 441 and 2240 respectively. The patches contain the corner points detected by the Harris corner detector. To verify the usefulness of the modified KLT tracker, we evaluated the convergence region ratio and an average tracking error for KLT tracker with and without the scale warp parameter. Varying the initial guess point of the tracker from -7 pixels to 7 pixels for both x and y, we checked the convergence region, where the tracker is converging over a searching region and calculated the average tracking error when converged. To consider that the robot moves forward, the images are scaled down to 0.7 times.

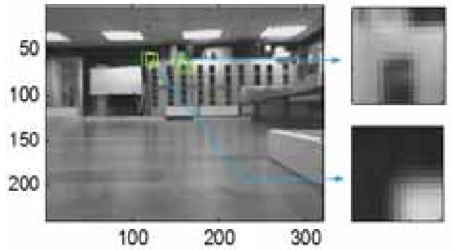


FIG. 2. CAPTURED IMAGES AND PATCHES

Table 1 Convergence region ratio and average tracking error

	Invariant patch		Variant patch	
① tracker without scale	87.6%	0.45 pixel	93.9%	0.97 pixel
② tracker with scale	26.7%	0.00 pixel	71.1%	0.00 pixel

Table 1 shows the results of performance evaluation. The tracker ① in the table means the tracker with the warps of eq. (6) and the tracker ② does the one with the warps of eq. (1). For the scale invariant patch, the convergence region of ② is much smaller than ①. The tracker ① has a pixel error of 0.45 pixel, which is less than 0.5, thus ① can be a choice for this scale invariant patch. On the contrary, for the scale variant patch, the average pixel error of the tracker ① is larger than 0.5 pixel and the tracker ② has a comparable convergence region to the tracker ①. Therefore, the tracker ② is suitable for this scale variant patch. These experimental results show that while the original KLT tracker with a fixed warp has difficulty to track both kinds of patches, the modified KLT tracker to select the warps according to SII can handle them well.

IV. CONCLUSION

An improved KLT tracker to deal with scale changes is proposed. An index (scale invariance index) is introduced to determine how close the Hessian matrix is to a singular matrix when including a scale warp parameter. According to the index, either of two different sets of warp parameters is selectively used: one is the translation and scale warp parameters and the other is the translation parameters only. Through experiments, the efficacy of the algorithm is verified.

REFERENCES

[1] C.Harris and M. Stephens, "A combined corner and edge detector," in Proc. Of Fourth Alvey Vision Conference, Manchester, United Kingdom, pp. 147-151, 1988.

[2] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no.2, pp. 91-110, 2004.

[3] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," International Journal of Computer Vision, vol.56, no.3, pp.221-255, Feb. 2004.

[4] J. Klippenstein and H. Zhang, "Performance Evaluation of Visual SLAM Using Several Feature Extractors," in IEEE/RSJ international conference on Intelligent Robots and Systems, pp. 1574-1581, Oct. 2009.

[5] D. Chekhlov, M. Pupilli, W. Mayol-Cuevas, and A. Calway, "Real-time and robust monocular slam using predictive multi-resolution descriptors," In 2nd Int Symp on Visual Computing, 2006.

[6] J. Klippenstein and H. Zhang, "Quantitative, Evaluation of Feature Extractors for Visual SLAM," Fourth Canadian conference on Computer and Robot Vision, 2007.

[7] Ki-Whan Choi, and et al., "Monocular SLAM with Undelayed Initialization for Indoor Robot," Robotics and Autonomous systems, vol.60, pp. 841-851, June 2012.

Development of a Speech-distortionless Beamformer for Two-microphone Digital Hearing Aids

Jonghee Han[a], Kyeongwon Cho[b], In Young Kim[b], Sung Hwa Hong[c], Dong Wook Kim[a]
[a]Samsung Advanced Institute of Technology, P.O.Box 111, Suwon 440-600, Korea
[b]Department of Biomedical Engineering, Hanyang University, , Seoul, 133-791, Korea
[c]Department of ORL_HLS, Samsung Medical Center, Korea

Abstract-- **To enhance noise reduction performance for digital hearing aids (DHA), this study proposes a speech-distortionless beamformer which is realized by a sequential processing of fractional delay, subtraction and integration (FDSI). A Simulink model of this beamformer was constructed and its performance of speech quality enhancement was evaluated in various noisy circumstances including babble, car and traffic noises. In this simulation, null direction of the beamformer was fixed at 120° azimuth and noise source position was altered from 100 to 140° azimuth. The distance between two microphones also varies from 8 to 20 mm. Compared to conventional beamformers, the FDSI beamformer showed higher frequency-weighted segmental SNR (fwsegSNR) and lower weighted spectral slope (WSS) while perceptual evaluation of speech quality (PESQ) was almost same in every beamformer. These results imply that speech distortion was significantly reduced. Therefore, the FDSI beamformer has great promising for speech quality enhancement in DHA.**

I. INTRODUCTION

In the field of digital hearing aids (DHA), microphone array processing techniques have shown better noise reduction performance than single microphone techniques [1, 2]. Beamforming, which suppresses sound coming from unwanted direction, is one of the most prevalent techniques in microphone array processing. In DHA, the first-order differential array has been commonly used as a beamforming method [3]. The first-order differential array can attenuate directional noise to a great extent but the shape of speech signal could change due to the process of differentiation. In addition, an equalization filter such as a first-order butterwoth low-pass filter is required to get uniform frequency response [4]. Since low-frequency gain in the equalization filter is very high, low-frequency signal disturbances can be also strongly amplified and speech signal is also distorted. Recently, a broadband beamformer using optimization methods was developed to reduce frequency dependency [5]. However, the broadband beamformer cannot have directionality in very low-frequency region.

To overcome these limitations and enhance speech quality, a speech-distortionless beamformer is proposed in this study. The proposed method modified the first-order differential array by replacing the equalization filter with integrator. The performance of this new beamformer was evaluated based on various speech quality measures via computer simulation.

II. METHODS AND MATERIALS

A. Beamforming using Fractional Delay, Subtraction and Integration (FDSI)

A far-field sound acquisition model was used to model two microphones' signals of DHA. A fractional delay (internal delay, δ_f) was applied to one microphone signal. Then, the other signal was subtracted from the delayed signal. These two steps can be formulated as follow:

$$x_1(t-\delta_f) - x_2(t) = s(t-\delta_f) - s(t-\delta_\theta) \qquad (1)$$

where δ_θ is an external delay by sound propagation. When the difference between the internal and external delay is small enough, the signal s in the interval [$t-\delta_\theta$, $t-\delta_f$] can be assumed to be linear. Therefore, equation (1) can be rewrited as

$$x_1(t-\delta_f) - x_2(t) = (\delta_\theta - \delta_f)\frac{\partial s(t-\delta_\theta)}{\partial t} \qquad (2)$$

By applying integration on both sides of equation (2) in discrete domain, we can finally get

$$\frac{1}{q(1-\cos\theta_f)}\sum_{k=0}^{n}(x_1[k-\tau_f]-x_2[k]) = \frac{\cos\theta-\cos\theta_f}{1-\cos\theta_f}s[n-\tau_\theta] \qquad (3)$$

The left side of equation (3) represents a core procedure of the FDSI beamformer and the right side is a processing result which is a product of the original signal and a directional gain. This means that speech signal can preserve its own shape. By changing θ_f , null direction can be easily adjusted.

B. Simulation and Evaluation

Two sound sources were used in this simulation and mixed in an additive manner. The first source was a speech signal that was selected from among the ten sentences in the IEEE corpus [6] and assumed to be located at the front side ($\theta = 0$). The second source was chosen from among 7 car noises, 9 traffic noises and 18 babble noises and positioned at incident angles (θ) of 100, 110, 120, 130, and 140˚ azimuth. The speech and noise signals were mixed at the input signal-to-noise ratio (iSNR) of 0 dB. In order to evaluate the effect of the distance between microphones, the distance varied from 0.8 cm to 2.0 cm in 0.2 cm increments. For these simulation environments, three beamforming algorithms were simualated including a

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

directional microphone (DM) [3], a broadband beamformer (BBF) [5] and the FDSI beamformer.

For the processed signal, objective speech quality measures such as frequency-weighted segmental SNR (fwsegSNR), weighted spectral slope (WSS) and perceptual evaluation of speech quality (PESQ) were calculated because these speech quality measures are highly correlated with the subjective speech quality measure [7].

III. RESULTS & DISCUSSION

A. *Evaluation according to the noise position*

Fig. 1 shows the evaluation results according to the noise position. When a noise position is the same as the null direction, the FDSI beamformer showed more than 10 dB fwsegSNR enhancement, which is better than conventional beamformers. The bigger a difference between the null direction and a noise position was, the worse speech quality was. However, even when the difference was 20˚, the FDSI showed more than 5 dB fwsegSNR enhancement for every noise type. As to WSS, the FDSI beamformer showed the lowest value in every case. These results means that speech distortion is minimized in the FDSI beamformer.

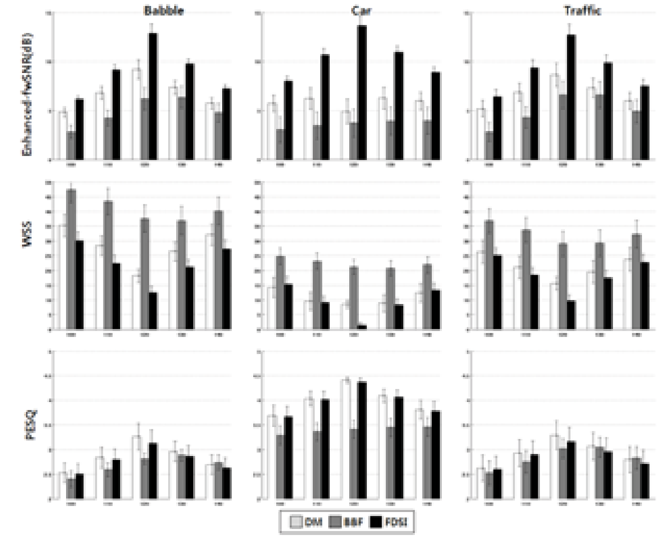


Fig. 1. Objective speech quality measures at various noise angles (100˚, 110˚, 120˚, 130˚, and 140˚) with a fixed null direction (120˚). The mean values and standard deviations of the enhanced-fwSNR, WSS and PESQ were compared.

B. *Evaluation according to the distance between microphones*

Fig. 2 shows the evaluation results according to the distance between microphones when a noise position corresponds to the null direction. For babble and traffic noises, the FDSI was not significantly different according to distance except that PESQ increased slightly. With car noise, the fwSNR decreased as the distance increased.

IV. CONCLUSION

In this study, a speech-distortionless beamformer which is capable of preserving the original signal shape was developed through the combination of subtraction and integration and evaluated by Simulink simulation. When the distance between microphones is small just like the case in DHA, the FDSI beamformer showed better performance than conventional ones based on SNR and WSS. Overall, the proposed beamformer has promise to contribute to the noise reduction system in DHA.

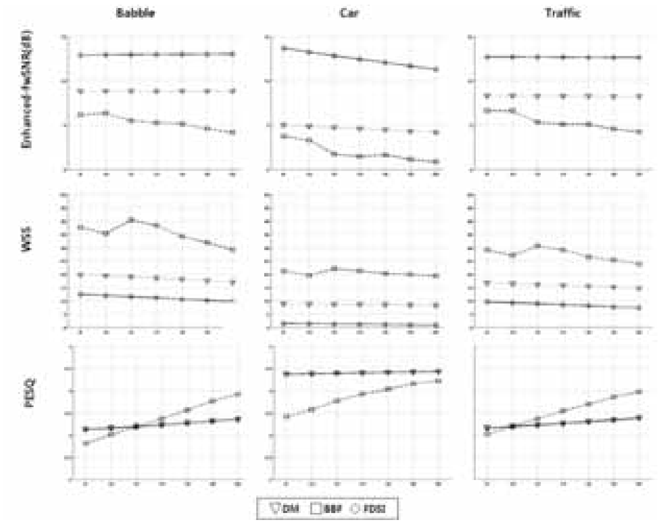


Fig. 2. Objective speech quality measures according to the distance between two microphones. The x-axis indicates the distance in mm.

ACKNOWLEDGMENT

This work was supported by grand No. 10031741 from the Strategic Technology Development Program of Ministry of Knowledge Economy.

REFERENCES

[1] W. Soede, A.J. Berkhout, F.A. Bilsen, Development of a Directional Hearing Instrument Based on Array Technology, *Journal of the Acoustical Society of America* **94** 785-798 (1993)

[2] R.W. Stadler, W.M. Rabinowitz, On the Potential of Fixed Arrays for Hearing-Aids, *Journal of the Acoustical Society of America* **94** 1332-1342 (1993).

[3] J.M. Kates, *Adaptive and Multimicrophone Arrays, Digital Hearing Aids* (Plural Publishing, Abingdon, 2008)

[4] M. Buck, M. Robler, First Order Differential Microphone Arrays For Automotive Applications, *7th International Workshop on Acoustic Echo and Noise Control, Darmstadt, Germany*, 2001.

[5] E. Mabande, A. Schad, W. Kellermann, Design of Robust Superdirective Beamformers as a Convex Optimization Problem, *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process., Taipei, Taiwan* 77-80 (2009)

[6] E.H. Rothauser, W.D. Chapman, N. Guttman, M.H.L. Hecker, K.S. Nordby, H.R. Silbiger, G.E. Urbanek, M. Weinstock, IEEE recommended Practice for Speech Quality Measurements, *IEEE Transactions on Audio Electroacoustics* **17** 225-246 (1969)

[7] Y. Hu, P.C. Loizou, Evaluation of objective quality measures for speech enhancement, *IEEE Transactions on Audio Speech and Language Processing* **16** 229-238 (2008)

Dialogue Enabling Speech–to–Text User Assistive Agent with Auditory Perceptual Beamforming for Hearing-Impaired

* Seongjae Lee, * Sunmee Kang, ** Hanseok KO, **Jongseong Yoon, **Minseok Keum
* School of Electronics and Computer Engineering, Seokyeong University, Seoul, Korea
** School of Electrical Engineering, Korea University, Seoul, Korea

***Abstract*—A novel approach for assisting effective bidirectional communication of intentions between people of normal hearing and hearing-impaired is presented in this paper. In particular, we demonstrate the intelligent hearing assistive utility by incorporating a speech-to-text interface such that the intended speaker's speech is directionally contained by means of a multichannel acoustic beamformer. Clear understanding of the normal hearing person's speech can reduce the burden on the hearing impaired in terms of required explicit response under context sensitive dialogue scenarios. Hence, hearing impaired friendlier interface device perceiving the utterances of normal hearing person is highly desirable. The proposed interface is a portable device with directional sound selection capability that performs speech-to-text task of normal hearing person's speech in diverse dialogue contexts. The relevant experimental results have confirmed that the proposed interface design is a feasible approach for realizing an effective and efficient intelligent agent, as it reduces the amount of work required from hearing impaired user who is interacting with the normal hearing people.**

I. INTRODUCTION

A novel approach for assisting effective bidirectional communication of intentions between people of normal hearing and hearing-impaired is by incorporating a speech-to-text (STT) interface such that the intended speaker's utterances are perceptually contained by means of a multichannel acoustic beamformer. Clear understanding of the normal hearing person's speech can reduce the burden on the hearing impaired in terms of required explicit response under context sensitive dialogue scenarios. Hence, hearing impaired friendlier agent device is highly desirable.

A previous study that utilized a speech recognizer [3] for STT task showed limitations in its effectiveness mainly due to the absence of background noise cancellation and insufficient number of vocabularies reserved for the speech recognition. In addition, the system designed was for very limited situations and did not pay attention to the user interface requirements.

Therefore, this paper proposes a portable audio perceptual device equipped with an advanced speech recognition module incorporating directional sound selection capability that performs STT function of normal hearing person's speech in diverse dialogue contexts. The proposed system is embedded in a mobile device with a user-friendly design such that it will be able to render effective assistance to the hearing-impaired by offering a more efficient and natural way of dialogue.

II. AUDITORY PERCEPTUAL USER ASSISTIVE AGENT

A. Structure of Auditory Perceptual User Assistive Agent

Fig .1 illustrates the overall structure of the proposed auditory perceptual user assistive agent system. The system is composed of three main components: device layer, algorithm layer and user interface layer. The device layer includes a multi-channel microphone for speech input and a speaker for speech output. An algorithm layer includes a STT module for recognizing the intended speaker's speech and a preprocessing module for directionally channeling the intended speaker's voice to the system with auditory perceptual function. A user interface layer allows the user to friendly control the system through a screen effectively.

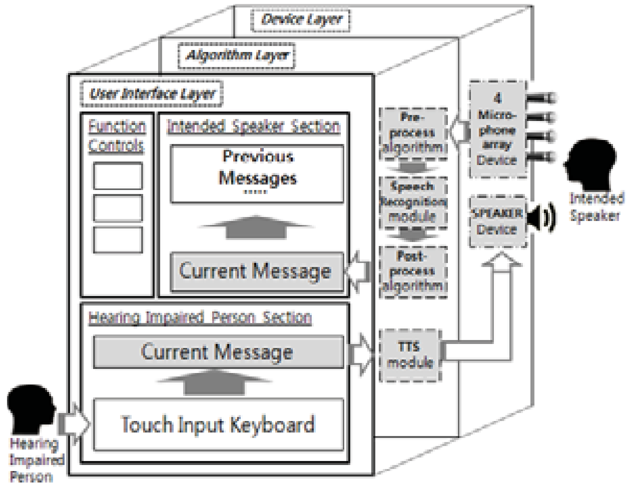


Fig. 1. Overall structure of the proposed system

B. Speech-to-Text

The preprocessing component in the STT module is essentially an auditory perceptual beamformer that clarifies the intended speaker's speech by suppressing the background noise prior to transferring into the speech recognition engine. It consists of TF-GSC (Transfer Function Generalized Sidelobe Canceller) [4] and Post-filter based OM-LSA (Optimally Modified Log Spectral Amplitude estimator) [5] . TF-GSC is a GSC based beamforming algorithm, which contains the sound imported through a multi-channel microphone in a specific direction as "desired signal" while defining the sound imported in other directions as background noise and cancelling it. It also compensates the characteristics of transfer function among microphones for sharp reduction of sound sources from unintended directions. OM-LSA is a Bayesian estimator-based algorithm, which is appropriate for non-stationary noise environment. It distinguishes the speech from noise based on the probability of speech presence. The preprocessing module in the proposed system first implements primary noise cancellation by applying TF-GSC to the signal imported through the multi-channel microphone. It then implements OM-LSA to

This research was supported by Ministry of Health & Welfare R&D (A111189) and in part by Business for Academic-industrial Cooperative establishments funded Korea SMBA in 2011 (00045925)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

the output and finally feed the noise-cancelled speech signal to the speech recognizer for performing the STT task.

The postprocessing module's goal is to extract keywords from the recognized speech by analyzing the content of the keywords assumed important for dialogues in several but still limited contextual situations in real life such as in doctor's office or pharmacy. The Spoken Language Understanding technique is applied to detecting keywords from those dialogues. In order to extract the keywords from continuous speech in each situation, the postprocessing module extracts the key words, which match each context by screening the speech from LVSCR (Large Vocabulary Continuous Speech Recognition) [6]. This method first implements a semantic tagging with the morpheme sequence extracted from the morpheme analysis and then exports the keywords by comparing with the existing keyword group which was set up for each specific situations. .

C. Hearing-Impaired Friendly User Interface

The main goal of constructing the user interface for the proposed system is to function as an agent between the hearing-impaired and the normal hearing person and provides a user-friendly manipulating environment for the hearing-impaired person. Accordingly, the user interface assists the bilateral communication by incorporating both STT and TTS modules into the system and provides a touch screen interface embedded in a portable device.

Fig. 1 illustrates the overall flow of the proposed user interface. First, the hearing-impaired user informs the intended speaker that the user is in need of using the proposed system with a gesture. Second, the intended speaker can transfer intention to the hearing-impaired person via speech, and then the user interface clarifies the client's speech and displays the extracted keywords on the screen using STT module. These two steps are independently carried out until the end of the conversation.

The user interface screen of the proposed system is constructed based on the Hearing-impaired person Scenario Modeling for user-friendly manipulation. It mainly consists of a context configuration menu and a dialogue menu. The place of context configuration menu illustrated in Fig 2 is continuously rearranged according to the frequency of use to improve the user convenience. The dialogue screen is the core element of the User Interface. In order to maximize the user readability, common vocabularies among the extracted key words are displayed in pictogram format. Also, the icons would be displayed on the left side of the screen if the user is right-handed and vice versa.

III. EXPERIMENTAL RESULTS

The system evaluation is performed in order to test the levels of functional improvement in noise cancellation, key word spotting after the preprocessing and postprocessing steps, and ease of use. For the preprocessing module, four microphones were used to provide 30 degree acoustic beamwidth which relate to effective speech recognition performance in noisy environment scenarios. For the STT component, the test was conducted by using 100 isolated words from the ETRI DB. Speech babble from NOISE EX-92 (CMU) was used as the noise database. For the postprocessing module performance evaluation, two types of language models with a recognition system of 2,000 words and 20,000 words were used. Tables I and II show the results (WER: 8.5% with 4.6 dB gain) from the performance evaluations done for the preprocessing and the postprocess modules, respectively. Table III indicates that the ease of use is satisfactory compared to the existing STT mobile application for normal hearing person [7].

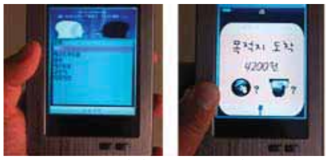


Fig. 2. Prototype of the proposed user interface

TABLE I
Performance evaluation of the preprocess module

Word Error Rate(WER) [%]		Input Signal SegSNR [dB]	Output Signal SegSNR [dB]	SegSNR Improvement [dB]
Without noise Reduction	GSC + OM-LSA			
18.4	8.7	10	16.6	6.6
12.6	8.2	15	17.9	2.9

TABLE II
Performance evaluation of the postprocess module

Language Model	Scale of the recognition system		Success rate Improvement [%]
	2,000 words [sentence accuracy]	20,000 words [sentence accuracy]	
Bi-gram	71.9%	82.0%	10.1
Tri-gram	74.0%	80.9%	6.9

TABLE III
Evaluation of the user interface (worst : 0 best : 7)

USEFULNESS				EASY OF USE			
Existing	3.7	Proposed	5.9	Existing	3.5	Proposed	6.2
EASE OF LEARNING				SATISFACTION			
Existing	3.5	Proposed	6.2	Existing	2.4	Proposed	6.1

IV. CONCLUSION

This paper proposed a context-based dialogue assistant system for people with hearing impairment. The preprocessing module with TF-GSC and OM-LSA was used to improve the recognition quality, and the postprocessing module was used for keyword spotting. As a result, the proposed system was proven to have robust performance against background noise compared with existing speech recognition systems. Also, the user-friendly interface of the proposed system is expected to improve the user satisfaction level.

References

[1] Jounghoon Beh, Robert H. Baran, and HANSEOK KO, "Dual Channel Based Speech Enhancement Using Novelty Filter for Robust Speech Recognition in Automobile Environments", IEEE International Conference on Consumer Electronics., pp243-244, January, 2006

[2] Seokyeong Jeong, Kyoung Won Min, and HANSEOK KO, "Fast Decoder Design of Connected Word Speech Recognition for Automobile Navigation System", IEEE International Conference on Consumer Electronics., pp215-216, January, 2006

[3] Y. Zhao, X.Zhang, R-S. Hu, J. Xue, X. Li, L.che, R.hu, L. Schopp, "An Automatic Captioning System for Telemedicine", ICASSP IEEE, 2006

[4] S. Gannot, "Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech", IEEE trans Signal processing, Aug, 2001

[5] Israel Cohen, Baruch Berdugo, "Speech enhancement for non-stationary noise environments", Signal processing, June, 2001

[6] Huggins-Daines, D., Kumar, M., Chan, A., Black,A.W., Ravishankar, M. and Rudnicky, A.I."PocketSphinx: a free real-time continuous speech recognition system for hand-held devices." Proc. ICASSP 2006, IEEE Press (2006), 185-188.

[7] www.nuancemobilelife.com/apps/dragon-dictation , "Dragon Dictation"

Wireless Powering Management by Analog Circuitry Based In-Band Signaling Controller

Dong-Zo Kim, Ki Young Kim, Young-Ho Ryu, Nam Yoon Kim, Yun-Kwon Park, and Sangwook Kwon
Future IT Research Center, Samsung Advanced Institute of Technology, Yongin 446-712, Korea

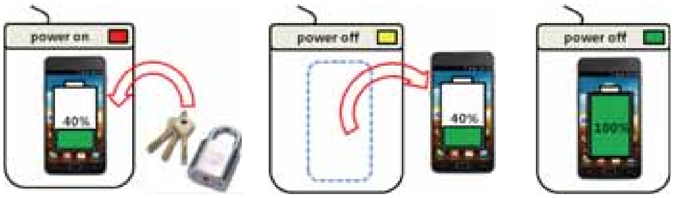
***Abstract*—Efficient wireless charging control scheme for powering a single device unit is presented. The proposed scheme uses in-band communication and analog circuits to control the clock signal lengths. Since it can generate the control data packets without MCU, it is easy to implement a one-chip of the RX parts. We have verified the performance of the in-band communication for the WPT system while a mobile device is in charging.**

I. INTRODUCTION

The resonant magnetic coupling can significantly enhance the efficiency and transfer range of a wireless power transmission (WPT) system and the applications of the WPT are currently ranging from mobile or biomedical device charging platform to wireless electric vehicle charging [1]. Communication and control are needed for supplying stable power to target devices by monitoring the charging status. Complicated charging protocol can enhance the operation stability for wireless charging. But in case of a single device charging, i.e. TX/RX is 1:1, this complicated charging protocol is inefficient. And a micro control unit (MCU) in the RX part is needed for generating communication and control data, but the design and fabrication processes of RX IC and MCU IC are quite different from each other, so it is difficult to implement a one-chip IC solution. Accordingly, in this paper, the efficient charging control scheme is proposed in powering single device unit without MCU for a one-chip implementation of the RX part. The proposed control scheme uses in-band communication [2-3] and analog circuits to control the clock signal lengths.

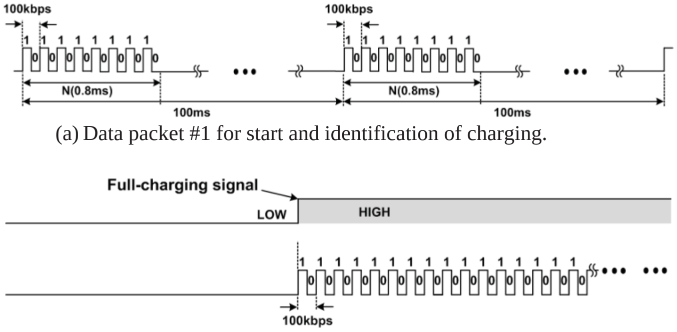
II. SYSTEM DESIGN AND MEASUREMENT

There are three potential use-cases for wireless powering a single device unit shown in Fig. 1. At first, in order to transmit power, the TX system should know whether a target device or foreign metal objects. Then, if a target device in charging leaves the charging area, TX system should stop powering. Finally, if a target device reaches a full-charging status, TX system should finish the powering. To distinguish these 3 cases, the device needs to offer at least two kinds of data packets to TX system as shown in Fig. 2. The N-data packet (100kbps) in Fig. 2(a) is generated every 100ms for TX system to recognize the target device and to transmit power consistently. This data packet #1 can cover the conditions of Fig. 1(a) and (b). When the target device reaches the full-charging, the continuous data packet in Fig. 2(b) is generated. Then, the TX system recognizes the status of the device and finishes the powering. This data packet #2 is for the condition of Fig. 1(c).



(a) Start (b) Stop (c) Finish

Fig. 1. Three use-cases for wireless powering single device unit.



(a) Data packet #1 for start and identification of charging.

(b) Data packet #2 for completion of charging.

Fig. 2. Proposed control data packets using clock signal lengths.

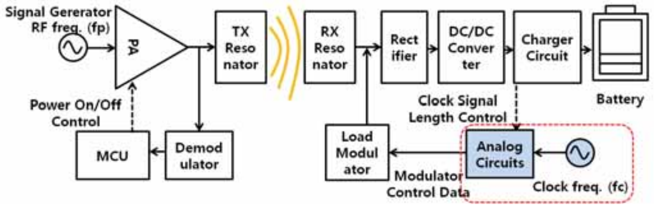


Fig. 3. Concurrent wireless power and signal transfer system with the proposed in-band communication and control.

Fig. 3 shows the proposed in-band communication system. The analog circuits control the length of 100kHz clock signal (fc) to generate two data packets shown in Fig. 2. The data packets are sent to the load modulator and the modulated TX waveform is generated, and the TX system demodulates those data packets.

Fig. 4 shows a schematic of analog circuit to generate the proposed data packets shown in Fig. 2. When the power is received, the 3-bit counter is enabled every 100ms by 10Hz sampling clock (fs), it makes 8-clock data packet (100kHz) which is corresponding to data packet #1 shown in Fig. 2(a). When the device reaches the full-charging (Vout>Vref), the continuous data packet (100kHz) shown in Fig. 2(b) is generated.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

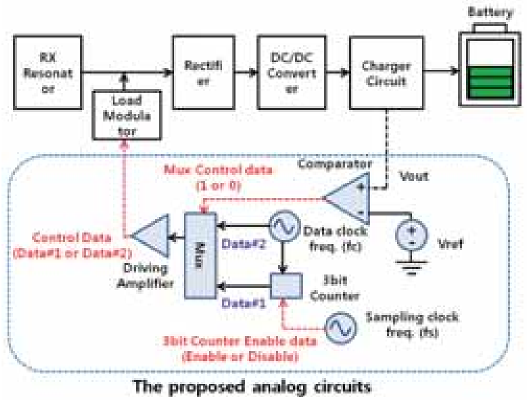
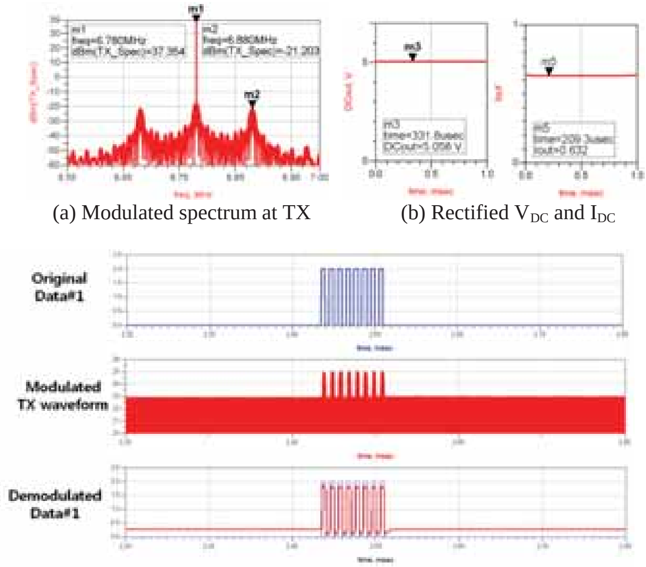


Fig. 4. Schematic of analog circuitry generating the proposed data packets based on clock signal lengths.



(a) Modulated spectrum at TX

(b) Rectified V_{DC} and I_{DC}

(c) Data waveforms

Fig. 5. Simulation results of the WPT system with in-band communication.

Fig. 5 shows the simulation results of the in-band communication in the proposed system by using ADS Ptolemy tools. This simulation intended to analyze the demodulation performance for the data packet #1 under wireless charging of a mobile phone. Fig. 6 shows the experimental setup for the measurement of power transfer efficiency and the performance of the proposed in-band communication. The input of the RX system was connected to a RX resonator and the output was connected to the battery charger of a mobile phone. The dimension of the RX resonator is 4.5cm×6.5cm and it was embedded in the rear case of a mobile device while a TX resonator with 15cm×15cm was used and the efficiency between the resonators is ~80%. The received power is 3.27W (4.96V and 660mA) while transmitting power is 5.4W (37.35dBm) at 6.78MHz and the total efficiency is ~55% when the efficiency of the TX system is ~90%. Because the occurrence of the ON state of the load modulation is few (0.8ms per 100ms), the decrease in power transfer efficiency with proposed in-band communication is less than 1%. Fig. 7 shows the measured three waveforms for the data transmission and reception between the TX and the RX systems.

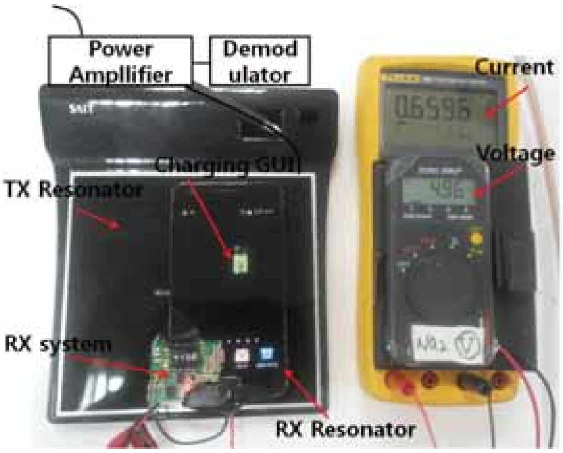


Fig. 6. Experimental setup.

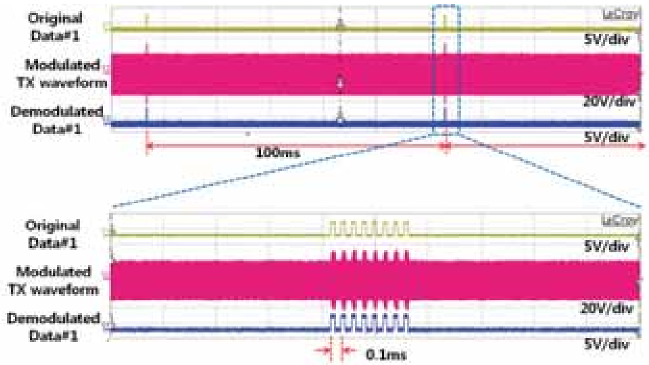


Fig. 7. Experimental results of in-band communication.

The original data packet #1 (yellow) of 100 kHz is generated every 100ms and it makes the modulated TX waveform (magenta). And the demodulator in the TX system makes the demodulated data packet #1 (blue). The measured results are good agreements with the simulated results shown in Fig. 5.

III. CONCLUSION

In this work, a simplified wireless charging management scheme using in-band communication has been proposed for one-chip implementation of RX part. The analog circuits to generate the control data based on clock signal lengths were also designed and fabricated. We have successfully verified the performance of the in-band communication for the WPT system while a mobile device is in charging.

REFERENCES

[1] A. Kurs, A. Karalis, R. Moffatt, J. D. Joannopoulos, P. Fisher, and M. Soljacic, “Wireless power transfer via strongly coupled magnetic resonances,” *Science*, vol. 317, no. 5834, pp. 83-86, July 2007.

[2] J. –H. Cho, and P. H. Cole, “An NFC transceiver using an inductive powered receiver for passive, active, RW and RFID modes,” *IEEE Trans. Ind. Electron.*, vol. 57, no. 5, pp. 456-459, May 2010.

[3] Z. Tang, B. Smith, J. H. Schild, and P. H. Peckham, “Data transmission from an implantable biotelemeter by load-shift keying using circuit configuration modulator,” *IEEE Trans. Biomed. Eng.*, vol. 42, no. 5, pp. 524-528, May 1995.

A New Rate-2 2×2 STBC with Low Complexity ML Detection

Sung Ik Park*, *Member, IEEE*, Heung Mook Kim*, *Member, IEEE*, Namho Hur*, and Jeongchang Kim**, *Member, IEEE*

*Terrestrial Broadcasting Technology Research Team, ETRI, Daejeon, 305-350, Korea
**Dept. of Electronics and Communications Engineering, Korea Maritime University, Busan, 606-791, Korea

***Abstract*--In this paper, we propose a new rate-2, 2×2 space-time block-code (STBC) with low complexity maximum likelihood (ML) detection. We focus on the operation under low signal-to-noise ratio (SNR) regions. Numerical results show that the proposed STBC outperforms the conventional rate-2, 2×2 STBC of WiMAX under low SNR regions. Also, the ML detection complexity of the proposed STBC is significantly reduced.**

I. INTRODUCTION

Recently, a rate-2, 2×2 space-time block code (STBC) achieving full diversity, referred to as Golden code, was proposed in [1]. Also, a variant of the Golden code, referred to as *Matrix C*, was included in the IEEE 802.16e-2005 WiMAX standard [2]. Though the Golden code achieves the optimum diversity-multiplexing trade-off, the maximum likelihood (ML) detection complexity with Q-ary quadrature amplitude modulation (QAM) is proportional to Q^4, which quickly becomes impractical as Q increases. In order to reduce the detection, several multiple-input multiple-output (MIMO) detection algorithms were applied to the detection of the Golden code [3]-[5]. In [6], on the other hand, a code of the reduced detection complexity was developed with a slight sacrifice of the performance compared to *Matrix C*.

For the practical communication systems operating at low signal-to-noise ratio (SNR) regions with powerful error correction codes, since the diversity gain is realized at high SNR regions, maximizing the coding gain is more important than achieving the diversity gain. In this paper, we propose a new rate-2, 2×2 STBC with low complexity ML detection. We focus on the operation under low SNR regions. Numerical results show that the proposed STBC outperforms the conventional rate-2, 2×2 STBC of WiMAX under low SNR regions. Also, the ML detection complexity of the proposed STBC is significantly reduced compared to *Matrix C*.

II. SYSTEM AND CHANNEL MODELS

We start with the *Matrix C* of WiMAX [2]. The signal matrix is given as $\mathbf{X} \triangleq \begin{bmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{bmatrix} = \frac{1}{\sqrt{1+r^2}} \begin{bmatrix} s_1 + jrs_4 & rs_2 + s_3 \\ s_2 - rs_3 & jrs_1 + s_4 \end{bmatrix}$ where the signals s_1, s_2, s_3, s_4 denote Q-ary modulated symbols and $r = (-1+\sqrt{5})/2$. Then, the signal $x_{n,t}$ is transmitted on the n-th transmit antenna at the t-th time interval.

We assume that the signals transmitted on distinct transmit antennas experience independent Rayleigh fading. Also, we assume that the channel response does not vary significantly during the transmission of a signal matrix. Then, the matched filter output at the m-th receive antenna on the t-th time interval is then given by $y_{m,t} = (\sqrt{\rho}/2)\sum_{n=1}^{2} h_{m,n} x_{n,t} + w_{m,t}$, m=1,2, t=1,2. Here, $w_{m,t}$ is the contribution of the additive white Gaussian noise (AWGN) with a double sided power spectral density of 1/2 and zero mean. The channel coefficients $h_{m,n}$ denote the complex channel gain between the n-th transmit and m-th receive antennas and are zero mean complex Gaussian random variables with variance 1/2 per dimension.

Finally, the transmit power of each transmit antenna is normalized so that the total average transmit power is equal to the case of a single transmit antenna system. Therefore, the average received SNR at each receive antenna is ρ. Then, the received signal matrix can be written as

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & y_{1,2} \\ y_{2,1} & y_{2,2} \end{bmatrix} = \sqrt{\frac{\rho}{2}} \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{bmatrix} \begin{bmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{bmatrix} + \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} \quad (1)$$

where $\mathbf{H}=\{h_{m,n}\}$ is the channel response matrix. The received signals can be rearranged as $\mathbf{y} = (\sqrt{\rho}/2)\bar{\mathbf{H}}\mathbf{s} + \mathbf{w}$ where $\mathbf{y} = [y_{1,1}, y_{1,2}, y_{2,1}, y_{2,2}]^T$, $\mathbf{s} = [s_1, \cdots, s_4]^T$, $\mathbf{w} = [w_{1,1}, w_{1,2}, w_{2,1}, w_{2,2}]^T$ and

$$\bar{\mathbf{H}} = \frac{\sqrt{2}}{\sqrt{1+r^2}} \begin{bmatrix} h_{1,1} & h_{1,2} & -rh_{1,2} & jrh_{1,1} \\ jrh_{1,2} & rh_{1,1} & h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} & -rh_{2,2} & jrh_{2,1} \\ jrh_{2,2} & rh_{2,1} & h_{2,1} & h_{2,2} \end{bmatrix}.$$

At the receiver, assuming perfect channel state information, the ML detection rule is given as

$$(\hat{s}_1, \hat{s}_2, \hat{s}_3, \hat{s}_4) = \underset{\mathbf{s}=[s_1,s_2,s_3,s_4]^T}{\arg\min} \left\| \mathbf{y} - \frac{\sqrt{\rho}}{2}\bar{\mathbf{H}}\mathbf{s} \right\|^2 \quad (2)$$

III. PROPOSED RATE-2, 2×2 STBC

We present a new rate-2, 2×2 STBC with good performance under low SNR regions. By modifying the structure of *Matrix C*, we propose a new 2×2 signal matrix as follows:

$$\mathbf{X}_{\text{new}} = \begin{bmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{bmatrix} = \frac{1}{\sqrt{1+r^2}} \begin{bmatrix} s_1 + rs_2 & s_3 + rs_4 \\ s_2 - rs_1 & s_4 - rs_3 \end{bmatrix} \quad (3)$$

This research was supported by the Korea Communications Commission (KCC), Rep. of Korea, under the support program supervised by the Korea Communications Agency (KCA) [Development of Wake-up Emergency Alert Broadcast System Using DTV].

Corresponding author: Jeongchang Kim (jchkim@hhu.ac.kr)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Then, using $\mathbf{X}_{\text{new}}$, the received signal matrix is given as

$$\mathbf{Y}=\sqrt{\frac{\rho}{2}}\begin{bmatrix}h_{1,1} & h_{1,2}\\ h_{2,1} & h_{2,2}\end{bmatrix}\mathbf{X}_{\text{new}}+\begin{bmatrix}w_{1,1} & w_{1,2}\\ w_{2,1} & w_{2,2}\end{bmatrix} \tag{4}$$

The received signals can be rearranged as $\mathbf{y}=(\sqrt{\rho}/2)\bar{\mathbf{H}}_{\text{new}}\mathbf{s}+\mathbf{w}$ where

$$\bar{\mathbf{H}}_{\text{new}}=\frac{\sqrt{2}}{\sqrt{1+r^2}}\begin{bmatrix}h_{1,1}-rh_{1,2} & rh_{1,1}+h_{1,2} & 0 & 0\\ 0 & 0 & h_{1,1}-rh_{1,2} & rh_{1,1}+h_{1,2}\\ h_{2,1}-rh_{2,2} & rh_{2,1}+h_{2,2} & 0 & 0\\ 0 & 0 & h_{2,1}-rh_{2,2} & rh_{2,1}+h_{2,2}\end{bmatrix} \tag{5}$$

denotes the equivalent channel matrix for the symbol vector **s**. Then, the ML metric can be computed as

$$(\hat{s}_1,\hat{s}_2,\hat{s}_3,\hat{s}_4)=\underset{\mathbf{s}=[s_1,s_2,s_3,s_4]^T}{\arg\min}\left\|\mathbf{y}-\frac{\sqrt{\rho}}{2}\bar{\mathbf{H}}_{\text{new}}\mathbf{s}\right\|^2 \tag{6}$$

From the code structure, we can know that the symbol pairs (s_1,s_2) and (s_3,s_4) are transmitted at first and second time slots, respectively. Hence, pairs (s_1,s_2) and (s_3,s_4) are decoupled and thus, the ML metric of (6) can be separated as follows:

$$(\hat{s}_1,\hat{s}_2)=\underset{(s_1,s_2)}{\arg\min}\left\|[y_{11}\quad y_{21}]^T-\left(\sqrt{\rho}/2\right)\bar{\mathbf{H}}_{\text{new},1}[s_1\quad s_2]^T\right\|^2 \tag{7}$$

$$(\hat{s}_3,\hat{s}_4)=\underset{(s_3,s_4)}{\arg\min}\left\|[y_{12}\quad y_{22}]^T-\left(\sqrt{\rho}/2\right)\bar{\mathbf{H}}_{\text{new},1}[s_3\quad s_4]^T\right\|^2 \tag{8}$$

where $\bar{\mathbf{H}}_{\text{new},1}=\frac{\sqrt{2}}{\sqrt{1+r^2}}\begin{bmatrix}h_{1,1}-rh_{1,2} & rh_{1,1}+h_{1,2}\\ h_{2,1}-rh_{2,2} & rh_{2,1}+h_{2,2}\end{bmatrix}$. Therefore, the ML detection complexity of the proposed STBC is proportional to Q^2.

IV. NUMERICAL RESULTS

We will only consider the case with two receive antennas. Figs. 1 and 2 show the average bit error rate (BER) performances of *Matrix C* and the proposed 2×2 STBC. In Fig. 1, the proposed STBC with 16-QAM and ML detection obtains the SNR gain of approximately 0.6 dB at BER=10^{-1}. Fig. 2 shows the comparison of BER performances of *Matrix C* and the proposed STBC with zero-forcing (ZF) detection using QPSK, 16-QAM, and 64-QAM. The proposed STBC obtains the SNR gains of approximately 0.6 dB, 0.7 dB, and 0.8 dB at BER=10^{-1} for QPSK, 16-QAM, and 64-QAM with ZF detection, respectively. Note that the SNR gain of the proposed STBC increases with increasing modulation order. Numerical results show that the proposed 2×2 STBC improves the BER performance under low SNR regions compared to *Matrix C* defined in IEEE 802.16e-2005 standard.

V. CONCLUSIONS

In this paper, we proposed a new rate-2, 2×2 STBC with low complexity ML detection. Since the practical systems operate at low SNR regions with powerful error correction codes, maximizing the coding gain is more important than achieving the diversity gain. Therefore, we focused on the operation under low SNR regions. Numerical results show that the proposed STBC outperforms the conventional 2×2 STBC for WiMAX under low SNR regions. Also, the ML detection complexity of the proposed STBC is significantly reduced compared to *Matrix C* defined in IEEE 802.16e-2005 standard.

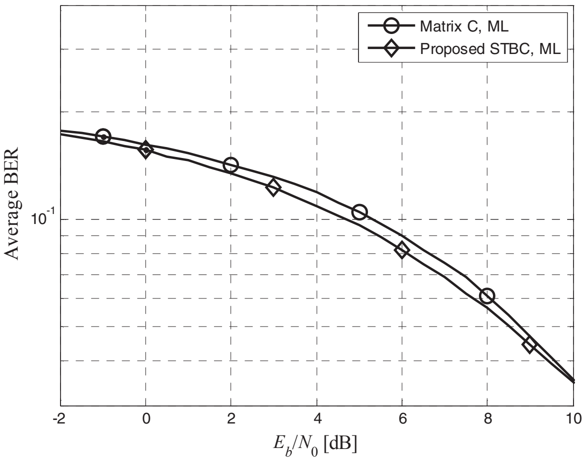


Fig. 1. Average BER performance of *Matrix C* and the proposed STBC with 16-QAM and ML detection.

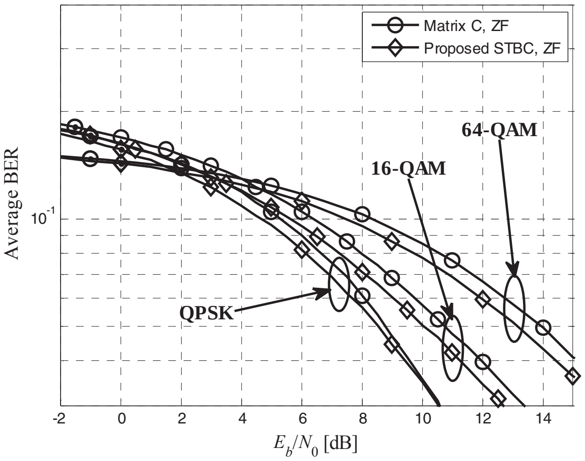


Fig. 2. Average BER performance of *Matrix C* and the proposed STBC with ZF detection.

REFERENCES

[1] J.-C. Belore, G. Rekaya and E. Viterbo, "The Golden code: A 2x2 full-rate space-time code with nonvanishing determinants," *IEEETrans. on Inform. Theory*, vol. 51, no. 4, pp. 1432-1436, April 2005.

[2] IEEE 802.16e-2005, *IEEE standard for local and metropolitan area networks-Part 16: Air interface for fixed and mobile broadband wireless access systems-Amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands*, Feb. 2006.

[3] L. Zhang, B. Li, T. Yuan, X. Zhang and D. Yang, "Golden code with low complexity sphere decoder," *IEEE PIMRC*, pp. 1-5, 2007.

[4] B. Cerato, G. Masera and E. Viterbo, "Decoding the Golden code: A VLSI design," *IEEE Trans. VLSI Systems*, vol. 17, no. 1, Jan. 2009.

[5] M. O. Sinnokrot and J. R. Barry, "Fast maximum-likelihood decoding of Golden code," *IEEE Trans. on wireless Commun.*, vol. 9, no. 1, pp. 26-31, Jan. 2010.

[6] S. Sezginer and H. Sari, "Full-rate full-diversity 2 2 space-time codes of reduced decoder complexity," *IEEE Commun. Lett.*, vol. 11, no. 12, pp.973-975, Dec. 2007.

Implementation of Interference Cancelling Repeater based-on Software Defined Radio in Long Term Evolution

Jongmin Kim[1], Minkyu Sung[2], and Jichai Jeong[3], Senior Member, IEEE
[1]Digital Signal Processing Team, ADRF Korea
[2]Department of computer and radio communications engineering, Korea University
[3]Department of Brain and Cognitive Engineering, Korea University

***Abstract*-- We implement an interference cancellation system (ICS) relay used in software defined radio (SDR) based long term evolution (LTE) technology. The paper aims to confirm essential conditions of an adaptive filter using the conventional least mean square (LMS) algorithm for the applications of LTE service, the next generation wireless communication, based on orthogonal frequency division multiplexing (OFDM) algorithm.**

I. INTRODUCTION

In mobile communications, relays have been advanced to process the feedback signal for removing the oscillation especially when same frequency signal is used in transmit antenna. Feedback cancellation technology particularly cancels out the interference by creating feedback signals through a real-time channel modeling on an external environment using adaptive filters, and then realized to process in real-time using a digital signal processing technology. We proposed a feedback cancellation method for the orthogonal frequency division multiplexing (OFDM) based LTE technology using the least mean square (LMS) algorithm which is already applied in wideband code division multiple access (WCDMA). Furthermore, by performing laboratory test, the interference cancellation system (ICS) [1] performance for OFDM based LTE technology is investigated and verified.

II. INTERFERENCE CANCELLATION REPEATER

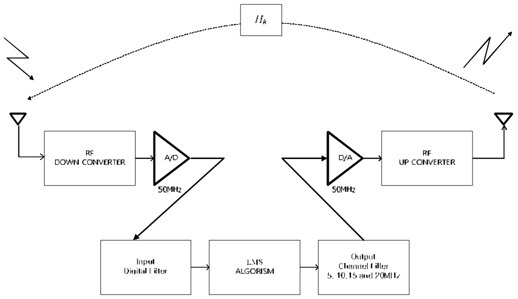


Fig. 1. Block diagram of ICS digital relay

Fig. 1 illustrates the block diagram of an interference cancellation digital relay. The received radio frequency signals are down converted to intermediate frequency (IF) band through mixer. After that, the ADC converts that signals to digital signals. To remove feedback signals which have the same frequency compared with received signal, the interference cancelling is performed by using adaptive filtering based on the LMS algorithm.

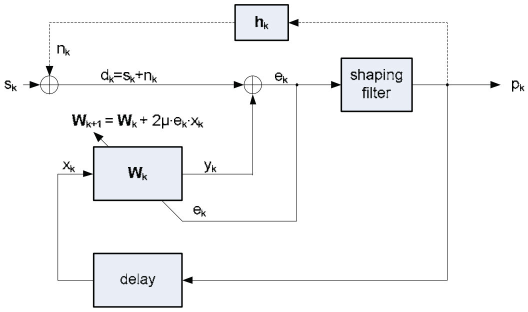


Fig. 2. Basic Structure of LMS

Fig. 2 shows the block diagram of the adaptive filter based on LMS algorithm. The feedback signal n_k is created by radiated output signal s_k at the transmit antenna through the feedback path h_k. This feedback signal is combined with input signal d_k. Due to the feedback signal, the repeater will go into oscillation if interference cancelling process is not performed.

To remove feedback signal n_k, the estimated feedback signal y_k is obtained as (1) by using adaptive filter weight W_k and delayed output signal x_k.

$$y_k = W_k^T \cdot x_k \tag{1}$$

In addition, by subtracting the input signal d_k and estimated feedback signal y_k, the estimated error e_k is obtained as

$$e_k = d_k \cdot y_k \tag{2}$$

In the adaptive process, the tap-weight vector is used to seek the minimum of the error performance surface using gradient method where the minimum point can be found theoretically by obtaining an expression for the gradient and setting it equal to zero. With this simple estimation of the gradient, the LMS algorithm for updating the steepest descent weight is expressed by

$$W_{k+1} = W_k + 2\mu \cdot e_k \cdot x_k \tag{3}$$

The bound on μ for convergence of the weight vector mean is derived as

$$0 < \mu < \frac{1}{\lambda_{\max}} \tag{4}$$

This expression is the standard LMS algorithm [2].

This work was supported in part by Brain Korea 21 project in 2012, World Class University program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea (R31-10008), and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0023063).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

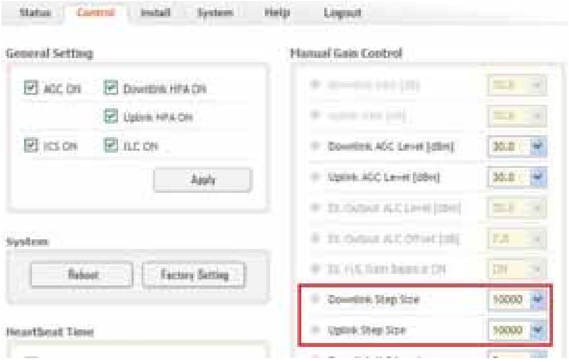
III. LABORATORY

A specification of a SDR-30-700-ICS repeater which is designed to be used in DSP and SDR modulations is summarized in Table I.

TABLE I.
SPECIFICATION OF SDR-30-700-ICS REPEATER

Electrical specifications	SDR-30-700-ICS
Downlink frequencies	Lower A: 728 – 734 MHz Lower B: 734 – 740 MHz Upper C: 746 – 757 MHz
Uplink frequencies	Lower A: 698 – 704 MHz Lower B: 704 – 710 MHz Upper C: 776 – 787 MHz
Filtering	Lower A, Lower B, and Upper C
Gain	90dB of max gain
Output power	30dBm max output power

SDR-30-700-ICS repeater is a kind of modular type relays with an input power of -60dBm, an output power of 30dBm and a gain of 90dB. As shown in Fig. 3, the performance of ICS was examined from the remote site through Web GUI. The output and μ-value in (3) were adjusted to the "step size" item.



General Setting

☑ AGC ON ☑ Downlink HPA ON
☑ Uplink HPA ON
☑ ICS ON ☑ ILC ON

Apply

System

Reboot Factory Setting

HeartBeat Time

Manual Gain Control

Downlink AGC Level [dBm]	30.0
Uplink AGC Level [dBm]	30.0
DL Output ALC Level [dBm]	30.0
DL Output ALC Offset [dB]	7.0
DL I/Q Gain Balance ON	ON
Downlink Step Size	10000
Uplink Step Size	10000

Fig. 3.SDR-30-700-ICS Web-GUI

Using SDR-30-700-ICS repeater, the performance of ICS was evaluated in DSP system according to the step size μ by Web-GUI. The performance of ICS is shown in Table II.

TABLE II.
ICS PERFORMANCE OF SDR-30-700-ICS

In Band IMD	> 50dB
System Delay	< 6.5 usec
EVM Normal	< 6% rms
Cancellation Windows Size	> 3usec
Doppler	Isolation = G-0dB, 10Hz
Direct Feedback	Isolation = G-10dB
EVM Max.	16-QAM < 12.5% rms

The laboratory test environment examining the function of ICS is shown in Fig. 4. Input signal was inserted using a signal generator and loop back path was assembled using output of ICS repeater. In the study, the μ–value in (3) was verified according to WCDMA and OFDM (16-QAM) based LTE systems. The performance of ICS is evaluated by error vector magnitude (EVM). In case of WCDMA and LTE, EVM followed a standard specified in the 3GPP (3rd Generation Partnership Project), which requires below the EVM of 12.5%, respectively [3]. EVM value is defined as the ratio of error vector e_k of root-mean-square (RMS) value to original signal R_k of RMS value n WCDMA and LTE signals

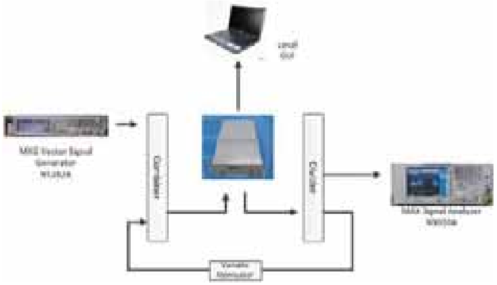

Local GUI
MXG Vector Signal Generator N5182A
Combiner
Divider
MXA Signal Analyzer N9020A
Variable Attenuator

Fig. 4. Laboratory Test Environment

Table III shows the performance capacity of WCDMA and LTE of ICS operation when feedback signal is present, and the set value of quantitative assessment.

TABLE III.
CHARACTERISTICS OF μ AND EVM VALUES ACCORDING TO THE SIGNALS

Gain 90	WCDMA(Downlink)		LTE(Downlink) / 16-QAM	
Isolation	μ	EVM	μ	EVM
G+15	1600	8.0%	12000	5.9%
G+10		8.0%		5.9%
G+5		8.0%		5.9%
G-0		8.0%		6.0%
G-5		8.1%		6.2%
G-10		8.4%		6.8%
G-15		9.2%		9.6%

In order to make the ICS performance identical, the μ-value varies in WCDMA and LTE depending on the various signals. Therefore, each μ-value must be defined appropriately depending on a specific signal.

IV. CONCLUSIONS

We implemented a ICS relay applicable to SDR-based LTE technology. The study is performed utilizing LMS algorithm over OFDM based LTE signals by using SDR-30-700-ICS. Moreover, we confirmed that an efficient communication service can be carried out by applying the proposed method of ICS's performance. The test results showed that the LMS algorithm can significantly improve the performance of ICS operated in a stable manner especially in case where oscillation occurs due to the feedback signal.

REFERENCE

[1] M. Lee, B. Keum, Y. S. Shim, and H. S. Lee, "An Interference Cancellation Scheme for Mobile Communication Radio Repeaters", *IEICE Trans. Commun.*, vol. E92-B, no. 05, pp.1778-1785, May. 2009.

[2] Y. K. Won, R.-H. Park, J. H. Park and B.-U.Lee, "Variable LMS algorithms using the time constant concept", *IEEE Transactions on Consumer Electronics*, Vol. 40, Issue 3, pp.655-661, Aug. 1994.

[3] 3GPP.TS25.106 V5.8.0, *UART repeater radio transmission and reception*, ETSI 2004

Forward Link ACM for Satellite Communication Public Test-bed via COMS

Joon-Gyu Ryu*, Sung-Yong Hong**, and Deock-Gil Oh*
*Satellite Broadcasting & Telecommunication Convergence Team, **ChungNam National University, The Korea

Abstract — **This paper present the detailed design and the algorithm for forward link ACM(Adaptive Coding & Modulation) system to improve the link availability and system throughput in satellite communication services. To implement ACM system, the channel prediction and MODCOD decision methods is simulated. The channel prediction result shows that the 99% of predicted values in LMS(Least Mean Square) algorithm is within 3dB. Also this paper proposes the appropriate system architecture for ACM transmission**.

I. Introduction

The demand for high-speed satellite communication is rapidly growing worldwide and future broadband satellite systems are expected to operate at higher bands, e.g. Ka-band(20~30GHz) because the lack of Ku-band resources. According to this trend, the DVB study the transmission scheme for high symbol-rate satellite services something like time-slicing.

Although this will provide additional capacity for new multimedia services, the propagation impairments can be greater, making it difficult to select a single physical/link design that is suited to all traffic and propagation conditions. Adaptive designs are, therefore, attractive, especially ones that can exploit the flexibility.

An Adaptive Coding/Modulation (ACM) scheme, included in the second generation DVB satellite DVB-S2 standard [1,2], implements a set of Modulation/Coding(MODCOD) waveforms at the physical layer. The use of DVB-S2 ACM is important for the forward link in two-way satellite systems using DVB-RCS to improve the system throughput[3,4].

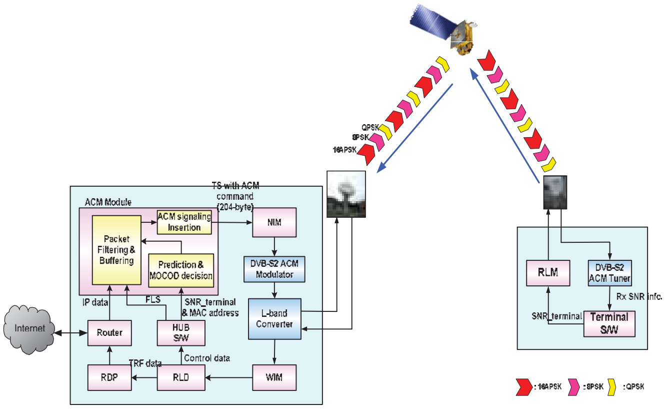


Fig. 1. The link budget result using COMS

II. Link BudGET

In satellite communication, the RTT(Round trip Time) and system processing delay of satellite link takes over 500msec. The frequent transition of MODCOD can degrade the BER performance. Also, the satellite communication network covers the wide area and each terminal experience the different channel condition. In order to apply most of MODCOD in ACM system, the Hub has to prepare large MODCOD buffers. For this reason, the appropriate number of MOCOD should be selected for reducing the transition frequency of MODCOD and system load.

In this paper the appropriate MODCOD is chosen by link budget. Table I shows the system parameters for link budget. When link budget is calculated, the first multi-beam satellite in Korea, COMS(Communication, Ocean and Meteorological Satellite), is targeted.

TABLE I
SYSTEM PARAMETERS

Forward link	Return link	Unit
DVB-S2 ACM	DVB-RCS	
QPSK/8PSK/16APSK	QPSK	
27	0.5/1/2/4	Msps
Hub Ant. Size 7.2	Terminal Ant. size1.2	m
HPA 175	HPA 10	Watt

The lowest MODCOD of forward link in this paper is QPSK LDPC 1/2 because the maximum link availability of return link is 99.93% at QPSK Turbo 1/2. The selected MODCOD is as follows, 16APSK LDPC 3/4, 8PSK LDPC 3/4, QPSK LDPC 3/4, QPSK LDPC 1/2.

III. Satellite forward link ACM system

In order to implement the ACM scheme, the VSAT system should include basically the process as follows.

- The terminal sends the SNR information to HUB every second.
- This information is delivered to ACM module with MAC address in HUB.

Figure 2 shows the block diagram of ACM module which consists of Database block, Packet filtering block, Channel Prediction & MODCOD decision block, Buffer block, Packet

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

processing block and ACM signaling insertion block. The DB block store and update the terminal information. e.g. MAC address, SNR, MODCOD. The Channel Prediction & MODCOD decision block predict the future SNR value about three seconds before and decide the appropriate MODCOD. Based on this decision, the Packet filtering block classifies the receiving IP data. This classified IP data is buffered and transform to TS data at Packet processing block. The 188-byte MPEG data is changed 204-byte MPEG data to fit the commercial ACM modulator's interface.

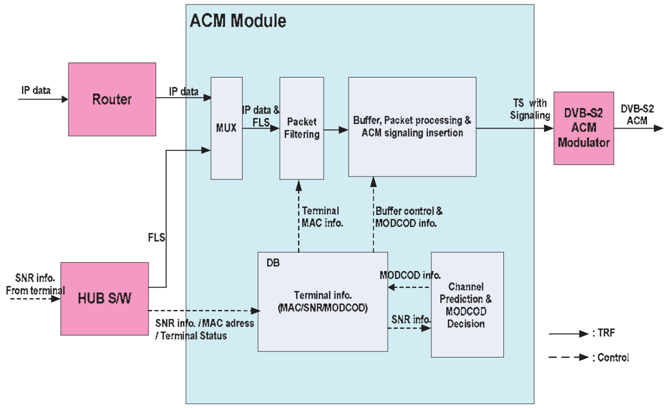


Fig. 2. The configuration of ACM module

A. SNR Prediction algorithm

In order to implement adaptive rain fade compensation according to the channel conditions, it is required to predict the channel quality accurately in advance with consideration of the round-trip delay the amount of the signal quality, e.g. SNR value. The SNR variation in a satellite link includes variation of rain attenuation and comparatively fast scintillation. Because, generally, the SNR variation due to scintillation is much faster than the response speed of an adaptive system, the prediction scheme needs to filter out this fast variation. The efficient prediction scheme consists of four functions including discrete-time low-pass filtering(LPF), rain-fade prediction, mean-error correction and prediction margin allocation.

One prediction algorithm uses two constant weight values for two end points of the observation period and assumes that future variation of the signal level will remain the same as the previous variation. This method is defined as slope based prediction(SBP). Another prediction algorithm employ the variable weights using adaptive filtering prediction(AFP) such as the LMS[5]. In this case, the weights are updated at every second.

B. MODCOD Decision algorithm

The chosen MODCOD is the one with the highest threshold that is lower than the observed SNR(including the fade margin). A hysteresis loop is normally used to avoid oscillations in choice of the MODCOD, which would result when the observed SNIR was near a threshold.

C. Simulation Results

To simulate the SNR prediction & MODCOD decision, the real rain attenuation values which is measured by Ka-band rain attenuation measurement system is used.

Figure 3 shows how the MODCOD tracks the changing channel conditions. It is safe to choose a lower ModCod than required, but this reduces the spectral efficiency, consuming more satellite capacity for the same data. A higher MODCOD than required has a threshold close to the estimated SNIR (small fade margin). This conserves satellite capacity, but reduces the probability of successful packet transmission.

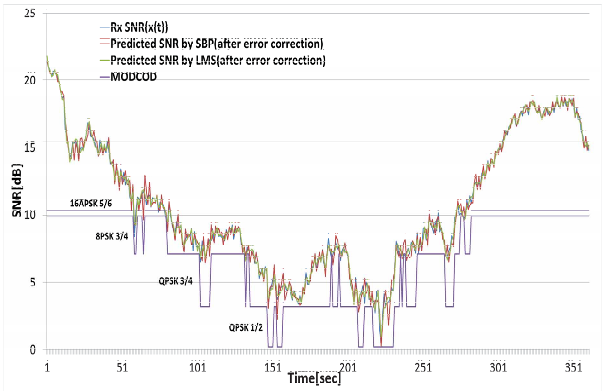


Fig. 3. The simulation results of ACM module

IV. CONCLUSION

This paper proposes an efficient adaptive rain attenuation compensation scheme for satellite communication systems. The appropriate configuration of ACM module and MODCOD is proposed. The 4 step of MODCOD is decided by link budget.

The proposed compensation algorithms will play an important role in providing satellite communication services which are both economical and of high quality.

REFERENCES

[1] Alberto Morello, Vittoria Mignone,"DVB-S2:The Second Generation Standard for Satellite Broad-band Services", PROCEEDINGS OF THE IEEE, Vol. 94, No. 1, January 2006

[2] ETSI EN 302 307, v1.1.2, "Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications" European Telecommunications Standards Institute (ETSI), June 2006.

[3] Georgios Gardikis, Nikolaos Zotos, and Anastasios Kourtis, "Satellite Media Broadcasting with Adaptive Coding and Modulation" , International Journal of Digital Multimedia Broadcasting, Volume 2009

[4] Hermann Bischl, Hartmut Brandt, Tomaso de Cola, et al, "Adaptive coding and modulation for satellite broadband and networks: From theory to practice", Int. J. Commun. Syst. Network, vol 28, March 2009

[5] Sooyoung Kim Shin, Kwangjae Lim, Kwonhue Choi, and Kunseok Kang, "Rain Attenuation and Doppler Shift Compensation for Satellite Communications", ETRI Journal, Volume 24, Number 1, February 2002

A Seamless Channel Handover Method for Service Continuity in Super Wi-Fi

Myeongyu Kim, Youchan Jeon, Sangwon Park and Jinwoo Park
School of Electrical Engineering, Korea University, Seoul, Korea

***Abstract*--A key technical challenge in the Super Wi-Fi applications is how to provide a seamless Internet service even when a Super Wi-Fi user should give up the channel in use due to an active incumbent user. In this paper, we propose a channel handover method to support service continuity of Super Wi-Fi, in which AP selects an available channel and provides the channel information to MSs. Performance evaluation shows that the proposed scheme is superior to the conventional Wi-Fi in channel handover delay.**

I. INTRODUCTION

Super Wi-Fi is a newly emerging wireless Internet technology, which constitutes Wi-Fi networks using TV White Spaces. TV White Spaces are the unused portions of the TV spectrum. FCC has approved operation of unlicensed radio transmitters in the broadcast television spectrum [1]. So far industry and standardization bodies have shown a particular interest in using the TV White Spaces for providing broadband services through extended Wi-Fi like connectivity known as Super Wi-Fi. The IEEE 802.11af working group has been set up to define a standard for implementing Wi-Fi-like networks in TV bands [2].

Super Wi-Fi supports the usage of an available channel in TV White Spaces and uses channel numbers specified by regulatory bodies. TV services operate on 50 channels in the VHF and UHF portions of the radio spectrum. As shown in Figure 1, an AP has the authority to control the operation of MSs after obtaining an available TV channel for use at its own location. One of available TV channels is chosen to be the AP's channel. There are incumbent users in TV White Spaces such as TV stations, microphones, etc. Microphones are used ranging from small-scale lecture rooms to large-scale music and sporting events. The TV White Spaces suffer from temporal variation due to the widespread use of wireless microphones.

In Super Wi-Fi, key challenge is dealing with the sudden appearance of an incumbent user on a channel such as a wireless microphone. If the incumbent user appears on the channel occupied by an AP, both the AP and MSs should vacate the channel and move to a new available channel immediately to minimize interference. In conventional Wi-Fi [3], the MSs should find the new channel using scanning. However, the scanning method gives rise to obstacle in Super Wi-Fi services due to the significant scanning delay. Super Wi-Fi should guarantee service continuity when a Super Wi-Fi user gives up the channel in use due to the active incumbent user.

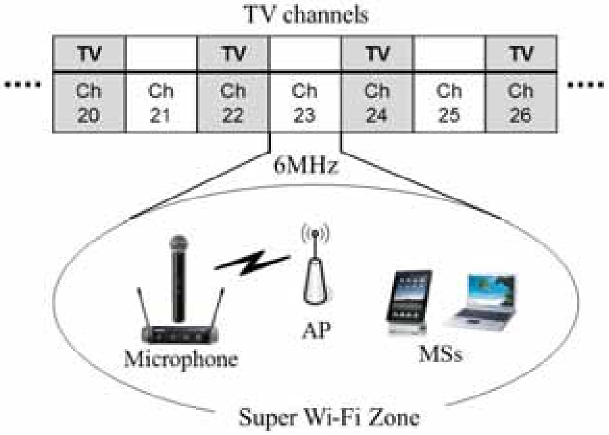


Fig. 1. Illustration of Super Wi-Fi architecture.

In this paper, we propose a channel handover method to support service continuity in super Wi-Fi. In the proposed scheme, AP selects a new channel and provides the channel information for handover to MSs considering incumbent users. Therefore the proposed scheme provides the method not only to support service continuity but also to minimize interference to incumbent users. We verified the superior performance of the proposed scheme in channel handover delay.

II. PROPOSED SCHEME

We propose a channel handover method to support service continuity for Super Wi-Fi considering incumbent users. Both AP and MS transceiver use Noncontiguous OFDM (NC-OFDM) [4]. AP and MS transceivers can avoid interference to incumbent users by inactive subcarriers within their vicinity. Figure 2 shows an example of proposed channelization when a microphone appears. The bandwidth of wireless microphone signals is 200 kHz, much smaller than that of a TV channel (6 MHz). The frequency band of the wireless microphone is set to inactive subcarriers and the other is set to active subcarriers. Through these active subcarriers, it is possible to decode data between AP and MSs.

The procedure for handover method in the proposed scheme is shown in Figure 3. When an incumbent user appears on the channel, the AP transceiver and the MS

This research was partly supported by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) [KCA-2012-08-911-05-001] and partly by the BLS project from the Seoul Metropolitan City [WR080951, Seoul R&BD Program].

978-1-4673-1363-6/13/$31.00 ©2013 Crown

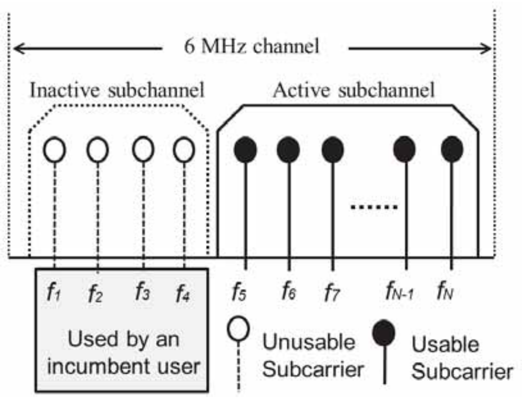


Fig. 2. Example of proposed channelization.

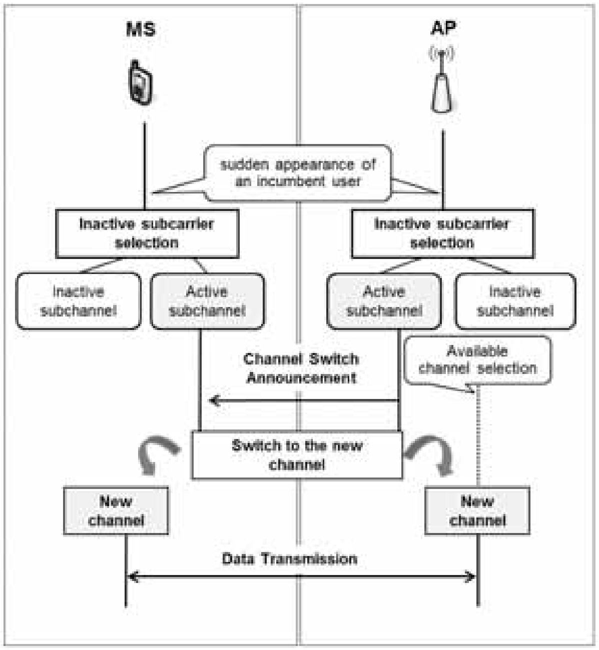

MS
AP
sudden appearance of an incumbent user
Inactive subcarrier selection
Inactive subcarrier selection
Inactive subchannel
Active subchannel
Active subchannel
Inactive subchannel
Channel Switch Announcement
Available channel selection
Switch to the new channel
New channel
New channel
Data Transmission

Fig 3. Procedure of proposed scheme.

transceiver are able to set up inactive subcarriers in frequency band of the incumbent user to have no interference to the incumbent user. The AP selects an available channel as a new channel not used by neighboring APs and incumbent users. Then, the AP broadcasts a channel switch announcement to inform the new channel using active subchannel. After the MSs receive the channel switch announcement, the AP and the MSs move to the new available channel immediately.

To summarize, the AP selects an available channel and provides the channel information to MSs having no interference to incumbent user. Therefore the AP and the MSs carry out seamless handover for Super Wi-Fi service.

III. PERFORMANCE ANALYSIS

To evaluate the performance of the proposed scheme, it is assumed that all the MSs and APs always have data frames to transmit like Bianchi's model [5], i.e., the traffic load is saturated. It is considered that the number of available TV channels is 20. In addition, it is assumed that the channels used by neighboring APs are randomly distributed and considered as unavailable channels for handover. The channel handover delay time in the proposed scheme can be included periods that AP selects a new channel due to incumbent users and MSs discover the channel and MSs move to that channel. The signaling cost can be defined as the amount of signal messages required when the AP scans channels for handover and MSs discover the new channel.

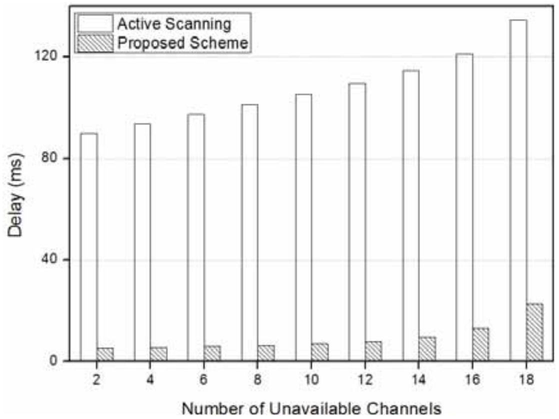


Fig. 4. Comparison of channel handover delay.

Figure 4 shows the channel handover delay time over the number of unavailable channels. The channel handover delay time of the active scanning increases as the number of unavailable channels increases because of unnecessary probe responses from neighboring APs. In the proposed scheme, the channel handover delay time maintains at less than 25ms. For active scanning, the signaling cost increases greatly as the number of MSs increases. However, in the proposed scheme, signaling cost remains constant according to the number of MSs.

IV. CONCLUSION

We propose a channel handover method to support service continuity of Super Wi-Fi, in which AP selects an available channel and provides the channel information to MSs without interference to the incumbent users. Therefore the AP and the MSs carry out seamless handover for Super Wi-Fi service. We verified the superior performance of the proposed scheme in channel handover delay.

REFERENCE

[1] FCC, ET Docket No. 08-260, "Second Report and Order and Memorandum Opinion and Order," Nov. 2008.

[2] IEEE P802.11af™/D1.04, "Amendment 4: TV White Spaces Operation," Oct. 2011.

[3] IEEE 802.11, "Part 11: Wireless LAN medium access control (MAC) and the physical layer (PHY) specifications," *IEEE Standard 802.11*, June 2007.

[4] R. Shamir, "An efficient implementation of NC-OFDM transceivers for cognitive radios," *in Proc. IEEE CROWNCOM*, June 2006.

[5] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE J. Sel. Areas Commun.*, vol. 18. no. 3, Mar. 2000.

A Mobile Agent Framework for Ubiquitous Media Access

Craig M. Gelowitz, *Member*, *IEEE*, Luigi Benedicenti, *Member*, *IEEE* and Raman Paranjape, *Member*, *IEEE*

University of Regina, Saskatchewan, Canada

Abstract - **The trend in increased storage capacity of personal media devices has paralleled the increase in network capable and digital media capable devices such as televisions, computers, cell phones and other personal media devices. This implies that media will continue to be stored across a variety of personal network enabled devices despite available cloud-based media storage and sharing solutions.**

There has been a demonstrated desire to establish a convergent and ubiquitous media access experience among distributed media storage and devices. This is supported by the numerous attempts and development of media access applications, systems and techniques by industry and academia.

The framework design in this paper demonstrates that the identified issues and limitations of existing media access solutions can be overcome through the utilization of a software agent media framework design.

I. INTRODUCTION

The quantity of personal media stored in digital form has followed an increasing trend over the past several years. This is mainly due to the vast number of devices available for creation and storage of media as well as the decreasing cost of those devices.

In order to share media content between devices, one of the current trends is to use some kind of cloud-based storage service such as Flickr, YouTube, Dropbox and Facebook which have become increasingly popular for sharing personal media online. However, those services require the user to upload their media content in advance of making the content available to other devices. In addition, uploading large amounts of data by utilizing the constrained upstream bandwidth of a typical Internet service connection can be a time intensive process. There are also numerous privacy and security issues associated with these types of services that users must understand and contend with.

A variety of conceptual frameworks, applications and protocols have been developed that attempt to provide a more ubiquitous media access experience for the end-user between devices. However, available industry solutions are often enabled through proprietary software or hardware that requires their own specific software or hardware combinations. Other less-proprietary solutions that have emerged such as UPnP [1] and DLNA [2] neglect the ability to share media between devices outside of a local area network.

All of the current paradigms provide only a partial solution for media access and sharing between devices. Ideally, users can access and consume any personal media data regardless of the location of the device where the data is stored or the location of the preferred destination device. In this ideal situation, there is no need to upload media to cloud-based storage or be under any constraint such as a local network or device compatibility.

II. AGENT TECHNOLOGY

Software agent technology has its roots in early research on distributed computing and artificial intelligence. Over the years, the agency concept has matured and become well established in scientific literature. Agents and agent systems have been developed to perform a wide variety of tasks such as monitoring, controlling resources, modeling, simulation, data mining, autonomous decision making and acting as proxies on behalf of end users [3, 4].

Logical separation of agent responsibilities and the ability of agents to communicate provide advantages to distributed systems development [5]. This research focuses on agency as a mechanism to enable a ubiquitous media service framework. Each agent in the framework contributes to the goal of overcoming real-time media access and delivery issues between devices through agent mobility, communication and collaboration as opposed to attempting to demonstrate agent intelligence in the traditional sense.

Mobile software agents, through their characteristic abilities, enable the proposed solution and form the middleware of the conceptual framework. Their autonomy provides the mechanism for dynamic adaptation and decision making that satisfies a wide range of possible scenarios and potential outcomes. Their migratory ability allows them to travel to-and-from networked devices for gathering situational and contextual information. Their communicative abilities provide the information architecture necessary to provide real-time media services.

III. AGENT FRAMEWORK COMPONENTS

There are several issues that are addressed by the proposed framework including device accessibility, network properties, bandwidth constraints, device attributes, and

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

media context. The framework design consists of five fundamental components:

1) Public Agent Repository: the origin of the agents in the framework

2) Gate Agents: agents responsible for registered devices

3) Bridge Agents: generic agent communication relays

4) Service Agents: agents responsible for handling media requests

5) Public Resources: publically addressable resources and the user interface

It is through these five core components a generalized framework solution is established. It is these components that are combined to provide accessibility, flexibility, media adaptation, contextual awareness and ubiquity to the proposed framework.

A. Public Agent Repository

Mobile agents in the framework originate at the agent repository and are provided access to the public network for migration. The repository enables mobile agent migration paths out from the agent repository to devices that participate in the framework.

Figure 1 is representative of common Internet networks where local network devices utilize communication paths to the public network through NAT (network address translation) gateways. The agent framework addresses multiple networks and LAN devices by providing a migration path from the agent repository. The path enables access to internal LAN devices through a series of multiple hops as illustrated on the right of Figure 1.

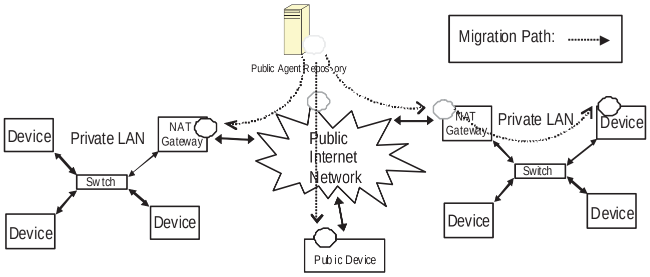


Figure 1: Public Agent Repository

B. Gate Agents

Gate agents, as their name implies, are the gate keepers of information and network access for participating devices in the framework. The gate agents are responsible to provide decentralized information for participating remote devices. This provides scalability and dynamic availability to the framework.

Gate agents exist on the public facing device for each local network and are tasked with being aware of their local executing environment and their local network context. Gate agents discover local device information and local network properties dynamically. When applicable, gate agents utilize broadcast communication within a local network to discover participating devices dynamically. The inclusion of gate agents decentralizes the information with respect to available participating devices. This enables the framework to grow without unnecessarily centralizing information about available devices and media. The gate agents and their information hierarchy mimic the DNS hierarchy where the responsibility for device information is decentralized into sub-levels of local network responsibility.

C. Bridge Agents

Bridge agents are generic modular agent instances that provide an abstracted communication mechanism for effectively bridging media service communication among networks and devices. The bridge agents form a middleware communication layer for media service agent communication. This middleware communication layer is an agent message relay mechanism that enables communication between multiple networks and end-devices.

As illustrated in Figure 2, bridge agents are initiated and migrate to selected networks and devices as a result of media service requests by users. In the figure's example, the bridge agents illustrated are the relays that establish the data path between the two service agents at the end-devices.

When a media service is initiated, bridge agents migrate from the public agent repository and provide peered communication channels for media service agents. This provides scalability to the framework. The dynamic peering between media service agents allows simultaneous media sessions to increase among individually managed networks without affecting the fundamental operation of the framework. Like in other peering systems, simultaneous sessions are effectively distributed among the participating devices and networks.

In addition to enabling these peering relationships, the bridge agents act as generic middleware communication instances that abstract communication issues away from the service agents. All of the bridge agents within the framework are generic agent instances that perform identical functions. The only difference between individual instances of bridge agents are the physical devices where they execute and the agents they are responsible for with respect to message forwarding.

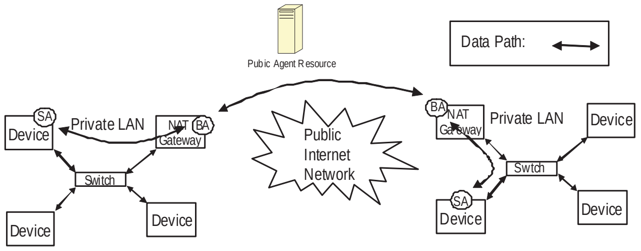


Figure 2: Bridge and Service Agents

D. Service Agents

Service agents are autonomous agents in the framework that are responsible for providing media services among the framework's participating devices. The service agent's primary goal are to determine the overall situational context, make decisions to account for imposed constraints and provide real-time media services.

Service agents are initiated and executed dynamically as a result of media service requests by users in the same manner as the bridge agents. Service agents respond to media service requests by migrating out from the public agent repository to the selected devices. The service agents are registered with their respective bridge agents to establish peered communication as indicated in the previous section.

In order to provide real-time media services between participating devices in the framework, the available instantaneous bandwidth is evaluated by service agents. Service agents are given this responsibility because the communication path maybe limited by network constraints at any network node between them.

In addition to determining the approximate one-way bandwidth constraint for media transfer, the service agents are responsible for understanding the situational context of their surroundings. The conditions evaluated by the service agents include device properties, media availability and media properties. Service agents examine their respective device properties including processing power, screen resolution, device conditions and device capabilities. This inspection determines the varying device capabilities the service agents must account for as part of their decision process. This provides flexibility to the framework because the service agents are responsible to adapt to varying device constraints.

Lastly, with respect to context, service agents are responsible for determining media availability and media properties dynamically through their execution on local devices. This is accomplished by the inspection of the local device's media file system for available media and media properties such as media type, resolution, bit rate, format, codec, file name and file size.

The information gathered with respect to media and device context is shared among the service agents to assist in the collaborative decisions required to provide successful media services in the presence of imposed constraints. This collaborative understanding of device and media context among the service agents enables the decisions and the subsequent actions by the service agents to provide real-time media services.

The service agents, through information sharing and collaboration, enable the appropriate decisions based on the emergent properties of their discovered situational context. Agent decisions that enable real-time media services can range from no adaptation of media data (because the situational constraints do not warrant it) to complete transformation of media data through transcoding techniques that account for imposed constraints. The fundamental purpose of the service agents in the framework is to share emergent contextual information about the devices involved, the network properties and the available media. In this way, the agents make the appropriate decisions and act on the information to successfully provide real-time media services. Information collaboration is utilized by the service agents to enable decisions that account for the constraints imposed by the dynamic situational contexts they encounter.

E. Public Resources

The framework design includes the addition of public resources as client/server mechanisms for the framework's agents and its users. The framework's agents, as clients, utilize public resources to respond to user interactions and provide mechanisms for requested media services.

The user interface into the framework is established through a publically accessible web server. The service agents respond to user interaction through the web server resource. Service agents also communicate with publically available resources to provide mechanisms such as remote procedure calls, downloadable executables, system information, web services or storage that may assist in providing real-time media services.

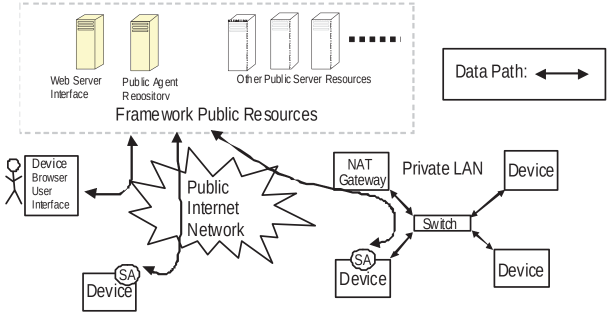


Figure 3: Public Resources

The framework's publically accessible resources include, at a minimum, a public agent repository and a publically accessible web server for providing the user interface into the framework. The framework design defines a publically accessible web enabled interface to enhance the flexibility and accessibility for its users. Users are intended to utilize any web capable device for user interaction with the framework.

The "other public server resources" in the figure are not defined in the framework design to provide flexibility to individual framework implementations. Framework implementations may offer a variety of client/server resources to aid in providing media services. Publically available resources provide service agents, in the absence of local functionality, media service functionality alternatives. When media service requests are evaluated by service agents, the agent decisions are made after considering all of the available mechanisms provided by the framework implementation.

IV. RESULTS

A prototype was implemented to realize, validate and examine the design of the conceptual framework. Empirical observations of the framework design's methodologies have been gathered to validate the framework design. This has included bandwidth estimation, transcoding, agent messaging and the subsequent effects of these methods on accessibility, real-time constraints, response time and usability.

As an example, the agent messaging in the prototype achieves real-time delivery of media by dividing the media to be delivered into defined segment sizes and transfers those segments as agent messages. As the segments are received by the destination agent, the destination agent displays the media to the user. Because there is an inherent amount of overhead in agent messaging, agent message passing can reduce the throughput achievable.

To compensate for this, agent message size can be adjusted by the agents so that the overhead of agent messaging is reduced. The following figure illustrates that message size adjustment modifies the achievable throughput. The throughput achievable by agent message passing converges with the achievable transcoding rate if the agent message size is increased.

The graph shows that an increase in agent message size results in the convergence of the rate at which the media can be transcoded. The message passing transfer rate converges but does not attain the actual transcoding rate. This is due to the inherent delay of processing the media segment for delivery. The graph is intended to show that if the transcoding parameters are adjusted to meet real-time requirements, agent message size can also be adjusted to maintain or exceed those real-time requirements. There is however, a trade-off with the reaction time of the prototype.

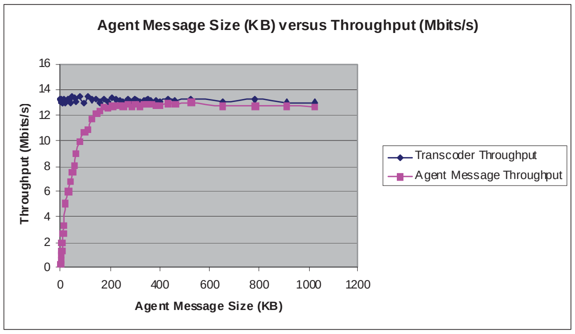


Figure 4: Agent Messages vs. Throughput

V. CONCLUSIONS

The conceptual media framework was designed to be a multi-agent system architecture that enhances media access capabilities across individually managed networks and devices. The framework outlines the agent-assisted methodologies to enable access and transfer of media between participating devices in real-time.

The agents migrate to devices to provide the media access functionality to remote devices on-demand. This methodology enables the framework to utilize generic and reusable modular agent instances and reduces the number of agent types and the complexity of agent interaction. It accomplishes this by enabling framework functionality through generic agent responsibilities that adapt to individual device characteristics. Individual devices do not depend on specifically compiled and compatible agents. As such, the functionality of a framework implementation can also be changed or upgraded on-demand because any newly compiled agent instances will migrate out to selected devices from the agent repository upon initiation of a media service.

REFERENCES

[1] M. Jeronimo, J. Weast, *UPnP Design by Example*, Intel Press, 2003

[2] E. A. Heredia, *An Introduction to the DLNA Architecture: Network Technologies for Media Devices*, Wiley, June 2011

[3] Q. H. Mahmoud, L. Yu, "Making Software Agents User-Friendly", IEEE Computer, pp. 94-96, July 2006.

[4] D. B. Lange, M. Oshima, "7 good reasons for Mobile Agents", Communications of the ACM, Volume 42, No. 3, March 1999.

[5] S. Cranefield, M.K. Purvis, "An Agent-Based Architecture for Software Tool Coordination", in Proc. PRICAI Workshop on Intelligent Agent Systems, pp.44-58, 1996.

The Single Image Dehazing based on Efficient Transmission Estimation

Soowoong Jeong and Sangkeun Lee, *Member, IEEE*
The Graduate School of Advanced Imaging Science, Multimedia, and Film,
Chung-Ang University, Seoul, Korea

***Abstract*--In this paper, we propose a novel method for estimating the transmission and removing the haze from single image. This method is based on observation that the property of haze is widely spread. Thus its estimated transmission should be smoothly changed over the scene. We employed the local entropy and log function for estimating the atmospheric light and the smooth transmission. Experimental results demonstrated that the proposed method can be applied efficiently to outdoor vision applications or devices including cameras and camcorders with low complexity.**

I. Introduction

The human eye can recognize an object by incoming light which is reflected from the surface of the object. The most of outdoor scenes are degraded because the incoming light is absorbed and scattered by the particle of atmosphere such as haze, fog, and smoke. Furthermore, the incoming light is even blended with airlight [1]. In these reasons, an observed image reduces the contrast and the tonal information of an original scene. The degree of its degradation is decided by the depth which is the distance between an observer and objects. Attempting to estimate the depth is still a difficult problem as researched in previous works [1-3]. The most of dehazing algorithms employ prior information or assumption. In this paper, we propose a novel method to estimate the atmospheric light by employing the dark channel combined with local entropy. High value in dark channel tends to degrade the scene more than low value; here the meaning of the low value in local entropy refers to the stability of energy. In this work, it is assumed that degradation by haze is spread over the image, and its transmission should be smoothed. For this purpose, we apply a log function to a min channel. The log function is an effective method for smooth transmission and assures fast computational time. Fig. 1 shows the result of our proposed method versus an original corrupted image. We believe that the proposed method can be a good tool for eliminating the haze in the related fields even for consumer products including camera and camcorders requiring low complexity in hardware implementation.

II. Proposed Method

In computer vision, the formula of degraded image by haze is defined as follows [1-3]:

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

Where I(x) is an observed image, J(x) is a scene radiance, and A denotes the global atmospheric light.

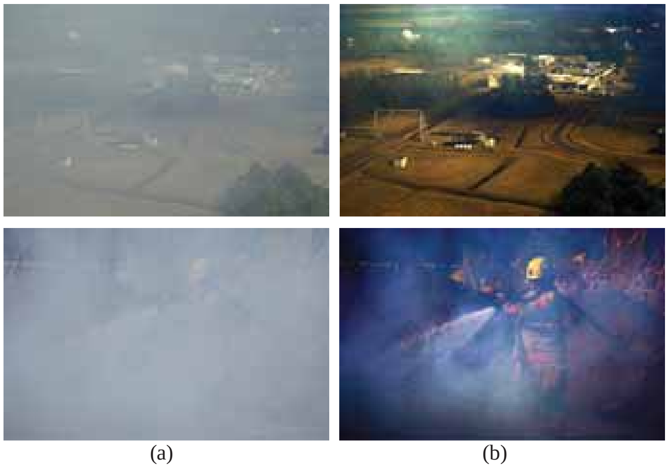


Fig. 1. Proposed method results: (a) original image, (b) result of proposed method.

A. Atmospheric Light Estimation

The most opaque region is estimated as atmospheric light A. In dark channel prior approach [1], the dark channel reflects the depth of an observed image I(x), and the most opaque region has the highest value in the channel. The dark channel is defined as follows [1]:

$$D(x) = \min_{c \in \{r,g,b\}} (\min_{y \in \Omega(x)} (I^c(y))) \tag{2}$$

where I^c is a color image, Ω is a local patch centered at x. However, the dark channel can miscalculate the opaque region by the influence of white object. Therefore, we employ the local entropy defined as follows:

$$E(x) = -\sum_{i=0}^{N} (\underset{y \in \Omega(x)}{p(y_i)} \times \log_2 (\underset{y \in \Omega(x)}{p(y_i)})) \tag{3}$$

Where p is the probability density in the local patch, N is the maximum range of an image. To reduce the computational time, we calculate the probability with non-overlapped blocks over an image. A low value in the local entropy indicates stable energy and also means that there is a high possibility of opaque region. We secure the candidates of atmospheric light with the top 0.1% of the brightest pixels in dark channel. Among the candidates, the lowest value in entropy is decided as the atmospheric light A.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Transmission Estimation

As we mentioned in Introduction Section, the transmission should be smooth enough. Therefore, we employ a log function that can estimate for the smooth transmission. The smooth transmission is defined as follows:

$$t(x) = 1 - \omega(\alpha \log(\min_{c \in \{r,g,b\}} (\frac{I^c(x)}{A^c}))) \qquad (4)$$

Where α is the correction factor calculated both with the mean of original min channel and the mean of modified min channel after applying the log function. The ω can be set according to application. In (4), the min channel reflects the coarse depth information. It is noted that the log function plays an important role of compressing the depth for smooth transmission under with the range of min channel is in [0, 1].

C. Scene Radiance Recovery

We can recover the original radiance image with the estimated atmospheric light and the estimated transmission by equation (5) that defined as follows:

$$J^c(x) = \frac{I^c(x) - A^c}{\max(t(x), t_0)} + A^c \qquad (5)$$

Where t_0 is set to 0.1, and it is used to avoid zero denominator value. If transmission is close to zero, then it may cause the noise in the recovered image.

III. Experimental Result

For the evaluation of the proposed approach, it is compared with He [1] and Tarel [2] schemes in terms of observable image quality and computational speed. It is noted that the sizes of the input image and local patches are set to 440×450 and 15×15, respectively.

The comparable results for quality comparison are shown in Fig. 2. The result of the proposed algorithm is comparable and even better for eliminating the haze. Additionally, Table I shows the computational time comparison under the implementation with the same commercial programming language for each algorithm. It is easily seen that the proposed scheme is much faster than other baseline algorithms. Specifically, our method required only 0.94 seconds for the whole processing.

TABLE I
Comparison of Computational Time

Method	Elapsed Time(s)
He's et al.	26.21
Tarel's et al.	3.67
Ours	0.94

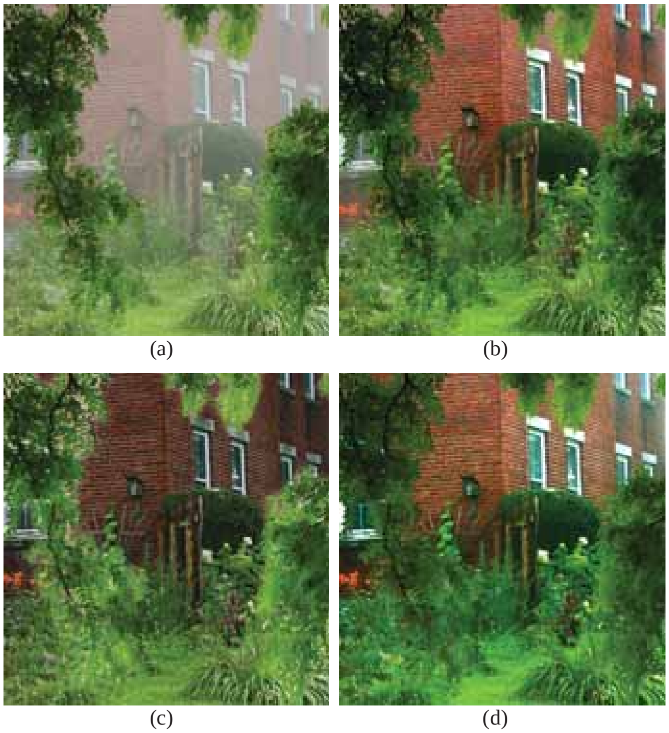


Fig. 2. Comparison with existing methods: (a) original image, (b) result of He's et al.[1], (c) result of Tarel's et al.[2], and (d) the proposed method.

IV. Conclusion

This paper proposed an efficient algorithm to effectively estimate the transmission image for single image dehazing. We modified the transmission map to reflect smooth changing property of the haze using a log function combined with entropy theory. The experimental results showed that our method outperformed the existing methods in terms of its complexity with keeping the restoration quality comparable to the baseline algorithms. Therefore, it is believed that the proposed approach can be widely used for outdoor vision applications or devices requiring low complexity.

Acknowledgment

This work was partially supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the MEST Korea government (No. 2012-0005409, 2011-0003496) and Korean Research Foundation under BK 21 project.

References

[1] K. He, J. Sun, and X. Tang, "Single Image haze Removal using Dark Channel Prior," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1956-1963, 2009.

[2] J. Tarel and N. Hautiere, "Fast Visibility Restoration from a Single Color or Gray level image," *Proc. IEEE International Conference on Computer Vision*, pp. 2201-2208, 2009.

[3] R. Fattal, "Single Image Dehazing," *ACM Transactions on Graphics*, vol. 27, pp. 1-9, 2008.

ADAPTIVE EVOLUTIONAL STRATEGY OF PARTICLE FILTER FOR REAL TIME OBJECT TRACKING

Clementine Nyirarugira and Tae Yong Kim

Graduate School of Advanced Imaging Science
Chung-Ang University, Seoul, Korea

ABSTRACT

In this paper, we propose an efficient real time tracker that uses a differential evolution strategy within the particle filter framework. Particles are strategically propagated based on the maximum a posterior (most likely) object location with genetic operators. This enables the use of a small sample size and alleviates the frequent sample degeneracy and impoverishment problems encountered in particle filters. We reduce the sample size considerable while improving the trackers accuracy. This makes the proposed tracker a good candidate for real time object tracking or an embedded resource constrained tracker.

1. INTRODUCTION

Human-computer interaction is evolving towards non-contact devices, using perceptual and multimodal user interfaces. A key component to effective interaction with devices is object tracking where the object may be a hand, a person, a head, or something that the user would like to interact with. In this paper, we develop a single object visual tracker based on a new particle filter that is inspired by a Differential Evolution (DE) algorithm. In particular, we use a DE algorithm to provide the particle filter with a better way to propagate samples by using genetic operators to model fast and slowly moving objects. This helps to control the number of particles used in tracking while improving tracking accuracy compared to classical particle filters. The new tracker is able to use a very small sample size while maintaining high accuracy in its tracking performance.

2. PARTICLE FILTER

2.1. Particle filter

A particle filter is a non-linear and non-Gaussian filtering method that is based on sequential importance sampling and involves the propagation of a set of particles in state space in order to approximate a posterior probability density function (PDF). As the number of particles increases, the particles gradually approach the state probability density function, and reach the optimal Bayesian estimate [1].

A common difficulty with particle filters is the degeneracy problem that occurs when all but a few particles have a negligible weight. A number of different approaches have been proposed to deal with degeneracy that include using a large number of particles, an adaptive resampling strategy, Particle Swarm Optimization [2], and genetic operators [3]. However, each of these has its own set of problems and limitations. In this paper, we incorporate differential evolution into particle filter framework to overcome sample impoverishment and degeneracy. This is done by propagating particles based on the most likely true state region (selection strategy). Simulation results show that it is possible to reduce the number of particles while avoiding the problem of degeneracy. As a result, visual tracking is improved and computational time is reduced.

3. AN ADAPTIVE EVOLUTIONAL PARTICLE FILTER ALGORITHM

When using a particle filter for tracking, it is important to efficiently draw the weighted set of particles (samples). A common conceptual base in DE is simulating the evolution of individuals (candidate solutions) by a processes of selection and perturbation. These processes depend on the perceived performance (fitness) of individuals, and the selection of the *best fit* individuals. Therefore, we propose a more strategic way of propagating particles using a maximum a posterior (MAP) estimator in combination with genetic operators. Unlike the approach used in [3], we propagate all of the particles at each time step without performing any resampling. The MAP helps avoid the convergence issues of DE and the resampling process in the classical PF and reduces the computational requirements.

3.1. The algorithm

The details of the algorithm are as follows. Given an estimate of the location of an object, $\widehat{x}_{t-1}$ at time $t-1$, the steps are

1. *Sampling*: Sample $\{x_{t-1}^i\}_{i=1}^N \sim p(x_0)$ centered at $\widehat{x}$.
2. *Differential Evolution*: Partition the samples into three disjoint subsets, S_1, S_2, S_3 with particles in S_1 under-

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

going mutation, S_2 undergoing arithmetic crossover, and S_3 left unchanged.

3. *Prediction*: Feed the three subsets of particles into the system transition model, $x_t^i \sim p(x_t^i | x_{t-1}^i)$.
4. *Likelihood*: Evaluate the particle fitness, $p(z_t|x_t) = p(z_t|x_t) \cdot p_{\text{subset}}$, where p_{subset} is the probability of mutation, crossover, or normal.
5. *Update*: Update the probabilities

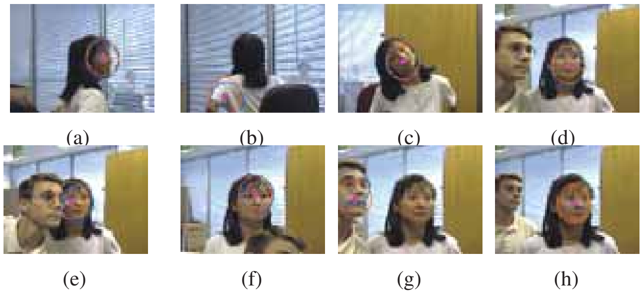
$$p(\mathbf{x}_t|\mathbf{z}_t) \approx \mathbf{w}_t^i = p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i) \cdot p(\mathbf{z}_t|\mathbf{x}_t)$$

and normalize the weights $\mathbf{w}_t^i$ so that they sum to one.

6. *Parameter Estimation*: Estimate the parameters, $\widehat{\mathbf{x}} = \arg\max_{x_t^i}(\mathbf{w}_t^i)$.

4. EXPERIMENTAL RESULTS

All experiments are done in HSV color space and the Bhattacharyya coefficient is used as a fitness measure. We first compare Sample Importance Resampling(SIR) to our algorithm. Although, a large sample size (300 particles) is used, the SIR tracker loses the object after a few iterations. However, using only 30 particles in the proposed tracker, the accuracy is much higher (results are not shown because of space constraint).


(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 1. Tracking result by the proposed tracker; the orange ellipse is the MAP state estimate and the blue one is the Mean.

Using the video sequences at http://vision.stanford.edu/birch/headtracker/seq/, Fig. 1 shows some of the challenges that are faced by the tracker. In Fig. 1(a), there is no face and the tracker makes an error by tracking the arm. The tracker, however, is able to track the correct object in the presence of similar object as seen in Fig. 1(d)-(f). Sometimes, after full occlusion of the object, a similar object may fool the tracker as seen in Fig. 1(g), but shortly thereafter the tracker recovers.

The proposed tracker overcomes the problem of sample degeneracy and is fast (see table 1). Using the proposed algorithm, we are able to reduce sample size by up to 90% while maintaining the tracking performance. In addition, we may skip 3 to 10 frames while maintaining the tracking performance, which means that the tracker is able to track fast-moving objects, and that it can be used in environments where the measurement are not computed at each time step.

Tracking Mode	Frame #	Rate Mean-MAP	Sample Size	FPS
Normal	400	90-92	600	32 Hz
			60	64 Hz
Skip(10)	40	85-90	600	32 Hz
			60	64 Hz

Table 1. Tracking Performance

5. CONCLUSIONS

We have presented a new candidate tracker for real-time object tracking. When particles are distributed strategically, tracking accuracy can be improved and the number of particles can be significantly reduced. Therefore, while providing good performance, we are able to reduce the computational resources considerably. The proposed tracker does not rely on re-sampling or the use of a large sample size to improve its accuracy. Genetic operators are used to efficiently spread particles. Thus, we are not concerned with convergence issues of DE. The proposed framework is able to solve the frequent problem of sample degeneracy and improve the speed of particle filters by reducing the sample size and producing efficient sample propagation.

ACKNOWLEDGMENT

This research was supported by MKE under the HNRC-ITRC support program supervised by the NIPA (NIPA-2010-C1090-1011-0010) and Basic Science Research Program of the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2011-0025855).

6. REFERENCES

[1] G. Sanjeev, A. S. Maskell, N. Gordon, and T. Clapp, ”A tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking,” *IEEE. Trans. Signal Processing*, vol. 50, pp. 174-188, Feb. 2002.

[2] B. Kwolek, “Particle Swarm Optimization Based Object Tracking,” *ACM Fundamenta Informaticae - Swarm Intelligence*, vol. 94, pp. 449- 463, Dec 2009.

[3] S. Park et al. ,”A new Evolutional Particle Filter for the Prevention of Sample Impoverishment,” *IEEE Trans. Evolutional Computation*, vol.12.No.4; August 2009.

[4] S. Saha, N. Bambha, S. Bhattacharyya, “Design and implementation of embedded computer vision system based on particle filter,” *ACM journal of Comp. Vision and Image Understanding*, vol. 114 pp. 1203-1214, Nov. 2010.

Efficient Asynchronous Re-sampling Implementation on a Low-power Fixed-point DSP

Markus Borgh[a], *Christian Schüldt*[b], *Ingvar Claesson*[b]

[a]Limes Audio AB, Tvistevägen 47, SE-90729, Umeå, Sweden.
[b]Blekinge Institute of Technology, Department of Electrical Engineering, SE-37179, Karlskrona, Sweden.

Abstract- **This paper presents an asynchronous re-sampling implementation on a low-power fixed-point DSP, which uses around 47% less computational resources compared to the solution provided by the DSP manufacturer, without compromising audio quality.**

I. INTRODUCTION

In many consumer electronics applications, audio is received from one source, processed in some manner, and then sent to a destination. In a conference phone connected to a computer via USB, for example, the conference phone receives audio through the USB connection, processes it and then sends it to an audio codec chip containing a digital-to-analog (D/A) converter for audio output on the loudspeaker. In many cases the USB interface and the audio codec interface have different clock sources, meaning that the clocks are not synchronized and may differ slightly in frequency, hence the need for *asynchronous re-sampling*.

This paper presents an asynchronous re-sampling implementation based on the approach in [1] and [2]. The implementation is made on a low-power fixed-point digital signal processor (DSP) [3].

II. INTERPOLATION BASED ASYNCHRONOUS RE-SAMPLING

One common approach to re-sampling is to use an interpolation based scheme. The procedure can easily be understood by visualizing an incoming digital signal, clocked by one source, conversion of this digital signal to an analog signal (D/A conversion) and then re-sampling of this analog signal using the clock source of the destination device (A/D conversion). In a digital software world, however, instead of carrying out the D/A - A/D conversion explicitly, the input signal is typically piecewise approximated by a polynomial of some order, and the digital output signal is computed from a point on the polynomial curve. For more in-depth details regarding this matter, the reader is referred to [1] and [2] and the references therein.

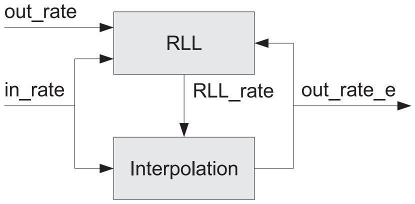


Fig. 1. Scheme of the asynchronous re-sampling.

A key component in the asynchronous re-sampling procedure is an estimator that measures the ratio between input and output sample rates. The estimated ratio is then used to control the re-sampling process. Typically a rate-locked loop (RLL), quite similar to a phase-locked loop (PLL) commonly used in ratio communication, is used for this. The RLL is essentially a control system that constantly measures the input/output sample ratio and adjusts the re-sampling ratio of the Interpolator to match accordingly. Figure 1 shows an illustration of this scheme.

III. REAL-TIME FIXED-POINT IMPLEMENTATION

The re-sampling algorithm in this paper was implemented in real-time on a fix-point digital signal processor [3], and the outline is described below.

In order to determine the rate between the number of incoming samples from the USB interface (`in_rate`) and the number of outgoing samples requested by the audio codec (`out_rate`), a simple RLL approach was used where counters were used to keep tack of the number of incoming and outgoing samples. The number of outgoing samples from the Interpolator (see figure 1) (`out_rate_e`) was also tracked and the difference between `out_rate` and `out_rate_e` was used for calculating the rate used by the Interpolator: `RLL_rate = in_rate / (out_rate + (out_rate - out_rate_e))`.

As for the Interpolator, the Candan approach [2] with a slight modification, where the delay dependent part is moved to the output section such that the filter state is independent of the interpolation delay [5], was used. The order of the polynomial interpolation was set to 3 and a Q2.13

The funding from the Swedish Knowledge Foundation (KKS) is gratefully acknowledged.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

fixed-point number format was used to get sufficient precision without the risk of saturation. To minimize the computational burden, the inner loop of the interpolator was written in assembly language (adopted from the floating-point C-code available from [5]), see listing below, utilizing the hardware circular buffering capability of the DSP.

```
  loop:
      BCC newinput, AC3 < 0           ; If d < 0

  newoutput:
      ; Calculate re-sample filter coefficients
      ; dk0, dk1 & dk2
      MOV #-8192<<#16, AC0
      SUB AC3, AC0                    ; k0 -> AC0
      ADD #8191<<#16, AC0, AC1
      ; (T2 contains QRECIP1)
      MPY T2, AC1                     ; dk1 -> AC1
      || ADD #16383<<#16, AC0, AC2
      MPYKR #QRECIP2, AC2             ; dk2 -> AC2

      ; Filter using calculated coefficients to produce
      ; output sample (Horner's method approach)
      MPYM *db_ptr+, AC2, AC2         ; dk2 in AC2
      SFTS AC2, #2
      ADD *db_ptr+<<#16, AC2
      MPY AC1, AC2                    ; dk1 in AC1
      SFTS AC2, #2
      ADD *db_ptr+<<#16, AC2
      MPY AC0, AC2                    ; dk0 in AC0
      SFTS AC2, #2
      ADD *db_ptr+<<#16, AC2

      ; Advance output time
      SUB dbl(*SP(ratio)), AC3        ; d -= RLL_rate

      ROUND AC2 || ADD #1, T0         ; numout++
      MOV HI(AC2), *out_ptr+          ; Store output

      BCC newoutput, AC3 >= 0

  newinput:
      ; Place new input sample in circular buffer
      ; (It is assumed that the input is downshifted
      ; by 2 bits to avoid saturation)
      MOV *in_ptr+, *db_ptr || ADD #1, T1 ; numin++

      ; Calculate the backward differences
      SUB *db_ptr-, *db_ptr2-, AC0
      MOV HI(AC0), *db_ptr- || SUB *db_ptr2-<<#16, AC0
      MOV HI(AC0), *db_ptr- || SUB *db_ptr2-<<#16, AC0
      MOV HI(AC0), *db_ptr

      ; Advance input time
      ADD #8191<<#16, AC3             ; d += 1.0

      ; Processed all input samples? If no, loop again.
      BCC loop, T1 < #48
  exit:
```

It the above listing, the temporary variables `T1` and `T0` hold the number of processed input and output samples, respectively. The pointer `db_ptr` refers to a hardware circular buffer and the pointer `db_ptr2` points to the same circular buffer with offset -1. The constants `QRECIP1` and `QRECIP2` correspond to the numbers 16384 (1.0/2.0 in Q15 format) and 10922 (1.0/3.0 in Q15 format), respectively.

A. Evaluation

The presented implementation was compared to the re-sampling provided with the audio software framework by the DSP manufacturer [4]. Comparison was made with respect to both audio quality and computational load.

A.1. Computational load

Measuring the number of required clock cycles for re-sampling one 1 ms audio frame sampled at 48 kHz at a `RLL_rate` between 0.95 and 1.05, gave that the presented solution required between 3763 and 3966 cycles, while the solution provided by the DSP manufacturer required between 7152 and 7269 cycles, depending on the `RLL_rate`. This means that the presented solution requires around 47% less clock cycles in this case.

A.2. Audio quality

The Total Harmonic Distortion + Noise (THD+N) was measured using tones of varying frequency which were re-sampled using both methods respectively. Both re-sampling methods gave $< 1\%$ THD+N for tones ranging between 20 Hz and 20 kHz, using a notch-filter of varying frequency and a Q-factor of 20 to separate the distortion from the pure tone.

Several listening tests were also conducted and no differences between the original signal and the re-sampled signal could be heard for any of the re-sampling solutions.

IV. CONCLUSION

This paper presented a fixed-point implementation of a asynchronous re-sampling algorithm on a low-power DSP. The implementation used around 47% less computational resources than the re-sampling implementation provided in the audio software framework by the DSP manufacturer. Subjective tests showed that listeners where unable to hear any difference in sound quality.

V. REFERENCES

[1] C.W. Farrow, "A continuously variable digital delay element," *Proc. IEEE Int. Symp. Circuits and Systems,* pp. 2641-2645, 1988.

[2] C. Candan, "An efficient filtering structure for Lagrange interpolation," *IEEE Signal Proc. Letters,* vol. 14, no. 1, pp. 17-19, Jan. 2007.

[3] *TMS320C55x DSP CPU Reference Guide,* Texas Instruments Inc., Literature Number SPRU371F, Feb. 2004.

[4] *C55x Connected Audio Framework,* Texas Instruments Inc., `http://www.ti.com/tool/c55x-audioframework`, [Online; accessed 09-July-2012].

[5] M. Boytim, *Asynchronous Sample Rate Conversion (ASRC) for Matlab (and C),* `http://home.comcast.net/ matt.boytim/asrc/`, [Online; accessed 09-July-2012].

Total Variation Regularization Algorithm for Video Stabilization in a Digital Camera

Wooram Son, Sungbin Hong and Seonghun Kim
Digital Imaging Business, Samsung Electronics Co., Ltd., Suwon, Korea

Abstract—This paper presents a motion modeling technique for digital video stabilization system. Our modeling approach may be applied for separation of the intentional camera motion and undesired shaky motion from estimated motion vector. The technique used for this separation is called total variation regularization technique and we believe that this motion modeling technique is well-suited for removal of unintentional motion noise. The experimental results establish the fact that our proposed methodology improves the quality of video stabilization.

I. Introduction

Video stabilization technique is one of the most important problems in handheld camera devices that remove undesired motions generated by camera-shake. Stabilization scheme generally consists of three main parts and processed in the following: motion detection, motion filtering, and motion compensation. In most case, video captured by handheld device contains user's intentional motion. Therefore, it is important to extract the intentional motions from detected motions by motion filtering. The performance of motion filtering part depends on the accuracy of differentiation between desired and undesired motions [1]. This paper introduces an effective motion modeling method based on the total variation regularization (TV) that can be used for modeling of intentional motions. TV regularization technique is used for solving inverse problems. This technique was originally introduced by Rudin, Osher and Fatemi [2]. Since, its inception the TV regularization technique has been applied to solve a wide range of imaging problems. We believe that the proposed motion modeling technique is well-suited for removing unintentional motion noise. The experimental results prove that our proposed methodology improves the quality of video stabilization.

II. Algorithms

A. Video Stabilization Procedure

Our video stabilization system is composed of three modules: motion detection, motion modeling, and motion compensation. The motion detection module generates estimated motion vector using a frame-by-frame comparison [3]. Then, the motion modeling module performs accumulation of motion vectors and applies filtering that allows separation of unintentional motion vector from accumulated motion vectors. Finally, output video frame is moved to the opposite direction of estimated unintentional motion vector by motion compensation module.

B. Total Variation Regularization

The basic TV regularization scheme can be defined as a minimization problem in equation (1) having a data fidelity term reflecting the noise motion characteristics of the detected motion provides an effective motion filtering. This expression is called the ROF model [2].

$$\inf_u F(u) = \int_\Omega |\nabla u| dx + \lambda \int_\Omega |f-u|^2 dx \tag{1}$$

, where f is the detected motion, u is an intentional motion, and λ is a positive scalar specifying the fidelity weight that should be adjusted to match the noise motion level [4] [5] [6].

C. Proposed motion modeling algorithm

As a first step, motion modeling module accumulates scalar magnitude value on the x and y-axis of motion vector that is generated by motion detection module as shown in equation (2). Hence, the TV regularization scheme which minimizes equation (3) estimates intentional movement on the x and y-axis from accumulated motion. Finally, unwanted noisy movements are calculated by equation (4).

$$\mathrm{AccMV}_{x,y}(k) = \sum_{i=0}^{k} \mathrm{MV}_{x,y}(i) \tag{2}$$

, where AccMV_x, AccMV_y are accumulated movement values along with the x-axis and y-axis respectively.

$$\begin{aligned} &\underset{\mathrm{IMV} \in BV(\Omega)}{\mathrm{argmin}} \ \| \ \mathrm{IMV} \ \|_{TV(\Omega)} + \\ &\frac{\lambda}{2} \int_\Omega (\mathrm{AccMV}(x) - \mathrm{IMV}(x))^2 dx \end{aligned} \tag{3}$$

, where IMV is an intentional movement, λ is a positive parameter controlling stabilization level. Equation (3) is performed as a one-dimensional minimization problem in all bounded variation (BV) search space.

$$\mathrm{FilteredMV}_{x,y}(k) = \mathrm{AccMV}_{x,y}(k) - \mathrm{IMV}_{x,y}(k) \tag{4}$$

, where $\mathrm{FilteredMV}_x$, $\mathrm{FilteredMV}_y$ are estimated unintentional movement along with the x-axis and y-axis respectively that is used for motion compensation of output video frame.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. Experimental Results

In our experiment, we have tested three detected motion vector series generated by motion detection module on compact system camera with 18-55mm lens (maximum zoom in); Test set (a): a video captured with handheld shots and no intentional movement, Test set (b): a video captured with natural handheld shots and intentional movement, Test set (c): a video captured on tripod with intentional movement [3]. The distance between camera and subject is 30-50cm. The distance function of TV regularization is the Gaussian model. The motion modeling algorithms were implemented and experimented on PC based simulation software. The Figure 1 relates to Test set (a) above which shows that the estimated intentional and unintentional movement along with horizontal-axis. The Figure 2 relates to Test set (b) above which shows unwanted movements are effectively reduced while preserving intentional movements. The Figure 3 relates to Test set (c) above which shows that our proposed scheme maintains the detected motion vector when it performs on steady shot situation. Table 1 is the comparative performance of distance models. All experiments were conducted for horizontal, vertical movement. Due to space constraints and ease of representation results of horizontal movement have been displayed.

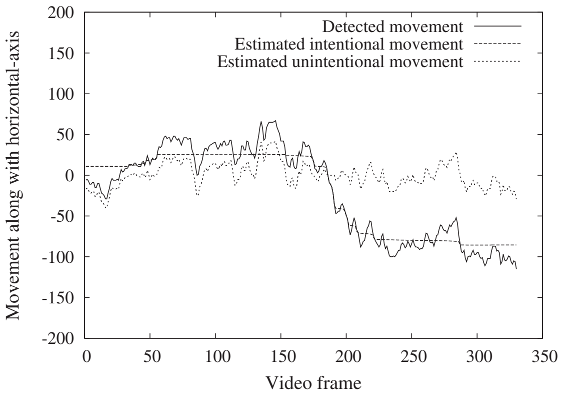


Fig. 1. Experimental results on detected motion vector series (a), The y-axis displays the movement along with horizontal-axis on camera. The x-axis represents time (time unit=33milliseconds).

IV. Conclusion

This paper presented is a new motion modeling algorithm for video stabilization system using total variation regularization. The experimental results demonstrate that the proposed algorithm could be a potential candidate for motion modeling algorithm of video stabilization system in consumer electronics. The hardware acceleration and comparative study with other algorithms are considered for future work.

References

[1] K. Lee, Y. Chuang, B. Chen, and M. Ouhyoung, "Video stabilization using robust feature trajectories," in Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009, pp. 1397–1404.

[2] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D: Nonlinear Phenomena, vol. 60, no. 1, pp. 259–268, 1992.

[3] Y. Lee and Y. Choi, "Method and apparatus for video stabilization by compensating for video direction of camera," Patent US 7 062 320, 05 17, 2012.

[4] S. Alliney, "A property of the minimum vectors of a regularizing functional defined by means of the absolute norm," Signal Processing, IEEE Transactions on, vol. 45, no. 4, pp. 913–917, 1997.

[5] T. Le, R. Chartrand, and T. Asaki, "A variational approach to reconstructing images corrupted by poisson noise," Journal of Mathematical Imaging and Vision, vol. 27, no. 3, pp. 257–263, 2007.

[6] P. Getreuer, "Rudin-osher-fatemi total variation denoising using split bregman," Image Processing On Line, 2012.

TABLE I
Comparative Performance of Distance Models

Model	Root mean square error		
	Test set (a)	Test set (b)	Test set (c)
Gaussian	12.22	23.86	2.82
Laplace	9.71	14.37	0.76
Poisson	71.71	193.42	0.76

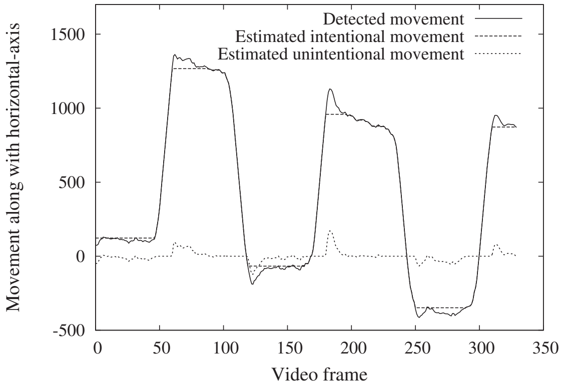


Fig. 2. Experimental results on detected motion vector series (b), The y-axis displays the movement along with horizontal-axis on camera. The x-axis represents time (time unit=33milliseconds).

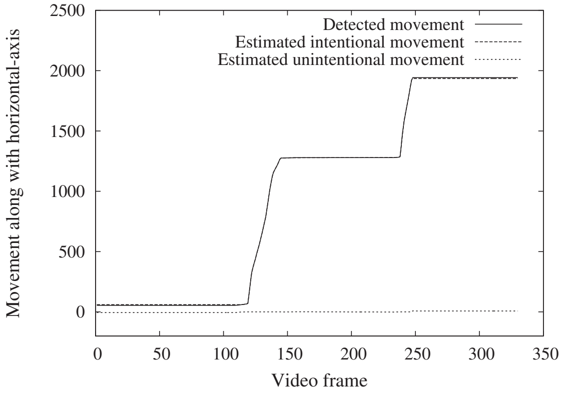


Fig. 3. Experimental results on detected motion vector series (c), The y-axis displays the movement along with horizontal-axis on camera. The x-axis represents time (time unit=33milliseconds).

A Markovian Algorithm for Creating Immersive Public-Speaking Audiences

Nicklaus THOMAS, *Student Member, IEEE,* David EVANS, Samuel RUSS, *Member, IEEE*

***Abstract*-- A person's view of this world is directly related to the type and amount of sensory data collected. Manipulating this sensory data would allow us to craft experiences in which we control every variable, granting an effective tool for immersion therapy. Consumer electronics has advanced to the point where consumer technologies, such as head-mounted displays and motion-tracking gaming systems, can be employed to create a cost-effective virtual reality (VR) system. One can now substitute sensory information using these technologies. This paper describes how one can gain a level of significant immersion within a virtual classroom for a participant that is given a speaking task, particularly a participant who stutters, and do so with low-cost consumer equipment.**

I. Introduction

After a review of modern, cost-effective VR technology and its uses for creating an environment for speaking, this will describe an algorithm for generating realistic audiences and show some of the results of the work.

II. Review of Current Technology

A. *Virtual Reality*

Virtual reality is a widely-used technology for applications such as education [1]. VR research has been widely published and is an active area of research for social scientists [2]. Because brain-imaging studies have confirmed the close correlation of VR-induced brain activity to physical brain activity, VR is now also recognized as a therapeutic tool [3],[4].

One method of presenting visual and audio stimuli to participants is a consumer-electronic head-mounted display (HMD) system [2], the fidelity of which can now be assisted through low-cost commercially available gaming motion-tracking systems.

B. *VR for Speaking Tasks*

A common application for VR is to create realistic audiences for persons performing public-speaking tasks. Even in the relatively early days of VR research, VR audiences have created emotional responses similar to actual audiences [5],[6]. In addition, more and less challenging virtual speaking tasks have influenced the frequency of stuttering in people who stutter (PWS) [8]. Recent research in creating realistic audiences has focused on making measurements of actual audiences to increase the fidelity of the simulation [7].

The focus on this work is on creating a simple, powerful algorithm that expresses realistic audience behavior.

C. *Therapies for Persons Who Stutter (PWS)*

Drawing on the potential of VR in speech therapy for PWS [8] and of VR to create realistic audiences, this work focuses on creating realistic audience-based public-speaking tasks for PWS. By using VR, the "audience" is repeatable, can be made available at any time, and will maintain client privacy, all of which are significant advantages for clinical studies of therapeutic efficacy across multiple patients.

The creation of the VR audience, discussed below, is the first step. Future work will be carried out to compare VR public speaking tasks and actual public-speaking tasks with PWS.

III. Virtual Reality Audience

A. *Hardware*

A high-end desktop computer was selected and attached to a HMD. A commercially available motion-tracking system was added to increase the immersion of the participant in the virtual environment. The HMD and motion-tracking system are kept in a separate room so that the operator can freely control the VR environment during patient use. The display and tracking system are shown below in Figure 1.

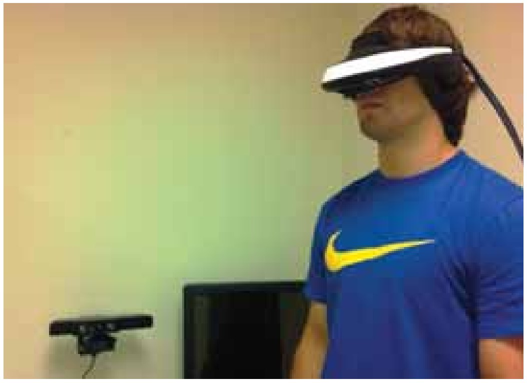


Fig. 1. The head-mounted display and tracking system are shown. The tracking system is on the left, and the user faces it in normal operation.

B. *VR Environment*

A VR environment was created using commercially available software, and realistic avatars representing a college-aged audience were purchased. The environment was then augmented using photos of doors and windows from university classrooms. Further, a 3D model replica of the desks used in the classroom were created and rendered in the virtual classroom. Thus the furniture, carpet, paint, and windows in

The authors are affiliated with the University of South Alabama. This work was sponsored by the University of South Alabama Faculty Development Council and the University's UCUR Undergraduate Research Program.

the simulation matched an actual classroom.

C. Algorithm for Audience Generation

An algorithm was needed to create an audience with some members that appeared bored and looked away from the speaker. A Markovian model of attention was created, as shown below in Figure 2.

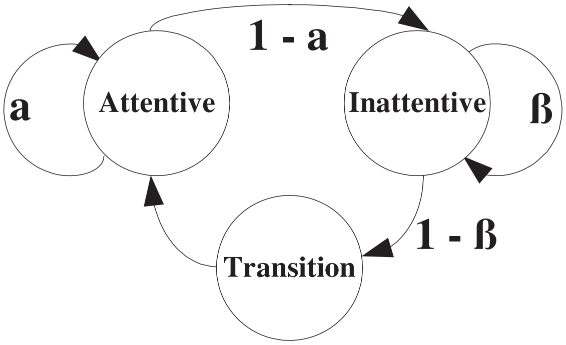


Fig. 2. Markovian model for audience member. The transition probabilities are selected to control how inattentive the audience appears.

The diagram shown in Figure 2 illustrates the state-based model used by each audience member. Each audience member has their own state machine and is in one of three different states. Periodically (e.g. every three seconds) the algorithm generates a random number for each audience member and "decides" whether the audience member should stay in the same state or change states. The α probability dictates how likely an audience member is to remain attentive. If α is lowered, each audience member is more likely to become inattentive and the audience will appear less interested.

The algorithm is designed to reflect the fact that people often alternate between "paying attention" and "daydreaming" during a lecture activity. The transition state was added so that a smooth head-neck motion back to the speaker could be added; without it, VR audience members tend to "snap" back to the speaker which is both distracting and unrealistic.

Besides being very simple to code, the algorithm is also extensible. Other behaviors, such as coughing or grooming, can be added to the model. Drawing on recent work correlating attentiveness to classroom seating [7], each member's transition probabilities can also be customized.

D. Results

An example screen shot from the simulation is shown below in Figure 3. The perspective is set to match a person standing in a level classroom.

Some of the audience members are paying attention (e.g. the member on the far left) and look at the participant. As the person wearing the head-mounted display moves, the audience members that are paying attention will track the motion and remain looking at the participant.

Other audience members are inattentive (e.g. the second and third members from the left). They "stare off into space". Because of the Markovian algorithm, they will do so for a brief time and then their gaze will return to the speaker.

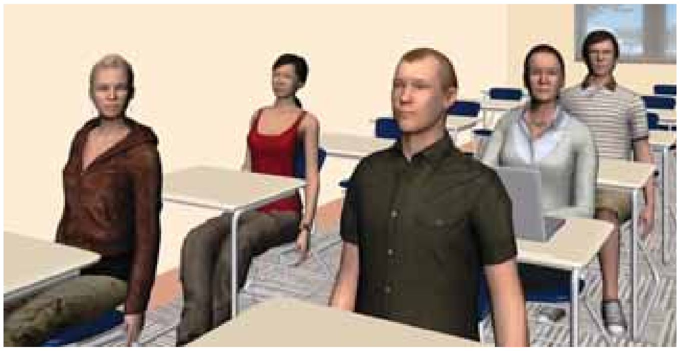


Fig. 3. Screen shot from actual VR simulation. Note how some of the audience members are looking at the speaker and some are not.

At any moment of time, most of the audience members are looking at the speaker and a few are looking away, and the members change their gaze over time because of the pseudo-random behavior of the algorithm. Setting the α probability lower or the β probability higher results in an audience that appears more inattentive. Thus there is direct, repeatable control over the appearance of the audience for clinical use.

IV. Conclusions and Future Work

The algorithm proved to be easy to implement and promises to be extensible. In addition to adding more realistic postures and behaviors, such as coughing and grooming, the VR environment will begin to be used in research and clinical settings, such as the treatment of stuttering.

Acknowledgment

We wish to acknowledge the support of the University of South Alabama Faculty Development Council and the university's University Committee on Undergraduate Research (UCUR) Undergraduate Research Program.

References

[1] H. Hoffman and D. Vu., "Virtual Reality: Teaching Tool of the Twenty-first Century?", *Academic Medicine*., vol.72, no:12, PP.1076-1081, Dec 1997.

[2] J. Fox, et al., "Virtual Reality: A Survival Guide for the Social Scientist," *J. Media Psychology*., vol.21, no:3, PP.95-113, June 2009.

[3] C. J. Bohil, et al., "Virtual reality in neuroscience research and therapy," *Nature Reviews: Neuroscience*., vol.12, PP.752-762, Dec 2011.

[4] G. Riva., "Virtual Reality in Psychotherapy: Review," *CyberPsychology & Behavior*., vol.8, no:3, PP.220-230, 2005.

[5] M. Slater, et al., "Public Speaking in Virtual Reality: Facing an Audience of Avatars," *IEEE Comput. Graph. Appl*., vol.19, no:2, pp.6-9, Mar/Apr 1999.

[6] D-P. Pertaub, et al., "An Experiment on Fear of Public Speaking in Virtual Reality," *Studies in Health Technology and Informatics*, Vol 81, 2001, pp. 372-378.

[7] S. Poeschl and N. Doering. "Virtual Training For Fear Of Public Speaking – Design of an Audience for Immersive Virtual Environments," *Proceedings of 2012 IEEE Virtual Reality (VR)*, pp. 101-102.

[8] S. B. Brundage, et al., "Frequency of stuttering during challenging and supportive virtual reality job interviews," *J. Flu. Dis*., vol.31, no:4, PP.325-329, Aug 2006..

Perceived Distortion Aware Backlight Dimming for Low Power and High Quality LCD Devices

Dong-Gon Yoo, *Student Member, IEEE*, and Young Hwan Kim, *Member, IEEE*
Division of Electrical and Computer Engineering, POSTECH

***Abstract*—This paper presents a perceived distortion control algorithm that considers the viewer's perception for the image distortion caused by reduced backlight brightness. The basic idea of the proposed method is to maintain the levels of image distortion for all local image areas below the distortion level that can be perceived by viewers.**

I. Introduction

It is known that the backlight unit accounts for the largest portion of the total power consumption of liquid crystal displays (LCDs), and therefore global backlight dimming is a very effective method of reducing their power consumption [1]. In general, global backlight dimming consists of three processes: image analysis, backlight modulation, and image compensation [1]. The image analysis process determines the clipping point by analyzing the input image, and then determines the dimming rate, which controls the backlight brightness. The backlight modulation process controls the backlight brightness using the dimming rate. The image compensation process compensates for the luminance loss by controlling the LC transparency. However, this process cannot compensate for image distortion that may occur for some pixels above a clipping point.

Recently, efforts have been invested to develop algorithms that can overcome the image quality degradation incurred by global backlight dimming [1]-[3]. However, these algorithms still experience image distortion in some local areas. In this paper, we propose a perceived distortion control algorithm that preserves the image quality even in local areas.

II. Proposed Perceived Distortion Control Algorithm

The proposed perceived distortion control (PDC) algorithm is intended to preserve image quality in all the local areas of the image. It controls the image distortion for the local area that is likely to experience the most significant image distortion as the backlight brightness decreases, so that the local area and the others are below the distortion level that viewers can perceive. The proposed method is divided into two steps as shown in Fig. 1. The first step is to find the local area that is likely to have the most significant image distortion as the backlight decreases. The second step is to determine the clipping point (I_{CP}) for the local area found in the first step.

To preserve the image quality in local areas, the amount of image distortion per local area should be limited. The proposed method uses two parameters to achieve this: just noticeable size (*JNS*) and just noticeable difference (*JND*). *JNS* denotes the local area size to limit the image distortion area size, and *JND* denotes the minimum amount of image distortion for the local area that viewers can perceive. The *JNS* and *JND* values are determined by the viewers because each viewer has a different sensitivity to image distortion. This sensitivity is affected by various parameters, such as display size, image resolution, viewing distance, and ambient brightness. We assume that viewers cannot perceive image distortion whose area size and amount are smaller than *JNS* and *JND*, respectively.

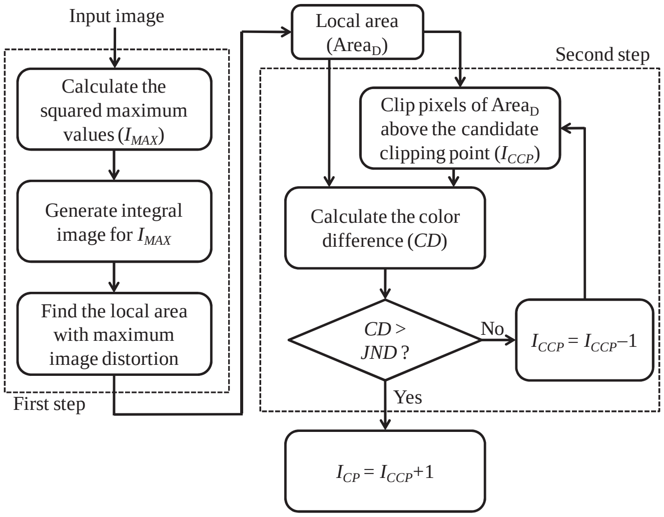

Input image
Calculate the squared maximum values (I_{MAX})
Generate integral image for I_{MAX}
Find the local area with maximum image distortion
First step
Local area (Area$_D$)
Second step
Clip pixels of Area$_D$ above the candidate clipping point (I_{CCP})
Calculate the color difference (*CD*)
CD > *JND* ?
No
$I_{CCP} = I_{CCP}-1$
Yes
$I_{CP} = I_{CCP}+1$

Fig. 1. Flowchart of the proposed perceived distortion control algorithm

The proposed method first finds the local area (Area$_D$) that is likely to have the most significant image distortion as the backlight decreases. As mentioned above, the proposed method limits the area size of the image distortion by using a block whose area size is *JNS*. In backlight dimming techniques, image distortion is caused by clipping the pixels above I_{CP}. Since I_{CP} is unknown and is determined at the end of the proposed method, we cannot calculate the image distortion for local areas. Thus, the proposed method estimates the amount of image distortion for local areas by assuming that the higher the RGB values, the more is the image distortion affected, represented as follows.

$$\begin{aligned} I_{MAX}(m,n) &= \max\{R(m,n), G(m,n), B(m,n)\}^2 \\ ED(B_{i,j}) &= \sum_{(m,n)\in B_{i,j}} I_{MAX}(m,n) \end{aligned} \tag{1}$$

where *R*, *G*, and *B* stand for the *R*, *G*, and *B* values of each pixel of the image, respectively. I_{MAX} denotes the square of the maximum among the RGB values of each pixel, *ED* denotes

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (20100023428) and IT Consilience Creative Program of MKE and NIPA (C1515-1121-0003).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

the estimated amount of image distortion of the local area, and $B_{i,j}$ denotes the block of which the left and top position is vertically the *i*-th and horizontally the *j*-th pixel, respectively. I_{MAX} values are generated before calculating *ED* to avoid redundant calculation. The proposed method finds Area_D by searching the local area with the largest *ED* value, as follows.

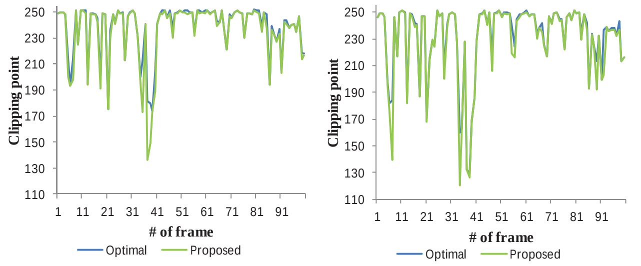
$$Area_D = \arg\max_{(i,j)}\{ED(B_{i,j})\} \qquad (2)$$

However, calculating *ED* for all local areas requires high computational complexity. Thus, we reduce the computational complexity using the integral image method [4]. We generate the integral image for I_{MAX}, and then calculate *ED* simply by using this integral image. By calculating the integral image for I_{MAX}, we can reduce the computational complexity of the *ED* calculation by more than 99%.

The second step of the proposed method is to control the image distortion of Area_D to below the distortion level that viewers can perceive. The distortion level that viewers can perceive is defined by *JND*. The proposed method considers *JND* as a just noticeable color difference in the *CIELab* color space [5]. First, the proposed method sets the candidate clipping point (I_{CCP}) to the maximum gray level of Area_D, and calculates the color difference between the original pixel values and the pixel values clipped above I_{CCP} for the Area_D in the *CIELab* color space. Then, the proposed method decreases I_{CCP} by 1 if the color difference is smaller than *JND*. This process is repeated until the color difference is smaller than *JND*. At the end of the repetition, I_{CCP} is the first clipping point that causes an image distortion perceivable by viewers. Thus, the proposed method determines I_{CP} as $I_{CCP} + 1$.

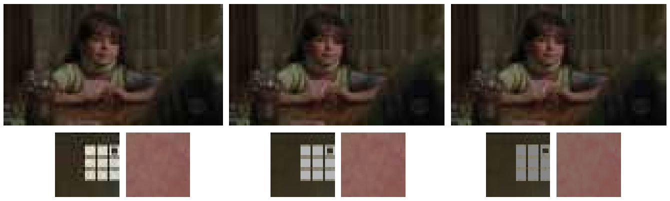
III. Experimental Results

We evaluated the performance of the proposed PDC algorithm by comparing it with the optimal clipping point of each image. The optimal clipping points were obtained by calculating the color differences between the original pixel values and the clipped values for all local areas whenever I_{CP} decreased by 1 and by confirming that the color differences for all local areas were below *JND*. We examined the proposed method using two *JNS*s: 32 × 32 and 64 × 64. *JND* was set to 2.3 [5]. We used 100 randomly captured images. Fig. 2 shows the clipping points for 100 test images using the optimal clipping points and those of the proposed method. In most images, the clipping points of the proposed method are comparable with the optimal clipping points. The average differences between the optimal clipping points and those obtained by the proposed method are 2.64 and 2.69 when the two *JNS*s were set to 32 × 32 and 64 × 64, respectively. Fig. 3 shows the sample image comparison when the sample image was the worst case in terms of the clipping point difference. Below the whole images, the left small block denotes the most distorted local area caused by the optimal clipping point, and the right small block denotes Area_D found by the proposed method. The optimal clipping point and that obtained by the proposed method were 181 and 136, respectively. In the whole image, we hardly find any image distortion. However, image distortion can be observed in the most distorted local area (left small block) of the proposed method. This is because Area_D was incorrectly predicted. To mitigate this problem in the proposed method, viewers should decrease *JNS*.



(a) *JNS* is 32×32 (b) *JNS* is 64×64

Fig. 2. Optimal clipping points and those obtained by the proposed method when *JNS* is (a) 32 × 32, and (b) 64 × 64



(a) Original image (b) Optimal result (c) Proposed method

Fig. 3. Comparison of (a) original image with the clipped images by (b) the optimal clipped point and (c) the proposed method when *JNS* is 32 × 32

IV. Conclusion

In this paper, we proposed a perceived distortion control algorithm to be used for backlight dimming. The proposed method reduces the backlight brightness considering the viewer's perception of the image distortion incurred by reduced backlight brightness. The proposed method first finds the local area of the image that is likely to have the most significant image distortion as the candidate clipping point decreases. Then, it determines the clipping point for the local area. In the experimental results, the average differences between the optimal clipping points and those obtained by the proposed method were about 2.6, and we confirmed that the proposed method preserves the image quality in the worst case.

Reference

[1] S.-J. Kang and Y. H. Kim, "Image integrity-based gray-level error control for low power Liquid Crystal Displays," *IEEE Trans. Consumer Elec.*, vol. 55, no. 4, pp. 2401-2406, Nov. 2009.

[2] N. Chang, I. Choi, and H. Shim, "DLS: dynamic backlight luminance scaling of liquid crystal display," *IEEE Trans. Very Large Scale Integration Systems*, vol. 12, no. 8, pp. 837-846, Aug. 2004.

[3] S.-J. Kang and Y. H. Kim, "Multi-histogram-based backlight dimming for low power Liquid Crystal Displays," *IEEE/OSA Journal of Display Tech.*, vol. 7, no. 10, pp. 544-549, Oct. 2011.

[4] P. Viola, and M. J. Jones, "Robust real time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.

[5] G. Sharma, "Digital Color Imaging Handbook," CRC Press, 2003.

A Comparative Evaluation of Perceptibility of Lip-Sync Errors for Different Shots and Angles in 3DTV and 2DTV

A. Fedina, E. Grinenko, and K. Glasman, *Member, IEEE*

***Abstract*- This paper describes an experimental study for 3D audiovisual programs. Experimental evaluation was conducted to give the quantitative answer to the question if any difference between the perceptibility of impairments caused by audio-to-video delays errors for different shots and angles for 3D and 2D multimedia programs for one and the same content exists.**

I. INTRODUCTION

The relative timing of sound and video components of the television signal is an important point of the viewers' perception of television programs. Expectations of the viewing public for the quality of television picture and sound are constantly changing. Besides the successful start of 3DTV, creation of new systems and algorithms media delivery and presentation there are changes of "cinematographic techniques" such as the choice of shot and angle of shooting. It can greatly influence the structure and meaning of a television program. Such kind of variety of shots can increase quality of impression making semantic emphasis. How does changing of shot and angles of shooting influence on evaluation of visibility of lip-sync errors in 3D and 2D content? The results of this study can be used to find the thresholds of delectability and acceptability of relative timing of sound and video for 3D audiovisual programs compared with the thresholds for 2D.

Audio and video production, distribution and broadcast digital television systems are complex arrays of compression, decompression, processing and storage devices. Each of these components in the chain causes a delay on the audio and video signals flowing through it. The perceptibility of relative time difference of audio and video components in 3DTV differs from 2DTV. The thresholds values of delectability and acceptability of audio-to-video delays errors given in the Recommendation ITU-R BT.1359-1 are only for 2DTV[1]. There are a lot of new factors affecting human perceptual experience in 3DTV. The most popular television programs use unusual shot sizes and angles to influence the meaning which an audience will interpret. The studying of mechanisms of perception of 3D images is not enough. There is no information about influence of shot and angle of shooting on perception of lip-sync error. The goal of this paper is to study the influence of lip-sync errors on audiovisual perceptual quality in 3DTV for different shots and angels.

II. EXPERIMENTAL PROCEDURE

A. *Experimental evaluation of visibility of lip-sync errors in 3D and 2D using active shutter glasses*

Audiovisual quality assessments of test sequences for stereoscopic 3D and 2D content were carried out [2,3]. The relationships between the relative time difference of audio and video components (in ms) and the MOS of audiovisual perceptual quality for 3D and 2D sequences are shown in Figure 1 when sound leading video. The confidence intervals shown in the figures are corresponding to the level of confidence probability equal to 0.75. As it described in Recomendation ITU-R BT.1359-1 the level of audiovisual perceptual quality which is equal to 4.5 can be considered as a threshold of detectability. It can be seen from experimental results shown on Figure 1 that the threshold of detectability when sound leading video for 2D images is about +50 ms and +27 ms for 3D images.

When we investigated user perception of lip-sync errors in 3D content different pictures (left and right images) were presented to the left and right eye of the user using active shutter glasses. When we investigated user perception of lip-sync errors in 2D content one picture (left image) was presented to the user using switched-off position of 3D active shutter glasses[4,5]. The glasses were used in this case only to make viewing conditions constant, such as color, brightness and contrast of the picture.

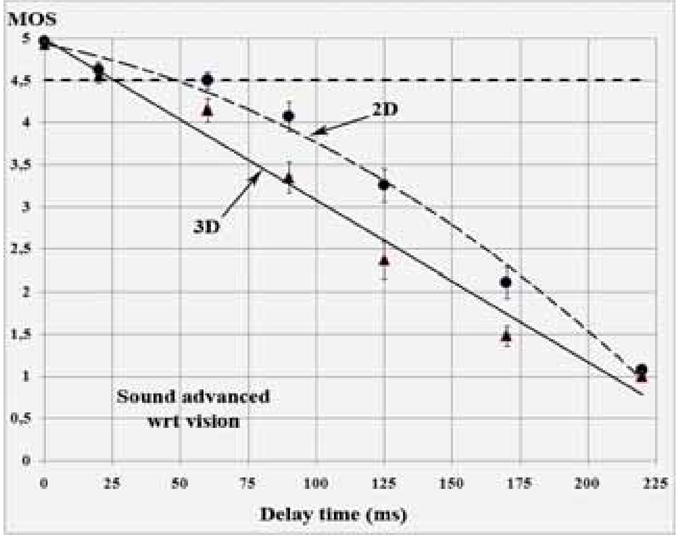


Fig. 1. The relationships between the relative time difference of audio and video components (in ms) and the MOS for sound advanced for 3D and 2D images

After the experimental evaluation of visibility of lip-sync errors in 3D and 2D content assessors were interviewed. Some

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

of the assessors said that there is an influence on human perception of lip-sync errors caused by usage of active shutter glasses. Due to shutter switching an assessor can see different phase of lips' movement. The question is if the difference in relative timing of audio and video perception in 3D and 2D is caused by the mechanisms of perception of the facial expression closed to realistic three-dimensional field or just by usage of active shutter glasses.

B. Experimental evaluation of visibility of lip-sync errors in 3D and 2D using anaglyph stereoglasses

The question is if the difference in relative timing of audio and video perception in 3D and 2D is caused by the mechanisms of perception of the facial expression closed to realistic three-dimensional field or just by usage of active shutter glasses. We decided to carry out new audiovisual quality assessments using anaglyph stereoglasses.

This method is free from any kind of stroboscopic effect which could be influence on human perception of lip-sync errors caused by usage of active shutter glasses. The results obtained in two series of experiments conducted using two different modes are shown on Figure 2.

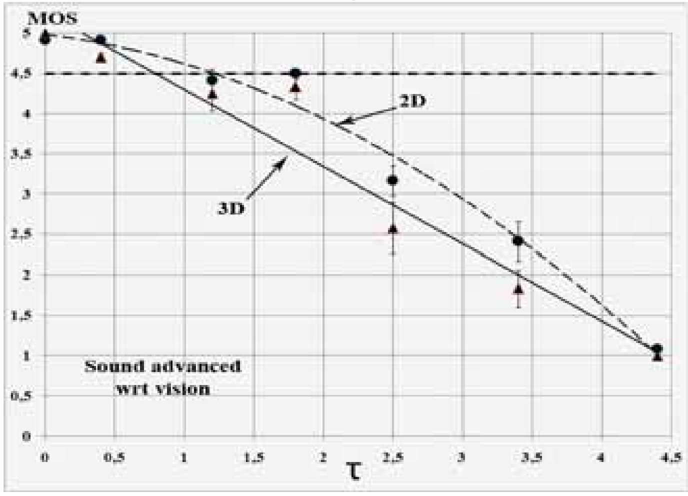


Fig. 2. The relationships between the relative time difference of audio and video components and the MOS for sound advanced for anaglyph 3D and 2D presented using anaglyph stereoglasses.

Anaglyph 3D effect was achieved by means of encoding each eye's image using red and cyan filters. When we investigated user perception of lip-sync errors in 2D content one picture (left image) was presented to the user using anaglyph stereoglasses.

The results of quality assessments prove that viewing public requires more accurate lip synchronization if television programs' viewers watch stereoscopic 3DTV. The threshold of detectability when sound leading video for 2D images is greater than 3D images by 50%. X-axis is the relative time difference of audio and video components. It is a normalized value. The normalization constant is equal to the threshold of detectability when sound leading video for 2D images.

C. Experimental evaluation of visibility of lip-sync errors in 3D and 2D for different shot size and angles of shooting

The most popular television programs use unusual shot sizes and angles to influence the meaning which an audience will interpret. (Fig.3). How does it influence on evaluation of visibility of lip-sync errors in 3D and 2D content?

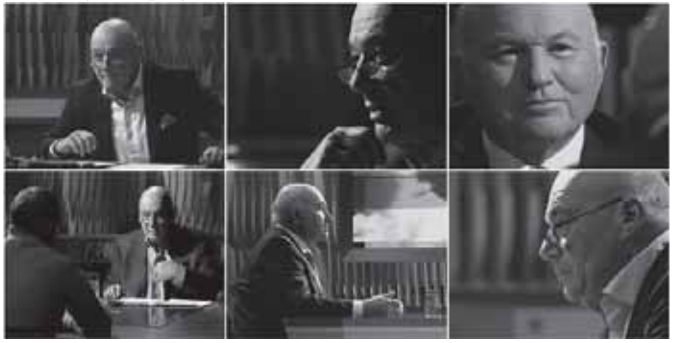


Fig. 3. A set of shots from one of the most popular television program in Russia

The most widely-distributed angle of shooting for newsreader is a medium close up: half-way between a mid shot and a close-up. Usually it covers the subject's head and shoulders (Fig.4 a). We decided to depart from the custom and to change the medium close shot to close-up shot: newsreader's head takes up the whole frame (Fig. 4 b). A comparative evaluation of perceptibility of lip-sync errors for medium close shot and close-up shot in 3D and 2D makes it possible to study the influence of head motion, facial expression and other factors on human perception.

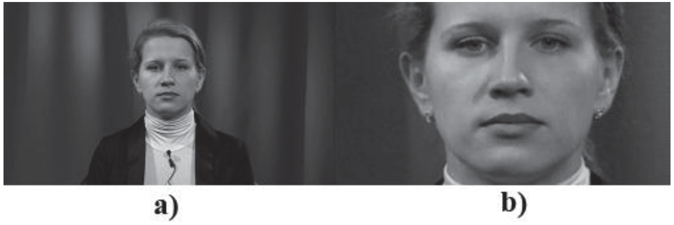


Fig. 4. Types of shots (a-medium close shot; b- close-up shot) for comparative evaluation of perceptibility of lip-sync errors in 3D and 2D.

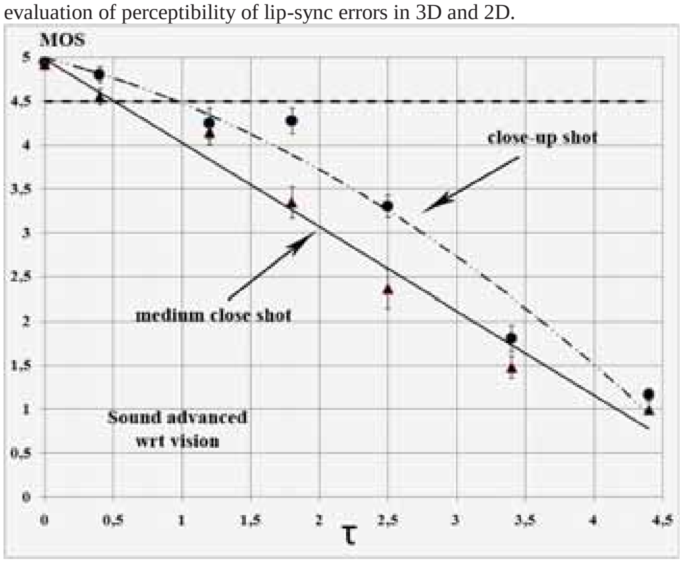


Fig. 5. The relationships between the relative time difference of audio and video components and the MOS for sound advanced for medium close shot and close-up shot for stereoscopic 3D images.

Audiovisual quality assessments of test sequences for two different shots for stereoscopic 3D and 2D content were carried out. The relationships between the relative time difference of audio and video components and the MOS of audiovisual perceptual quality for medium close shot and close-up shot in 3D are shown in Figure 5 when sound leading video. It can be seen from experimental results that the threshold of detectability when sound leading video for close-up shot is greater than medium close shot by 50% for 3D images. There is an idea that speaker head motion may be linguistically informative [6]. It can influence on audiovisual perception of talking face and detectability of lip-sync error because motion of the head is integrated with the system generating speech. The close-up shot makes less evident speaker head motion.

The results of quality assessments sequences for 2D content didn't give detectable difference in perception for medium close shot ant close-up shot. The difference is not statistically significant in both cases when sound leading or delayed to video for 2D images.

There are various degrees of medium close shot depending on which camera angle was chosen. The newsreader can be shot from different angles (Fig.6). The front view (0 deg) and side view (80 deg) of talking head can give us different assessments of perceptibility of lip-sync errors. We decided to study the influence of angle's changing on audiovisual perceptual quality. It was chosen three angles: 0 degrees, 40 and 80 degrees.

It can be seen from experimental results shown on Figure 7 that the threshold of detectability when sound leading video for 0 degrees (a front view) is greater than 40 and 80 degrees by 60% for 3D images. But there is no difference for 2D images which can be proven by statistical analysis.

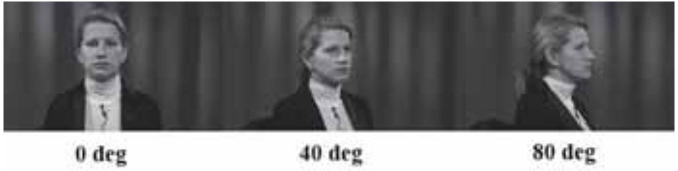


Fig. 6. Types of angles (rotation by 0, 40 and 80 degrees about an axis) for comparative evaluation of perceptibility of lip-sync errors in 3D and 2D.

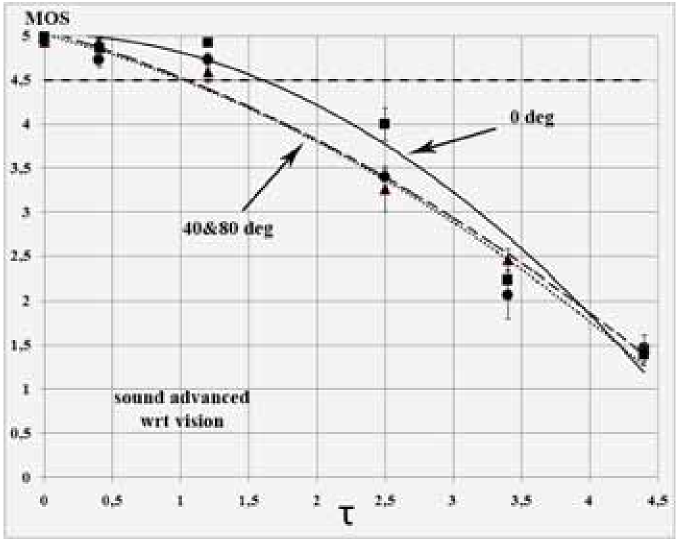


Fig. 7. The relationships between the relative time difference of audio and video components and the MOS for sound advanced for different angels for stereoscopic 3D images.

CONCLUSIONS

Degradations caused by audio-to-video delays errors are more noticeable for viewers in 3DTV than in 2DTV. The threshold of detectability when sound leading video for 2D images is about +50 ms and +27 ms for 3D images. The results of quality assessments prove that viewing public requires more accurate lip synchronization if television programs' viewers watch stereoscopic 3DTV medium close shot, half-turn position of newsreader (rotation by 40 degrees about an axis).

REFERENCES

[1] Relative Timing of Sound and Vision for Broadcasting. Recommendation ITU-R BT.1359-1.

[2] Recommendation ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures.

[3] EBU Technical Recommendation R37-2002: The relative timing of the sound and vision components of a television signal, the European Broadcast Union (2002).

[4] NVIDIA® 3D Vision™ technology. http://www.nvidia.com/object/3d-vision-main.html

[5] Stereoscopic Player.http://3dtv.at/Products/Player/Index_en.aspx

[6] T. Kuratate, K. Munhall, and P. Rubin, "Audio-Visual Synthesis of Talking Faces from Speech Production Correlation," EuroSpeech'99 Publication.

Memory Reduction Method of Luminance Compensation Algorithm for Mobile AMOLED Display Applications

Kyonghwan Oh, *Student Member, IEEE,* Nack-Hyun Keum, Oh-Kyong Kwon, *Member, IEEE,*
Department of Electronic Engineering, Hanyang University, Korea (e-mail: okwon@hanyang.ac.kr)

***Abstract*-- A memory reduction method of luminance compensation algorithm using luminance sensor is proposed for high resolution mobile active matrix organic light emitting diode (AMOLED) display applications. The proposed method can be applied to AMOLED panels which have poor luminance uniformity even though compensation pixel circuit is used. Measurement results show that the deviation of luminance error is reduced from 18.1% to 3.8% when the proposed algorithm is applied to 5.3-inch AMOLED panel using threshold voltage compensation pixel structure.**

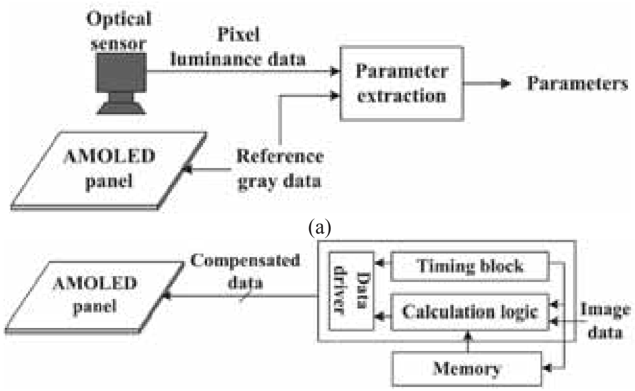
I. INTRODUCTION

The luminance non-uniformity of active matrix organic light emitting diode (AMOLED) display due to electrical variation of polycrystalline silicon thin film transistors (TFTs) is main problem to achieve high image quality display [1]. Previously, pixel structures and driving methods to compensate electrical variation of driving TFTs have been reported [2-4], but compensation capability is reduced as the display resolution is increased because the row line time which is used for compensation phase is reduced as the resolution of panel is increased. To solve the problem of limited compensation time, the parameter extraction and data adjusting method using light sensor have been proposed [5]. This method assumed that the relation of luminance and data voltage of AMOLED pixel is quadratic [5], however drain current of driving TFT is not quadratic function thus luminance error will be increased due to the large difference between modeling and real characteristic in low gray levels.

To improve compensation capability, a new model equation between luminance and data voltage is required. However, accurate model increases complexity of calculation logic block for data modulation and required memory due to extended the number of parameters. To solve this problem, a new compensation algorithm is proposed to simplify calculation logic block and reduce required memory. The proposed algorithm is verified by 5.3-inch AMOLED display panel of 800×1280 resolution format.

II. PROPOSED DRIVING SYSTEM

The proposed compensation algorithm consists of two steps, which are parameter extraction with luminance sensing step and displaying step with data modulation for compensation. Fig. 1(a) and 1(b) show the block diagram of parameter extraction with luminance sensing step and displaying step using data modulation.



(a)

(b)

Fig. 1. Block diagram of (a) parameter extraction with luminance sensing step and (b) displaying step using data modulation.

In the parameter extraction with luminance sensing step, same reference gray data is programmed to entire pixels and optical sensor measures the luminance of every pixel~~s~~. The parameter extraction block receives pixel luminance data and reference gray data, and extracts parameters to characterize luminance-gray data relation of each pixel. The extracted parameters are stored to external memory. This step is required to be performed only one time because once extracted parameters are stored to non-volatile memory and they can be used in displaying step with data modulation.

In the displaying step with data modulation, calculation logic block receives original image data and the extracted parameters from external memory and computes compensated data. The data driver programs the compensated data to each pixel.

III. COMPENSATION ALGORITHM

During the parameter extraction with luminance sensing step, the luminance-gray data of AMOLED pixel is modeled as,

$$L = \alpha\left(V_{data} - \beta\right)^{\gamma}, \qquad (1)$$

where L, V_{data}, α, β, and γ are output luminance, input gray level, slope factor, shift factor, and curvature factor of AMOLED pixel, respectively. In this model, α depends on channel mobility of driving TFT and efficiency of OLED, and β depends on the threshold voltage of driving TFT. The parameter γ depends on exponent factor of saturation current of driving TFT and curvature of luminance-current relation of OLED, and gamma correction for human eye perception. Equation (1) is the general model of voltage programming current source (VPSC) type AMOLED pixels, thus the proposed algorithm can be applied to any VPCS AMOLED

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

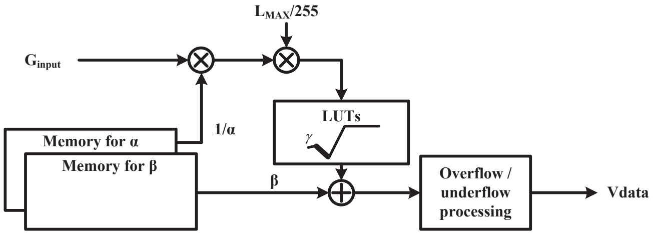


Fig. 2. block diagram of calculation logic.

panel.

To extract α, β, and γ parameters, more than 3 points of luminance of each pixel is measured, and nonlinear least square algorithm is used.

The modulated data can be expressed as,

$$G_{mod} = \beta + \sqrt[\gamma]{\frac{L_{MAX}}{\alpha}\left(\frac{G_{input}}{255}\right)}, \qquad (2)$$

where G_{mod}, G_{input}, and L_{MAX} are compensated gray data, original input gray level, and maximum target luminance, respectively. Then, the luminance of compensated pixel (L_{mod}) can be expressed as,

$$L_{mod} = \alpha\left(\beta + \sqrt[\gamma]{\frac{L_{MAX}}{\alpha}\left(\frac{G_{input}}{255}\right)} - \beta\right)^{\gamma} = L_{MAX}\left(\frac{G_{input}}{255}\right). \quad (3)$$

Equation (3) shows that the compensated luminance only depends on input gray level and maximum target luminance.

The computation of (2) should be performed in calculation logic block as shown in Fig. 1(b) however the second term of (2) is hard to realize simple arithmetic operation. To accomplish this operation, high speed digital signal processor in calculation block is required, but it increases system cost and chip size. This disadvantage makes hard to apply the proposed method to mobile AMOLED applications.

To solve this problem, γ parameter of every pixel is averaged then the second term of (2) can be realized in simple arithmetic operation and look-up-table. Fig. 2 shows the block diagram of calculation logic. However, averaged γ increases fitting error between sensed luminance data and fitted luminance curve, (1), because α and β parameters are optimized from raw γ parameters. To reduce fitting error, the second fitting is performed with respect to α and β. Eventually, the required memory is (the number of pixels) × 3(red, green, and blue) × 2(α and β) × 8bit(bit depth of gray scale).

Therefore, required external memory to store pixel parameters can be reduced and simple arithmetic operation is required by using averaged γ parameter.

IV. Measurement Results

The proposed luminance compensation algorithm is applied to 5.3-inch 800×1280 resolution format AMOLED display using polycrystalline silicon TFT backplane. The fabricated AMOLED panel uses threshold voltage compensation pixel structure [4]. The luminance sensing for parameter extraction in Fig. 1(a) is performed using 6 points of luminance intensity. The computation for nonlinear least square algorithm and generation of look-up-table are executed using a PC.

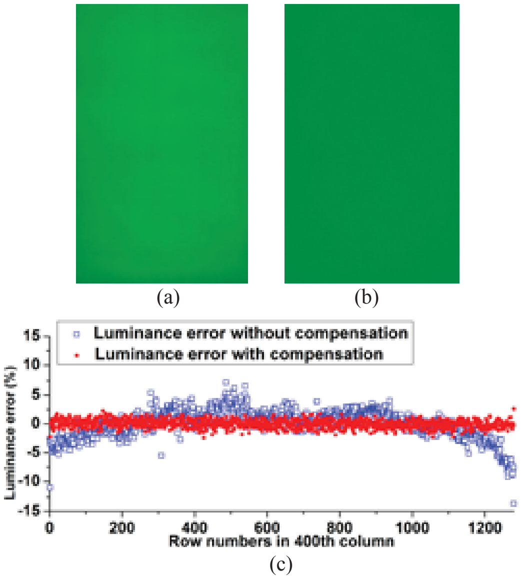


Fig. 3. Measurement result (a) without compensation method, (b) with compensation method of luminance in middle gray level (127 gray), and (c) luminance error on 400th column line.

Fig. 3(a), (b), and (c) show measured results of AMOLED panel without the proposed algorithm, with the proposed method, and luminance error on 400th column line in 127 gray level using self-compensation pixel circuits, respectively. The measured results show that the luminance error deviation is 18.1% without the proposed algorithm, and it is reduced to 3.8% with applying the proposed algorithm.

V. Conclusion

In this paper, a compensation algorithm to improve luminance uniformity of AMOLED display by using optical sensor and data modulation logic is proposed. The proposed algorithm can reduce required external memory size and simplify calculation logic block. Experimental results show that the deviation of luminance error is achieved to 3.8% whereas the luminance error deviation 18.1% without the proposed method in 127 gray level using the panel with self-compensation pixel circuit.

References

[1] X. Guo and S. R. P. Silva, “Investigation on the current nonuniformity in current-mode TFT active-matrix display pixel circuitry,” *IEEE Trans. Electron Devices*, vol. 52, no. 11, pp. 2379–2385, Nov. 2005.

[2] S. -H. Jung, W. -J. Nam, and M. K. Han, “A New Voltage-Modulated AMOLED Pixel Design Compensating for Threshold Voltage Variation in Poly-Si TFTs,” *IEEE Electron Device Lett*., vol. 25, no. 10, pp. 690-692, Oct. 2004.

[3] H. -Y. Lu, P. -T. Liu, T. -C. Chang, and S. Chi, “Enhancement of Brightness Uniformity by a New Voltage-Modulated Pixel Design for AMOLED Displays,” *IEEE Electron Device Lett*., vol. 27, no. 9, pp. 743-745, Sep. 2006.

[4] S. -M. Choi, O. -K. Kwon, N. Komiya, and H. -K. Chung, “A self-compensated Voltage Programming Pixel Structure for Active-Matrix Organic Light Emitting Diodes,” *Proc. Int. Display Workshop,* pp. 535-538, Dec., 2003.

[5] H. -J. In, K. -H. Oh, O. -K. Kwon, C. -H. Hyun, and S. -C. Kim, “Luminance Adjusting Algorithm for High Resolution and High Image Quality AMOLED Displays of Mobile Phone Applications,” *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1191-1195, Aug., 2010.

Edge Connectivity-Based Image Denoising for Digital TV Systems

Sung In Cho, *Student Member, IEEE* and Young Hwan Kim, *Member, IEEE*
Division of Electrical and Computer Engineering, Pohang University of Science and Technology

***Abstract*— This paper presents a new approach to denoising images that uses edge connectivity information. Using this information, the proposed algorithm calculates the amount of image details accurately and performs connected component-based noise filtering.**

I. Introduction

In video systems, image noise severely degrades image quality. Hence, image denoising, which refers to remove noise from images, is a very important part of image processing tasks for video systems. The goal of image denoising is to remove as much noise as possible with minimum loss of image detail. To achieve this goal, a variety of image denoising algorithms have been developed. Bilateral filtering [1] and non-local (NL) means filtering [2] are widely used for image denoising, because they can be easily implemented and they possess good noise removal capability. Although these algorithms have been successfully used, there is still scope for improvement in terms of image quality. Recently, a block matching and three-dimensional (3-D) (BM3D) filtering algorithm [3] showed outstanding performance in terms of image quality. However, this algorithm requires massive hardware resources and also involves very high computational complexity. Hence, it is inappropriate for use for digital TV systems, which require real-time processing capability and small amount of hardware resources.

In this paper, we propose a new image denoising algorithm to be used for digital TV systems that uses edge connectivity to effectively remove noise from a single image while consuming a small amount of hardware resources as compared to conventional image denoising algorithms. In our evaluation of the performance of the proposed algorithm, we used additive white Gaussian noise (AWGN) as the image noise model.

II. Proposed image denoising Algorithm

Fig. 1 illustrates the process of the proposed image denoising algorithm. The proposed algorithm consists of two parts: detail descriptor generation and connected component-based noise filtering. First, the detail descriptor is generated using a binary edge map and an eight-connected component of edges. Subsequently, connected component-based noise filtering is adaptively performed depending on the detail descriptor.

A. Detail descriptor generation

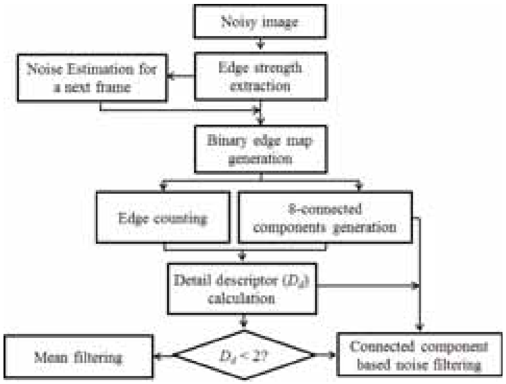

Noisy image
Noise Estimation for a next frame
Edge strength extraction
Binary edge map generation
Edge counting
8-connected components generation
Detail descriptor (D_d) calculation
Mean filtering
$D_d < 2?$
Connected component based noise filtering

Fig. 1. Process of the proposed image denoising algorithm

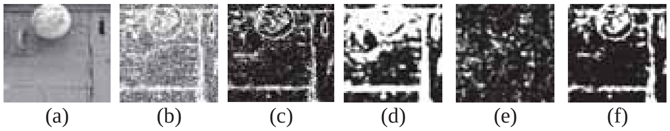


Fig. 2. (a) input image, (b) *ES*, (c) E_{binary}, (d) N_{edge}, (e) $N_{component}$, (f) D_d

In this section, the edge strength (*ES*) (Fig. 2(b)) is calculated using Sobel operators. Next, a binary edge map (E_{binary}) (Fig. 2(c)) is generated by classifying the *ES* values into 0 and 1 according to the threshold value (TH_{ES}) as shown below:

$$E_{binary}(i,j) = \begin{cases} 1, & ES(i,j) > TH_{ES} \\ 0, & otherwise \end{cases}, \tag{1}$$

where *i* and *j* denote the index of the pixel being processed, and TH_{ES} is the expectation of *ES* at a specific noise level. The noise level can be estimated from noisy smooth patches in the previous frame. E_{binary} can be used to calculate the detail descriptor ($D_d(i,j)$) of pixel (i,j), as follows.

$$D_d(i,j) = N_{edge}(i,j) / N_{component}(i,j), \tag{2}$$

$$N_{edge}(i,j) = \sum_{y=-adj}^{adj} \sum_{x=-adj}^{adj} E_{binary}(i+x, j+y), \tag{3}$$

where $N_{edge}(i,j)$ (Fig. 2(d)), and $N_{component}(i,j)$ (Fig. 2(e)) denote the number of ones and the number of connected components in an E_{binary} patch(i,j), whose center is in the pixel (i,j) and size is 9 × 9 pixels. *adj* denotes the range of the patch and is set to 4. In general, noise components have poor connectivity, whereas image details have strong connectivity. Hence, $D_d(i,j)$ (Fig. 2(f)), which denotes the average pixel numbers per connected component, can represent the amount of image details in a local image region very accurately.

This work was supported by the Brain Korea 21 project in 2012, IDEC, and LG Display.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Connected component-based noise filtering

Next, noise filtering is performed adaptively depending on the generated detail descriptor. If $D_d(i,j)$ is equal to or less than 1, pixel (i,j) is classified to be in the completely noisy smooth region, and mean filtering is performed. Otherwise, connected component-based noise filtering is performed as follows:

$$G(i,j) = \frac{1}{\left(Z(i,j)\right)} \sum_{x=-adj}^{adj} \sum_{y=-adj}^{adj} w(x,y) \cdot G\left(i+x, j+y\right), \tag{4}$$

$$Z = \sum_{x=-adj}^{adj} \sum_{y=-adj}^{adj} w(x,y), \tag{5}$$

where Z is the normalized factor and weight $w(x,y)$ is extracted by

$$w(x,y) = \begin{cases} e^{-\alpha \cdot Dif/D_d(i,j)}, & if\ label_{p(x,y)} = label_{p(i,j)} \\ \left(e^{-\alpha \cdot Dif/D_d(i,j)}\right)/8, & otherwise \end{cases} \tag{6}$$

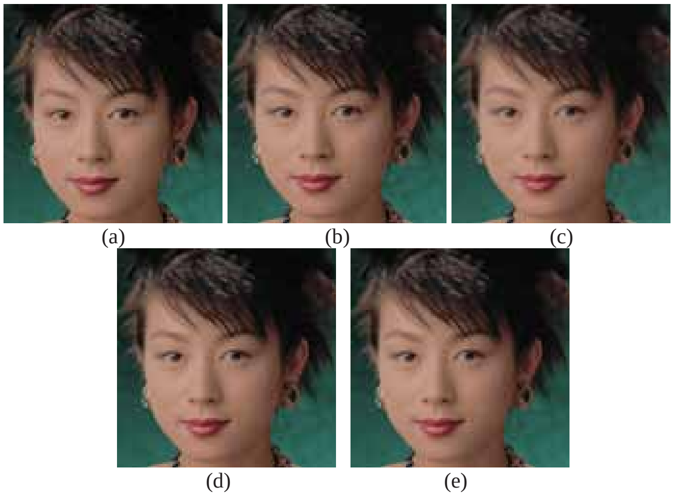
$$Dif\,(x,y) = \left|G(i,j) - G(x,y)\right|, \tag{7}$$

where $label_{p(\cdot,\cdot)}$ denotes the label number of a connected component of pixel $(\cdot,\cdot)$, and Dif denotes the similarity between pixel (i,j) and its neighboring pixels. α denotes the experimental smoothing factor varied with the noise level. In (6), depending on the amount of image details, the smoothing strength is adjusted by dividing $-Dif$ by $D_d(i,j)$. If the pixels are in the same connected component, they have similar characteristics. Hence, if a neighboring pixel has the same label number as pixel (i,j) and has a similar gray level, a large weight value is assigned.

III. EXPERIMENTAL RESULTS

We evaluated the performance of the proposed algorithm in terms of image quality using the peak signal to noise ratio (PSNR) and the structural similarity (SSIM) [4]. For the test image, we used a Kodak, ISO, and IEC image set that was captured by IEC video. The images were degraded by AWGN with 3%, 5%, and 10% standard deviations. As the benchmark algorithms, we used the bilateral and NL means filters. Some algorithms, such as BM3D, can provide better results than the benchmark algorithms, but these algorithms cannot be implemented for real-time processing with small amount of hardware resources. Given that our algorithm was directed at real-time processing with small amount of hardware resources, the benchmark algorithms were appropriate for comparison with the proposed algorithm. In the comparison, the parameters of each benchmark algorithm are adjusted to deliver the highest values of PSNR and SSIM for each noise level and to use the same size of line buffer.

As shown in Table I, the proposed algorithm has the highest PSNR and SSIM values at various noise levels. This implies that the proposed algorithm removed noise most effectively. The subjective evaluation showed that the proposed algorithm yielded good image quality compared with the benchmark algorithms, as shown in Fig. 3.


(a) (b) (c)

(d) (e)

Fig. 3. Image comparison of the conventional and proposed methods: (a) original image, (b) degraded image, (c) image by the bilateral filtering, (d) image by the NL means filtering, (e) image by the proposed algorithm.

TABLE I
AVERAGE PSNRS AND SSIMS OF PROPOSED AND BENCHMARK ALGORITHMS

Test images (The number of images)		Noisy image		Bilateral		NL means		Proposed	
		PSNR [dB]	SSIM	PSNR [dB]	SSIM	PSNR [dB]	SSIM	PSNR [dB]	SSIM
Kodak (24)	σ=3%	33.776	0.880	36.033	0.897	**37.016**	0.943	36.953	**0.945**
	σ=5%	30.845	0.747	34.120	0.825	34.316	0.894	**34.788**	**0.900**
	σ=10%	28.796	0.492	31.457	0.707	32.601	0.804	**32.754**	**0.805**
ISO (15)	σ=3%	34.304	0.864	38.960	0.956	38.172	0.956	**39.049**	**0.970**
	σ=5%	31.412	0.722	36.131	0.920	35.775	0.925	**36.388**	**0.946**
	σ=10%	29.385	0.46	32.224	0.774	33.500	0.881	**33.517**	**0.887**
IEC (20)	σ=3%	33.993	0.842	39.076	0.939	38.538	0.948	**39.408**	**0.960**
	σ=5%	31.090	0.679	37.205	0.905	35.919	0.904	**37.418**	**0.934**
	σ=10%	29.064	0.398	32.631	0.719	34.535	0.858	**35.084**	**0.874**

IV. CONCLUSION

In this paper, we proposed a new image denoising algorithm that uses edge connectivity information to separate image detail components from noise components. Using these separated image details, the proposed algorithm estimates accurately the amount of image details in a local image region. Based on this amount of image details, the algorithm subsequently performs connected component-based noise filtering. In the experiments, the average PSNR and SSIM of the proposed algorithm were up to 1.50 dB and 0.03, respectively, higher than that of the benchmark algorithms.

REFERENCES

[1] C. Tomasi and R. Manduchi “Bilateral filtering for gray and color images,” *Proc. IEEE Int"l Conf. Computer Vision*, pp. 839-846, Jan. 1998.

[2] A. Buades, B. Coll and J.-M. Morel “A non-local algorithm for image denoising,” *IEEE Computer Vision and Pattern Recognition*, 2005.

[3] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian “BM3D image denoising with shape-adaptive principal component analysis,” *Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[4] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, Morgan & Claypool, 2006, ch. 1.

Adaptively Partitioned Block-Based Backlit Image Enhancement for Consumer Mobile Devices

Nahyun Kim[1], Seungwon Lee[1], Ewoo Chon[2], *Member, IEEE,* Monson H. Hayes[1], *Fellow, IEEE,* and
Joonki Paik[1], *Member, IEEE*
[1]Image Processing and Intelligent Systems Laboratory, Graduate School of Advanced Imaging
Science, Multimedia, and Film, Chung-Ang University, Seoul, Korea
[2]1st R&D Center/CP group, Nextchip Co., Ltd., Seoul, Korea

***Abstract*-- In this paper, we present an efficient backlit region detection and enhancement method with low-computational cost for consumer mobile imaging devices. The proposed approach partitions an input image into multiple rectangular blocks, and uses a fuzzy logic-based optimal threshold to classify each block as backlit, background, or ambiguous. Backlit blocks are then processed using a contrast enhancement algorithm, whereas those block that are classified as ambiguous are further partitioned into four subblocks, and each subblock goes through further classification. The major advantage of the proposed method is computational efficiency and improved quality with minimized blocking artifacts compared with other block-based approaches.**

I. Introduction

As digital and mobile cameras are becoming popular in the consumer market, quality issues are raised in many consumer photos. Backlit images fall into one of the quality degradation issues because important objects may disappear in the backlit region if the illumination is not carefully controlled. Conventional backlit image enhancement methods include histogram equalization [1] and Retinex methods [2]. Spatially adaptive approaches have also been proposed, but as with the conventional methods, they are subject to blocking artifacts that degrade the quality of the image. The bi-histogram equalization enhances the backlit image while preserving the brightness of the entire image, but it is difficult to separate background and object regions because the threshold that bisects the entire histogram is determined using the simple average intensity of the image [3]. Retinex-based methods can enhance backlit images by reducing the illumination effect and stretching the contrast of reflectance, but its high computational cost and color distorting artifacts make its use in consumer applications difficult.

In this digest, we present an efficient backlit image enhancement algorithm using adaptively partitioned block-based backlit region detection. The proposed method classifies partitioned blocks as either backlit, background, or ambiguous blocks using fuzzy logic-based thresholds. The amount of backlighting is then estimated within each backlit block, and the ambiguous blocks are further partitioned into four subblocks and reclassified. After the adaptive block partitioning and classification process has been completed, a contrast enhancement algorithm is used to enhance the backlit blocks. Experimental results show that the proposed method can efficiently enhance the contrast of backlit images using adaptively partitioned blocks for fast backlit region detection without the use of a complicated segmentation algorithm.

II. Adaptively partitioned block-based backlit image enhancement

The proposed backlit image enhancement algorithm first converts the input color image into the HSV (Hue-Saturation-Value) color space. The V channel image is partitioned into rectangular blocks, and each block is then classified as either a backlit block, a background block, or a block that is ambiguous. The classification of a block is performed by applying thresholds to the largest and smallest intensities within the block. These thresholds are determined using a using fuzzy c-means (FCM) clustering algorithm. More specifically, let $X = \left[x_1 \mid x_2 \mid \ldots \mid x_N\right]$ be a $p \times N$ data matrix, where x_j is the j-th feature vector. The FCM algorithm then finds a set of cluster means by values, c_j, that minimize the quadratic error

$$J_{FCM} = \sum_{i=1}^{N} \sum_{j=1}^{C} \left(u_{ij}\right)^m \left\| x_i - c_j \right\|^2, \tag{1}$$

Here, N is the number of patterns, C is the number of clusters, , and m is a weighting exponent on each fuzzy membership [4]. In this work, C =2 because there are only two clusters, backlit and background. After FCM clustering, the two clustering centers c_1 and c_2 are given by

$$c_j = \frac{\sum_{i=1}^{N} (u_{ij})^m x_i}{\sum_{i=1}^{N} (u_{ij})^m}, \; j = 1, 2. \tag{2}$$

Each block is then classified and given a label as follows:

This research was supported by Basic Science Research Program through National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (2009-0081059), by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2011 (00045420-1), and by the Technology Innovation Program, Industrial Strategic Technology Development Program(Development of SR(Super Resolution) Image Scaler for 4K UHD), under Grant K10041900.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$B = \begin{cases} 2, \ \mathrm{b}_{\max} \le T_L \\ 1, \ T_H \le \mathrm{b}_{\min} \\ 0, \ otherwise \end{cases} , \tag{3}$$

where the label B=2 is used to indicate a backlit block, B=1 represents a background block, and B=0 corresponds to an ambiguous block. The variables $b_{\max}$ and $b_{\min}$ represent the maximum and minimum intensity values, respectively, of the block, and T_H and T_L are the thresholds that are used to distinguish between backlit and background regions. Those blocks that do not satisfy the conditions for backlit or background regions are classified as ambiguous, and these blocks are partitioned into four subblocks, and each of these subblocks are again classified using (3). This process is continued until all of the blocks are unambiguously classified.

Fig. 1 shows the result of backlit region detection, and Fig. 2 shows the step-by-step process of adaptive block partitioning. After completing adaptive block partitioning, the contrast of a backlit block is corrected using a median filter with adaptive weighting. As shown in Fig. 2(d), there is no blocking artifacts in the final enhanced image.

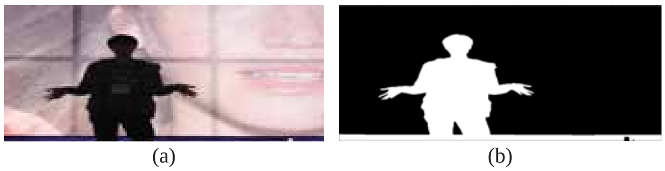
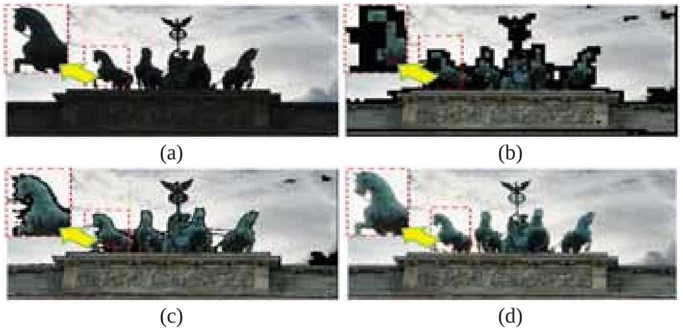

(a) (b)

Fig. 1. The result of the proposed backlit region detection; (a) input image and (b) the detected backlit region.

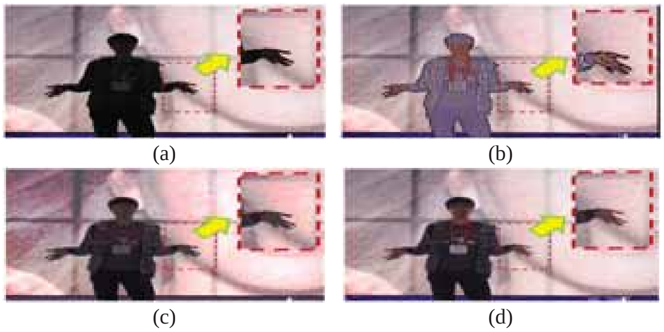

(a) (b)

(c) (d)

Fig. 2. The results of adaptive block partitioning; (a) input image, (b) result of the first partitioning into 64×64 blocks, (c) result of the second partitioning into 32×32 blocks, and (d) the finally enhanced backlit image

III. Experimental Result

To test our backlit image enhancement algorithm, we captured test images of size 1280x1024 that have backlit regions, and the performance of the proposed method was compared with adaptive histogram equalization (AHE) and bi-histogram equalization (BHE) methods.


(a) (b)

(c) (d)

Fig. 3. Experimental results of three different proposed methods; (a) input image, (b) result of AHE, (c) result of BHE, and (d) result of the proposed method.

Fig. 3 compares experimental results of three different backlit enhancement methods. Figs. 3(b) and (c) show results of existing AHE and BHE methods, respectively, and both result in unnatural saturation in the background. On the other hand the proposed method is able to enhance the backlit region without saturation in the background region.

IV. Conclusion

In this digest, we presented an efficient backlit region detection and enhancement method using a very low computational cost for consumer mobile imaging devices. The proposed method partitions an input image into regular blocks. Optimal thresholds for classifying backlit regions were adaptively selected using the FCM clustering algorithm. We then classify each block as being either a backlit or background block, or a block that is ambiguous. The enhancement of the backlit regions is performed by correcting the contrast using a median filter with adaptive weights. Experimental results show that the proposed method can efficiently enhance the backlit image without the use of complicated segmentation methods.

Reference

[1] J. Zimmerman, S. Pizer, E. Staab, J. Perry, W. McCartney, and B. Brenton, “An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement,” IEEE Trans. Medical Imaging, vol. 7, no. 4, pp. 304-312.

[2] J. Bai, T. Nakahuchi, N. Tsumura, and Y. Miyake, “Evaluation of image corrected by retinex method based on S-CIELAB and gazing information”, IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences, vol. E89-A, no. 11, pp. 2955-2961, 2006.

[3] Y. Kim, “Contrast enhancement using brightness preserving bi-histogram equalization,” IEEE Trans. Consumer Electronics, vol. 43, no. 1, pp. 1-8, 1997.

[4] S. Shen, W. Sandham, M. Granat and A. Sterr, “MRI Fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization.” IEEE Trans. information technology in biomedicine, vol. 9, no. 3, PP.459-467, September 2005.

A Feedback ANC Based Voice Enhancing Earmuffs System

Seong-Pil Moon, Tae-Ho Roh and Tae-Gyu Chang, *Senior Member, IEEE*

***Abstract*-This paper proposes a voice enhancing earmuffs system which is based on feedback active noise control. The performance of the proposed noise canceling algorithm is analytically derived. The noise reduction performance of the earmuff system is also experimentally tested to show its feasibility.**

I. INTRODUCTION

Passive earmuffs are widely used in noisy environment such as industry plants, construction sites and manufacturing factories. Simultaneous suppression of conversational voice sound is one of the big troubles of adopting the passive earmuffs.

In this paper, an active earmuffs system is proposed, where background noise is selectively suppressed while conversational voices are preserved as much by applying the feedback active noise control (ANC) algorithm. The earmuffs system is a headphones-based structure where the feedback ANC generates speaker sounds to acoustically cancel the undesired background noise.

Design of the feedback ANC requires comprehensive understanding of the effects of physical parameters including the two key parameters, i.e., the delay of electro-acoustic coupling path and the bandwidth of background noise. Performance of the proposed earmuffs system is analytically derived to examine the effects of the two design parameters and to provide an optimized design guide. To show the feasibility of the proposed active earmuffs system, an experimental earmuffs system is implemented and its performance is evaluated.

II. Feedback Active Noise Control Based Voice Enhancing Earmuffs System

A block diagram of the proposed earmuffs system is illustrated in Fig. 1, where the feedback ANC algorithm produces the anti-noise sound through the headphone-speaker and cancels the undesired sound (primary noise) coming into the earmuffs. The residual sound is sensed by the error microphone, and this sensed error signal is denoted by $e(n)$.

The electro-acoustic coupling path is defined by the path starting from $y(n)$ to $e(n)$. The path includes not only the acoustic coupling between the loud speaker and the microphone but also the electric dynamics for D/A converter, power amplifier, loud speaker, microphone, microphone preamp and A/D converter. This electro-acoustic coupling path is modeled as a transfer function $\hat{S}(z)$, which is called the secondary path estimation.

In the feedback ANC algorithm, the reference signal $x(n)$ is internally generated through the feedback path based on the error signal $e(n)$ and the filtered anti-noise signal $\hat{y}'(n)$ as

$$x(n) = \hat{y}'(n) + e(n) = s(n) * y(n) + e(n). \qquad (1)$$

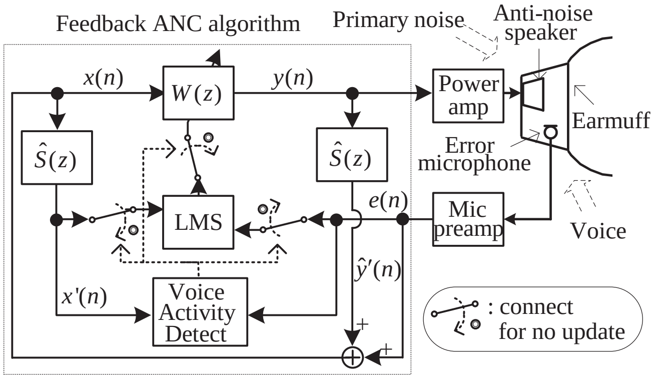


Fig. 1. Proposed FBANC based earmuffs system.

The adaptive filter $W(z)$ is updated by the normalized least-mean-square (LMS) algorithm based on the error signal $e(n)$ and filtered reference signal $x'(n)$ [3,4].

Selective suppression of the background noise is necessary for enhancement of voices. When voices are added to the background noise, the adaptive filter tries to tune itself to suppress the voices together with the background noise. The wrong update of the filter can be prohibited by blocking the update during the voice activity period as detected by a separate detection module. In this way, the adaptive filter can stay tuned to cancel only the background noise, being not disturbed by the occurrence of voices.

III. Performance Analysis Of The Earmuffs System

In this section, noise reduction performance of the proposed noise canceling algorithm is analytically derived in relation to the noise bandwidth and the secondary path delay.

To exclusively investigate the effects of the two parameters, noise and secondary path are modeled with a first order AR process and a Δ-sample pure delay, respectively.

Under the assumption that the estimation error in the pure delay secondary path model of $S(z) = z^{-\Delta}$ is negligibly small, the feedback ANC can be reduced to a Δ-step linear predictor [2]. Noise reduction performance of the FBANC can be obtained from its equivalent Δ-step linear predictor model, reflecting the noise bandwidth in terms of the AR model's parameters. A closed form equation of maximum available noise reduction is derived as the ratio between the average power of the primary noise $d(n)$ and that of the residual noise $e(n)$ as (2).

$$NR_{AR(1).\max} = \frac{1}{1-l^{2\Delta}} \qquad (2)$$

where Δ : delay of the secondary path,
l : AR coefficient, and
$1-|l|$: -3 dB noise bandwidth for $1-|l| \ll 1$.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

IV. Measurements and Results

The performance of the proposed earmuffs system is also experimentally tested to compare with the analytically derived results. For the experimental test, a headphones structured earmuffs system is implemented. A picture of the overall experimental setting is shown in Fig.2.

A dummy head wears the headphones, and through the dummy head, the sensing tip of the measurement microphone closely faces the headphone speaker. Primary noise is generated by a loud speaker located in front of the dummy head. The earmuffs algorithm is implemented on a floating-point TI TMS320C6x DSP system.

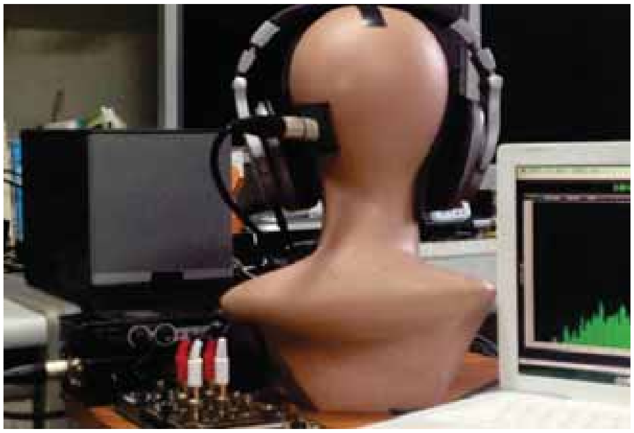


Fig. 2. Experimental setup of the FBANC based earmuffs system.

Primary noises are generated using AR models having three different -3dB bandwidths, i.e., 1Hz, 10Hz and 100Hz. The noises are sampled with the sampling frequency 96kHz and they are decimated to 16kHz for ANC operation. The same noise sound is applied two times to the earmuffs system, once without running ANC and once with running ANC. The microphone sensed residue signals are recoded using an independent computer. The noise reduction performance is measured by obtaining the ratio of the average powers of the two recoded residues.

The experimental results are summarized in TABLE I together with the analytic results. In the derivation of the analytic results, the excessive mean-square error of the LMS algorithm is added to the noise reduction equation (2). In this example, the secondary path delay is set to 0.25msec.

It is verified from the TABLE I that the analytic results and the experimental results agree very well. The relatively high difference, i.e., 3.76 dB, shown for the noise having 100 Hz bandwidth is caused by the magnitude distortion effect of the secondary path in the real experimental setup, which is not reflected in the pure delay secondary path model in analytic result (2).

Voice signals are added to the background noise signal having the bandwidth 10Hz. Update of the adaptive filter is selectively enabled to avoid voice talking periods as described in section II. The experimentally measured results are shown in Fig.3. The power spectrum of the voice added noise is plotted with the power spectrums obtained from the applications of two different ANC operations, i.e., the continuous update and the selective update.

TABLE. I. Noise Reduction Performance of The Proposed Earmuffs System

	3-dB Bandwidth of AR noise		
	1 Hz	10 Hz	100 Hz
Analytic results	24.67 [dB]	17.16 [dB]	7.23 [dB]
Experimental results	23.43 [dB]	17.47 [dB]	10.99 [dB]

From Fig.3, it is shown that the peak noise reduction achieved by the application of the ANC is more than 20dB. It is also verified that the proposed selective update of ANC filter significantly enhances voices as can be inferred from the noticeable enhancement of power spectrum in the frequency range between 400-800 Hz where most of the speech energy is distributed.

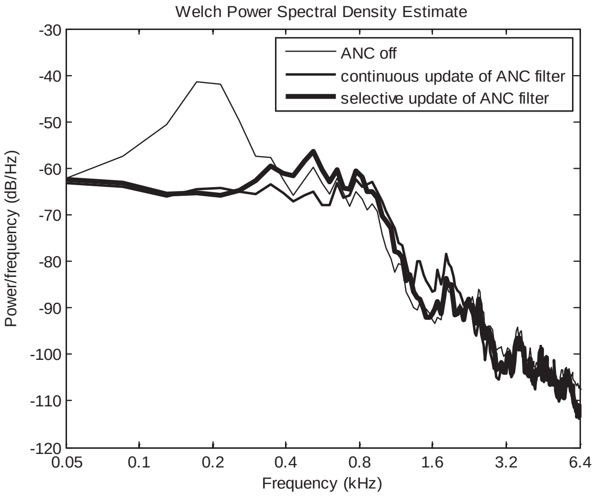


Fig. 3. Comparison of noise reductions for voice added background noise. The proposed ANC method shows the significant enhancement of power spectrum in the frequency range between 400-800 Hz where most of the speech energy is distributed.

V. Conclusion

This paper presents a FBANC based earmuffs system which selectively reduces background noise while preserving the clarity of voice. The noise reduction performance of the proposed noise canceling algorithm is analytically derived and experimentally tested as well, and their agreements are verified.

The feasibility of the proposed earmuffs system is shown by the experimental test, where the achieved levels of noise reduction for the three different noise bandwidths cases, i.e., 1Hz, 10Hz, and 100Hz, are high to show 23dB, 17dB and 11dB, respectively.

VI. References

[1] S. Vanit-Anunchai, "An implementation of active noise control for headphone using DSK TMS320VC5402," *Suranaree Journal of Science and Technology*, vol. 10, no. 4, pp. 266-274, Oct.-Dec. 2003.

[2] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems—Algorithms and DSP implementations*, New York: Wiley, 1996.

[3] D. R. Morgan, "An analysis of multiple correlation cancellation loops with a filter in the auxiliary path," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-28, no. 4, pp. 454-467, Aug. 1980.

[4] W. S. Gan and S. M. Kuo, "An integrated audio active noise control headset," *IEEE Trans. Consumer Electronics*, vol. 48, no. 2, pp. 242-247, May 2002

The Improvement of Mobile Phone Voice Quality by Bone-Conduction Device

Hyung-Woo Park, A-Ra Khil and Myung-Jin Bae*
Information and Telecommunication Department, Soongsil University, Seoul, Republic of Korea

***Abstract*-- This paper proposes a new way of reducing noise and enhancing speech signals of mobile phones by installing bone-conduction speakers in ordinary mobile phones. With this new system, the noise from surrounding environments can be reduced and eventually the quality and clarity of voice signals coming out of mobile phones can be improved. The improved voice quality of mobile phones was confirmed by the experiment which measured frequency responses as well as emotional responses before and after the activation of the proposed system.**

I. INTRODUCTION

With the recent advances in development of information and communication technology, the mobile devices, such as mobile phones, PMPs (portable media players), and MP3Ps (MP3 players) have been widely used. People use their portable devices in various places and they just want to use their devices more freely, whenever and wherever, regardless of the noise condition. Environment noise coming out of relatively noisy places exceeds 80dBA in average. Since the mobile phone makers usually manufacture the products with the sound generation capacity for higher than 100dB, the users do not have any difficulty in communicating on mobile phones[1].

However, if the users are being exposed to the loud sound through mobile devices for a long time, they may suffer noise induced hearing loss. In this paper, we propose the way of reducing noise and enhancing speech signals of mobile devices, mobile phones in particular, by installing bone-conduction speakers in ordinary mobile phones[1].

This paper is structured as following. Chapter 2 investigates the usual environments for using mobile phones. Chapter 3 proposes the method for installing a bone-conduction speaker into mobile phones. Chapter 4 examines the results of the experiment to evaluate the improved performance of mobile phones after installing the proposed system. Chapter 5 provides the conclusion.

II. ENVIRONMENTS NOISE OF MOBILE PHONES

We analyzed noise from various environments of using mobile phones and compared them to each other[1]. In each place, the noise was measured 10 times for 5 minutes each. Table 1 shows the average noise level for each environment. As shown in Table 1, the difference in average noise level between the quiet environment and noisy environment is about 20dB.

TABLE 1. AVERAGE NOISE LEVEL OF EACH ENVIRONMENT

environment	Level(dB A)	environment	Level(dB A)
Subway platform	85.7	Shopping mall	78.3
In Bus	80.4	Parks	55.1
In Subway	80.9	In office	61.2
Road side	77.9		

Figure 1 illustrates the frequency spectrum measured in a noisy environment (in public transportation) in comparison to the one measured in a quiet environment (in an office).

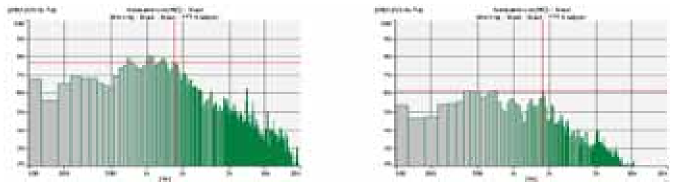


Fig. 1. Frequency response of noisy environment

III. MOBIL PHONE WITH BONE-CONDUCTION SYSTEM

The bone conduction system for mobile phones proposed in this paper is equipped with an additional bone-conduction speaker and an additional microphone, in order to receive and reduce the environment noise. Figure 2 is the block diagram which outlines the method proposed in this paper. An additional microphone receives the noise and then the noise passes through the bone conduction noise cancellation system. The bone-conduction speaker generates the vibration of anti-phase noise and eventually decreases the noise level. In this system, we first cancel the noise at cochlea by applying a destructive-interference with the anti-phase of environment noise, and then enhance voice signals by applying a constructive-interference with the in-phase of pre-emphasized signals. As the environment noise undergoes the continuous analysis and reduction process this way, the system actively reduces the noise.

In ordinary mobile phones, incoming voice is converted from electromagnetic signal to voice signal through the antenna and receiver, and then it is heard through the air coming out of speakers. Since the environment noise is transferred through the air, it travels to the opposite side of the ear, or it's received through the gap between the ear and a phone. Hence, voice signals are usually corrupted by noise and, to make the situation worse, louder noise makes voice signals unclear and further makes phone conversations uncomfortable.

* corresponding author : Myung-Jin Bae(mjbae@ssu.ac.kr)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

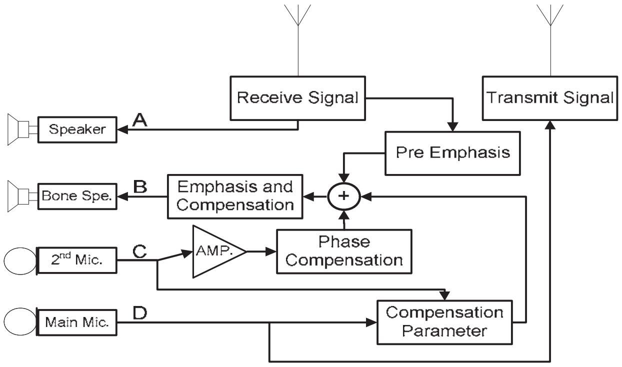


Fig. 2. Block diagram of proposed method

IV. Experiment and results

In order to evaluate the performance of the mobile phone installed with bone-conduction speaker as proposed in this paper, we conducted an experiment in the anechoic chamber where the background noise was 30dB. We set the 80 dB noise environment condition using monitor speakers. To evaluate the performance of the bone-conduction system, a vibration sensor and an additional microphone were attached to the phone. Firstly, we measured the frequency response and extracted the modeling parameter of bone-conduction compensation. Secondly, we tested the incoming sound quality of the mobile phone using the parameter of bone-conduction, and also evaluated the outcome of the proposed system with respect to emotional responses of the participants.

Figure 3 shows the performance of the phone before and after the activation of the proposed system. Frequency responses were measured by placing the mobile phone close to ears of the participants using Torso. As shown in Figure 3, the proposed system achieves the maximum attenuation of 19dB in noise interference at 700Hz.

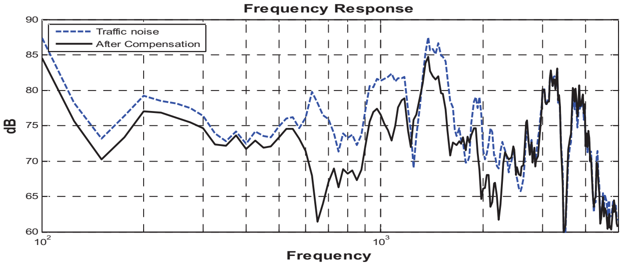


Fig. 3. Frequency response of before and after system in traffic noise

The emotional response evaluation proceeded under the same noisy condition of 80dB in the anechoic chamber. The difference in incoming voice levels of mobile phones were measured before and after the activation of the proposed system. Table 2 shows the results of emotional evaluation done by 20 participants. In noisy environments, participants set their average volume at 99.95dB for their mobile phones, but turned down the volume to 83.35dB with the proposed system. It means that the bone-conduction speaker system reduces 17dB of noise volume, consequently improving the voice quality by reducing the noise for 6dB more from the previous system proposed in our previous study[4]. It can be concluded that the bone-conduction speaker system is a better method for improving the voice quality of mobile phones by reducing noise more effectively.

TABLE 2 EXPERIMENT RESULT OF EMOTIONAL TEST

tester	before (dBA)	after (dBA)	tester	before (dBA)	after (dBA)
#1	99	85	#11	97	84
#2	102	84	#12	99	82
#3	98	74	#13	103	83
#4	97	82	#14	99	85
#5	99	83	#15	97	83
#6	103	89	#16	98	84
#7	101	83	#17	103	85
#8	99	81	#18	102	84
#9	98	83	#19	99	86
#10	101	84	#20	105	83
Average				99.95	83.35

V. Conclusions

Mobile phones can be used in a variety of environments. Since mobile phones are usually manufactured with the sound generation capacity of 100dB, users can use mobile phones even in noisy conditions. However, turning up the volume of phones in noisy environments causes noise-induced hearing loss to their users. In this paper, we proposed a new method to reduce noise effectively and enhance speech signals of mobile phones, by installing bone-conduction speakers.

In order to prove the improved performance of the new system, we installed the bone-conduction system in an ordinary mobile phone and conducted an experiment in order to analyze the frequency responses of the phone as well as emotional responses of participants before and after the activation of the proposed system. As confirmed in the results of the experiment, we succeeded in improving the voice quality of mobile phones by reducing noise for about 19dB at 700Hz. Emotional evaluation results also confirmed the improved performance of reducing noise for about 17dB.

In the future research, we will try to improve further the noise response rate of the bone-conduction speaker system and enhance its overall performances.

References

[1] H.W. Park, S.T. Lee and M.J. Bae, “A study on a SNR of an Available Telephone Conversation on Variable Noise Condition,” Proceedings of the ASK Conference, Vol.29 No.1(s), pp.36-37, 2010.

[2] H.W. Park, S.T. Lee and M.J. Bae, ″A Technique for Preventing Noise Induced Hear ing Loss Due to Mobile Phone Use Under Noisy Environment,″ The Journal of ASK, Vol.30, No.4, pp. 207-214, 2011.

[3] J.W. Kim and M.J. Bae, ″A Study on Hearing Loss According to Sound Pressure Level of The Ear-Phone″, KEE Transactions, Vol. 32, No. 1, pp.1086-1087, 2009.

[4] H. J. Kwon and M.J. Bae, “A Study on a Hearing Test to Measure Progress of Noise Induced Hearing Loss,” ASK, Vol.29, No.3, pp 184-190, 2009.

[5] M. J. Bae and S. H. Lee, *Digital Voice Signal Analysis*, Dong Young, Korea, 1998.

Multi-level Video Segmentation Using Visual Semantic Units

Huang-Chia Shih
Human-Computer Interaction Multimedia Lab
Department of Electrical Engineering, Yuan Ze University, Taiwan, R.O.C.

***Abstract*—This paper illustrates multiple levels of boundary for video segmentation from coarse to fine granularity, which include highlights, attentive visual change, and game status change. The boundaries between video shots are commonly known as scene change and the action of segmentation a video sequence into multiple shots is called scene change detection. However, different applications are suitable for different fine granularities of the shot boundary. In this paper, we take the content semantic into consideration to segment the video. The experimental results show the efficiency of the proposed method for sports programs.**

I. Introduction

The almost all applications in the first step video analysis is to segment the video into temporal "shot", which represents a partial period of video sequence that owns the same low-level or high-level properties. Video segmentation is a method to segment the video sequence into pieces of shots by determining the boundary. Many efforts have been made to analyze the content structure of video sequence with frame-indexed scores from the viewpoint of the temporal content similarity. Zhu *et al*. [1] proposed a scene segmentation and semantic representation framework for the efficient retrieval of video data by using many low-level and high-level features. Vlachos [2] employed the phase correlation to obtain a measure of content similarity for temporally adjacent frames and responds very well to simple cuts. In [3], a target distribution of the model parameters is constructed to model the probabilities for each video shot being declared as the scene boundaries, and the solution is achieved by performing the sampling from this target distribution using the Markov chain Monte Carlo (MCMC) technique.

II. Proposed Method

A shot is what captured by a camera between a record and a stop operation. Fig. 1 shows the schematic of video sequence with event and different granularities of content decomposition. For sports program, a number of events compose as a video sequence. An event includes the number of shots according to the category of event. It is possible to determine the event boundary by using the content analysis. In this paper, four visual semantic units of video are introduced. First, the attentive motion information provides the very important cue for the on-going game status. Second, the logo transition which is the special scene transition sandwiched between replays and highlights. Third, the outcome statistics in sports video are usually shown on the screen as superimposed caption box (SCB). Finally, the contextual information should be obtained. When the context of the SCB is changed, it implies the game status also changed.

A. Extraction of the Visual Semantic Units

1) Attentive Motion

Here, we illustrate a 1-D intensity projection method to represent the attentive motion. We divide the projection curve into small slices consisting of *N* values. Assume frame A is connected with frame B, now taking a slice of frame A and sliding it over the projection curve obtained from frame B. To calculate the *sum of absolute difference (SAD)* value for an offset of A. The displacement curve can be obtained by using the offset value, which used to determine the motion types such as static, zoom, pan, and tilt.

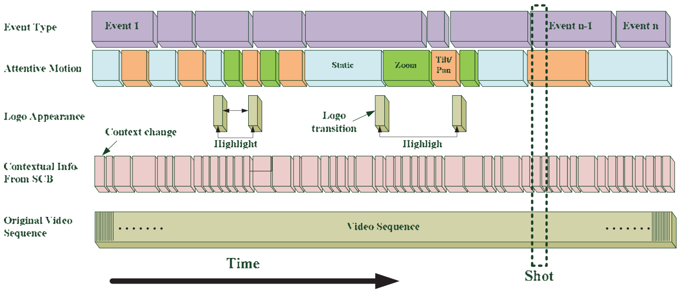


Fig. 1. The semantic units of a video sequence with event and contextual relations.

2) Logo Appearance

Because the temporal transitions between logo and replay are always large, and with special editing effects are applied to the logos, we call these transitions as "logo transitions". We compute the Hue and Intensity differences between two consecutive frames. Logo transitions remain the same during the entire game, hence we detect the logo transitions and the associated replay segments. A logo transition, usually less than one second long composes a set of consecutive frames that contain logo images of the broadcaster or a special event.

3) The superimposed caption box (SCB)

The SCB in sports videos is always stationary or sometimes may be locally dynamic instead of perceivable varying. Here, we combine color-based local dynamic and temporal motion consistency to locate the SCB from a group of frames (GoF).

4) Context Change

Since the SCB template does not change much during the entire video, we may use the SCB color model to identify its presence even if the SCB may be transparent. The similarity measure between the pre-stored SCB model h_R and the potential SCB is formulated by *Mahalanobis distance.* The representative SCB 1-D color histogram h_R is created by averaging the histograms of the segmented SCBs in a GoF. By comparing the color histogram of on-going video frame masked by h_k^{Mscb} with the representative histogram h_R, we may identify whether the content of the SCB has changed.

B. Segmentation Scheme

This paper illustrates three levels of boundaries for video segmentation from the coarse to fine granularity to define the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

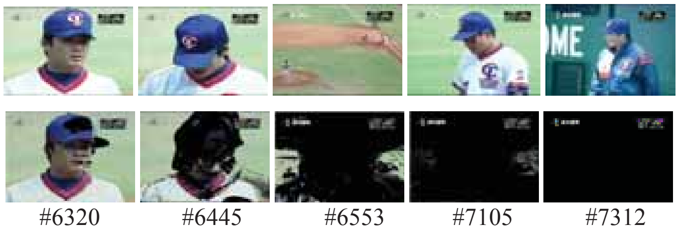


Fig.2. An example of the SCB appearance detection.

content that includes (1) Highlights (HL); (2) attentive visual change (AVC); (3) game status changes (GSC).

1) *Highlights (HL)*

For some types of sports programs usually comprise more than 80% break scenes. Normally, we may utilize the attentive motion information to determine that the shot is belongs to break (e.g., Static) or play (e.g., Pan/Tilt/Zoom). However, this determining relation will depend on the sports types. To propose a general approach, we use highlight instead of the plays-and-breaks. The highlight of the game should belong to the plays shot. We apply the logo appearance detection to infer the highlight event occurs. Because that the highlight is usually sandwiched by the logo scene transition, which is a compromised general method.

2) *Attentive visual change (AVC)*

In this paper, we propose a scene cut detection method, which composes the horizontal and vertical SAD value to detect the abrupt scene change and the gradual scene change. Here, we discuss four kinds of scene change scenarios: (1) abrupt changes, (2) fading changes, (3) dissolve, and (4) wipe. To find the scene change frame, our process need to determine frame k^* of which the sum of horizontal and vertical SAD values is the maximum for all possible center position n_o and offset s.

3) *Game status change (GSC)*

Here, we define two types of shot boundary associated with the SCB: *pop-out* and *content-change*. The former is determined when the SCB abruptly appears on the screen. The latter is defined when the on-screen SCB has changing content. The *pop-out* indicates globally changing, whereas the *content-change* represents locally varying. Since the context within SCB will usually change immediately is following the occurrence of a new event, which usually consists of one or more than one shots. Fig. 2 shows the example of SCB appearance detection.

III. Experiments

1) *Testing data and Setup*

For the evaluation of our system, we digitized 4 types of live sports programs about 8 hours including soccer, baseball, basketball, and tennis programs. The video frames used in our experiments were captured from the TV broadcasting programs of the Major League Baseball (MLB), National Basketball Association (NBA), FIFA, and Wimbledon Championships (Wimbledon) in the 2010 season. We manually annotated the ground truth with the video boundaries of pre-defined types such as the HL, AVC, and GSC (as see in Table 1).

Table 1
Test data set

Program	***Game***	***#Shots***		
		HL	AVC	GSC
Baseball	MLB	78	743	206
Soccer	FIFA	36	1,020	4
Tennis	Wimbledon	183	712	168
Basketball	NBA	61	552	74

2) *Objective Evaluation*

The performance was also evaluated in terms of True Positive Rate (TPR) and False Positive Rate (FPR). We performed the proposed method on the testing programs. As shows in Table 2, the AVC-oriented video segmentation method performs the more consistencies on the soccer game than the baseball program. On the contrary, the result of the GSC-oriented segmentation for baseball game is normally better than that of soccer. It is because that the baseball game involved much more contextual information within the SCB. We do not take the time remaining clock into account, the ground truth of the GSC for soccer is only represents the goal boundary (i.e. goal event). On the other hand, the results of the HL-oriented video segmentation for basketball and tennis programs show in Table 3. We performed the logo transition detection to label the start and end point as the duration of highlight occurs. As the result, it is unrelated to the types of sports, whereas it is only affected by the formation of the logo transition.

Table 2
Results of AVC & GSC-oriented video segmentation

Soccer / Baseball	**Results**		
	Ground Truth	***TPR* (%)**	***FPR* (%)**
AVC	1,020/743	88.1/86.1	8.6/13.9
GSC	4/206	100/96.6	0/2.4

Table 3
Results of HL-oriented video segmentation

Basketball / Tennis	**Results**		
	Ground Truth	***TPR* (%)**	***FPR* (%)**
HL	61/183	98.4/98.9	0/3.3

IV. Conclusion

We have illustrated three levels of boundary for video segmentation with different granularities to define the content including highlights, attentive visual change, and game status change. The proposed algorithm has referred the high level semantic meanings to improve the perceptual consistency between the human perception and the digital content.

References

[1] S. Zhu and Y. Liu, “Scene Segmentation and Semantic Representation for High-Level Retrieval,” *IEEE Signal Processing Letters*, vol. 15, pp. 713–716, 2008.

[2] T. Vlachos, “Cut detection in video sequences using phase correlation,” *IEEE Signal Processing Letters*, vol. 7, no, 7, pp. 173–175, 2000.

[3] Y. Zhai and M. Shah, “Video scene segmentation using Markov chain Monte Carlo,” *IEEE Trans. on Multimedia*, vol. 8, no. 4, pp. 686–697, 2006.

An Application-level Energy-Efficient Scheduling for Dynamic Voltage and Frequency Scaling

Keunjoo Kwon, Seungchul Chae, and Kyoung-Gu Woo
Samsung Advanced Institute of Technology, Samsung Electronics, Korea

***Abstract*—Power consumption in mobile devices is a critical issue with monitoring-based services. Operating systems in smartphones employ interval-based dynamic voltage scaling algorithms to reduce power consumption. To boost the effect of those algorithms, we propose an application-level scheduling algorithm which slows down the execution of the application deliberately and thus maintains low utilization rate of CPU. The experimental result shows the proposed algorithm saves up to 32% of power consumption.**

I. Introduction

Nowadays, many people carry their smartphones with them around the clock in daily life. Such pervasiveness of smartphones, combined with the recent advances of sensor and wireless communication technologies, enables new kinds of monitoring-based services such as remote healthcare services or context-based services. Those services involve smartphones to process data streams from sensors continuously, thus they may consume a significant amount of the power. Therefore, considering the limited capacity of the battery in mobile devices, energy efficiency of the continuous data stream processing is a critical issue for building practical monitoring-based service systems.

To reduce power consumption of the smartphones, general-purpose operating systems (OS) incorporate process schedulers with interval-based scheduling algorithms which utilize the dynamic voltage scaling capabilities of modern processors. Interval-based scheduling algorithms decide the supply voltage of the central processing unit (CPU) based on the utilization rate of the CPU. By the Ohm's law stating that power consumption is proportional to the square of the applied voltage, lowering the supply voltage could greatly save power consumption of the processor. However, the effect of the scheduling algorithms is diminished when applications greedily raise the utilization rate of the CPU. Therefore, applications should maintain low utilization rates in order to reduce power consumption.

This paper proposes an application-level scheduling algorithm to boost the effects of interval-based scheduling algorithms of the OS's process schedulers. The proposed algorithm deliberately slows down the processing of an application as much as possible by calling the sleep API. Such intentional slowdown can fit in with the continuous data stream processing in monitoring-based services, because the average workload is relative low compared to the maximum processing capabilities of smartphones. By calculating minimum required cycles, the algorithm does not miss the deadline of the processing while it is running slowly. The experimental results show that the proposed algorithm enables the OS scheduler to keep the voltage of the process low and thus the power consumption can be reduced up to 32% when the system workload is 30% of the maximum capacity.

TABLE I. Measured Power Consumption of A Smartphone

CPU Clock	CPU Utilization rate	CPU State	LCD State	Power (mW)	Power/Cycle (nW)
200MHz	0%	Awake	Off	202.05	-
200MHz	100%	Awake	Off	328.79	0.634
500MHz	0%	Awake	Off	205.53	-
500MHz	100%	Awake	Off	548.09	0.685
1.2GHz	0%	Awake	Off	373.91	-
1.2GHz	100%	Awake	Off	1859.00	1.237

II. Data Stream Processing Environments in Mobile Platforms

A. Processor Characteristics

CPU is the major source of power consumption in the mobile platform [1]. Even though display and graphics modules also may consume a significant amount of energy with interactive applications, monitoring-based applications usually involve CPU-intensive background processing and occasional user interactions. To reduce the power consumption of processors, manufacturers have been providing a dynamic voltage and frequency scaling (DVFS) method which enables softwares to manage the supply voltage. The method is based on the physical principle of semiconductors whereby the switching rate of a transistor decreases when an input voltage decreases, and thus the operating frequency of a processor has to be reduced. Energy consumed by a processor is expressed according to Ohm's law, as given by the equation

$$P = C \cdot V^2 \cdot F \tag{1}$$

where P is the power consumption, C is the load capacitance, V is the supply voltage, and F is the operating frequency. The power consumption of the processor is proportional to the cube of the supply voltage, because the operating frequency has to be scaled according to the supply voltage. TABLE I shows the externally measured power consumption of a smartphone model used in our experiments. The device's processor has capability of DVFS from 200MHz to 1.2 GHz. The power consumption per cycle is the difference of power between the utilization rates 100% and 0% divided by CPU clock, which means the energy required to do a certain amount of work besides the basic power consumption. The power consumption per cycle of 1.2GHz is almost two times larger than that of 200MHz.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Scheduling Algorithms of Operating Systems

A wide variety of DVFS algorithms has been proposed for the process scheduling in OS. Those can be categorized into two classes. The one is task-based scheduling and the other is interval-based scheduling [2].

Task-based DVFS scheduling algorithms determine the lowest required operating frequency, based on the estimated worst case execution time and the given deadline of each task. These algorithms aim to meet the deadline by estimating the execution times of tasks accurately [3]. But those scheduling algorithms can work only with the real-time OSs, where the deadlines of tasks are declared in advance.

Interval-based DVFS scheduling algorithms monitor the utilization rate of the CPU during a pre-defined interval and change the frequency according to the utilization rate. If the utilization rate is higher than the given upper threshold, it steps up the frequency. When the utilization rate becomes less than the given lower threshold, it steps down the frequency. Those algorithms are widely adapted to general-purpose OSs such as Windows, Linux and Android OS, since they don't require modifications of applications.

Figure 1 illustrates exemplary cases of a DVFS with the measurements in TABLE I. For a simple explication, power consumption of idle state is not considered. Assuming that a workload of 12G cycles is given, the CPU with 1.2GHz takes 10 seconds to complete it and thus consumes 18.59J as shown in Figure 1 (a). If the CPU works with 500MHz for 12 seconds and then the frequency is raised to 1.2GHz, 15.872J is consumed totally as in Figure 1 (b). In the case the deadline is known as 25 seconds, the frequency may stay in 500MHz for 24 seconds consuming 13.154J as in Figure 1 (c), which is 70% of the case in Figure 1 (a).

C. Mobile Data Stream Processing Applications

Mobile devices are equipped with various embedded sensors generating large amounts of data stream. Smartphones usually incorporate GPS, accelerometer, and magnetometer as embedded sensors. External sensors also can be easily connected to the device through wireless communications. Especially, healthcare services may involve various kinds of external measurement devices such as electrocardiography (ECG) sensors, glucose meters or pulse oximeters. For example, a smartphone connected with an ECG sensor can act as a cardiac monitoring device as shown in Figure 2, performing feature extraction and arrhythmia detection [4].

Processing data streams from sensors in mobile environments can be characterized as continuous processing and under-loaded workloads. For example, context-inference applications process incoming sensor data continuously, in order to identify noticeable changes of context from perceived information [5]. Healthcare applications monitor medical conditions of a patient by applying algorithms to the measured signals [4], [6]. Meanwhile, the data transmission rates of those sensors are relatively low due to the low sampling rate or the aggregated transmission. For example, the sampling rates of accelerometers in recent smartphone models are around 100Hz and ECG sensor's sampling rates are around 250Hz. To reduce power consumption of wireless communication, external ECG sensors transmit the measured signals less often than a second. Furthermore, the workload of data processing is designed to be considerably less than the maximum capacity of the system. Otherwise, continuous full utilization of CPU would drain the battery too fast, so that smartphones fail to fulfill the role of a communication device.

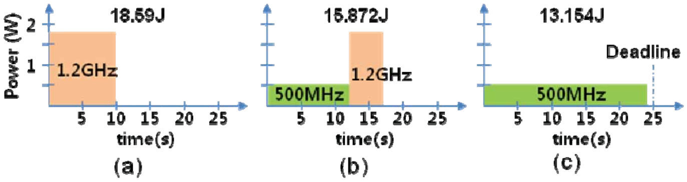


Figure 1. Examples of Dynamic Voltage and Frequency Scaling

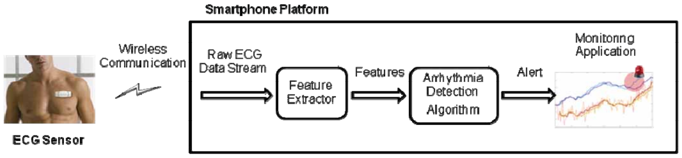


Figure 2. Healthcare Application with an Electrocardiography Sensor

III. Energy-Efficient Scheduling Of Data Stream Processing

A. Scheduling of Execution

To reduce power consumption, we propose an application-level scheduling algorithm working in conjunction with the interval-based DVFS scheduling algorithms of OSs. While many interval-based DVFS algorithms have been proposed, those algorithms are focused on avoiding unnecessarily high supply voltages by setting the CPU clock appropriate to applications' demand. However, the scheduler of the OS cannot help raising the supply voltage of the CPU, in the case that an application requests high computational power within short period of time. To overcome such limitations of interval-based DVFS algorithms, applications should cooperate on maintaining low utilization rates of CPU. To exploit off-the-shelf smartphones, we took an application-level approach without modifying OS. Given a fixed amount of workload, the slower an application executes, the lower DVFS scheduling algorithms can adjust the supply voltage of the CPU.

To control the speed of execution in the application level, our approach inserts a scheduling algorithm into application's processing codes, calling the sleep API after executing some amounts of processing per every unit time. Two things are prepared before the run-time. First, the data processing code is reorganized into interfaces which allow incremental execution. For example, an ECG processing code can be made to be called sample by sample, where the number of execution per second equals to 250 in the case of 250Hz sampling rate. Second, a unit time is determined to decide how often the scheduling algorithm to call the sleep API. During the run-time, the proposed scheduling algorithm executes a certain amount of data processing and sleeps for the rest of the unit time, by measuring the actual spent time of the code execution.

When the execution time exceeds the unit time, the scheduling algorithm skips the sleep.

When there are multiple data streams to process, the scheduling algorithm selects a data stream to process each time it schedule. As a selection algorithm, we employed the Earliest Deadline First (EDF) algorithm which always chooses the data stream having the earliest deadline. It is known that the EDF algorithm guarantees that all deadlines are met provided that the total CPU utilization rate is not more than 100%.

B. *Minimum required cycles per unit time*

In order to decide how slow the execution should be, the scheduling algorithm calculates the minimum cycles per unit time required to meet deadlines of all data stream processing codes. Because most sensors measure values periodically, we assume that the input data streams of sensor data processing applications are fed with measured values periodically. For each periodic data stream, the deadline of data processing has to be smaller than or same as the period of the stream. Otherwise the system cannot handle the increasing size of the input buffer, caused by processing delay.

The minimum required execution rate to meet the deadline can be called as the slowest execution rate. The slowest execution rate of a job is the total expected cycles to finish the job divided by the deadline of the job. The total required execution rate of multiple jobs is obtained by summing up the required execution rate of each job. The scheduling algorithm sets the job as the processing of a data set fed at a time from the stream and the deadline of the job as the period of the stream.

The total execution rate multiplied by the unit time becomes the minimum required cycles per unit time as the following equation

$$\left(\sum_{i=1}^{n}\frac{C_i}{D_i}\right)\times t_u \qquad (2)$$

where C_i is the expected number of cycles required to process each data set of an i-th stream, D_i is the deadline of an i-th stream, t_u is the unit time and n is the number of streams. The number of cycles to process each data is measured empirically and thus provided to the scheduling algorithm as a parameter.

IV. Experimental Results

A. *Experiment Configuration*

The prototype of our scheduling algorithm was implemented on the Android platform. The hardware used for the experiment is Samsung Galaxy S II and its power consumption was measured by an external power monitor [7]. As the scheduler of Android OS is based on Linux Kernel governor, the policy of the scheduler is set to "Ondemand" scaling from 500MHz to 1.2GHz [8]. Ondemand policy is one of interval-based scheduling policies, which is recommended by the experimental calibration for case that there are many background processes [9]. Considering the environment of continuous processing, the minimum clock is set to be a little higher than the lowest of the processor. The scheduling interval is 100 ms and utilization rate threshold for stepping up is 90%. To demonstrate the effect of the scheduling, we simulated an artificial data stream processing with looping of calculations. It is configured that a bulk of a data stream is fed every two seconds and the number of sensor data is controlled by a parameter for each experiment.

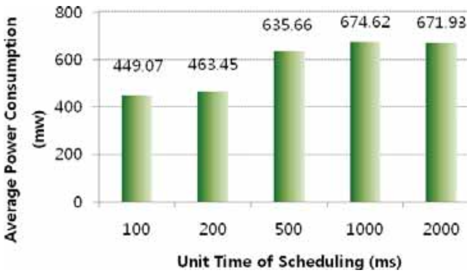


Figure 3. Average Power Consumption with Scheduling

B. *Experimental Results*

To demonstrate the effect of the unit time, we measured power consumptions with varying unit times. Figure 3 shows the average power consumption of the proposed scheduling algorithm with varying unit times from 100ms to 2000ms. The effect of the scheduling becomes maximized when the unit time becomes equal to the interval of the OS scheduler which is 100ms. With the unit time larger than the interval of the OS scheduler, the power consumption increases because it raises the utilization rate of CPU within in the interval of 100ms.

Figure 4 shows the average power consumption with and without the scheduling algorithm, along with various ratios of the workload to the maximum capacity. When the system is under-loaded (e.g. less than 70% of capacity), the processing with the proposed scheduling algorithm consumes less power than that without the scheduling algorithm. When the workload is 30%, the system uses 32% less power which means a 47% extended battery life. As the workload increases, the gain becomes smaller and the power consumption of the scheduling algorithm becomes as same as that without the scheduling.

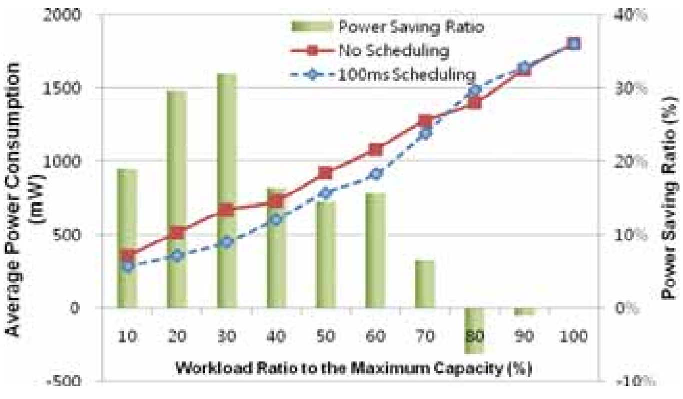


Figure 4. Power Consumption by Utilization rate

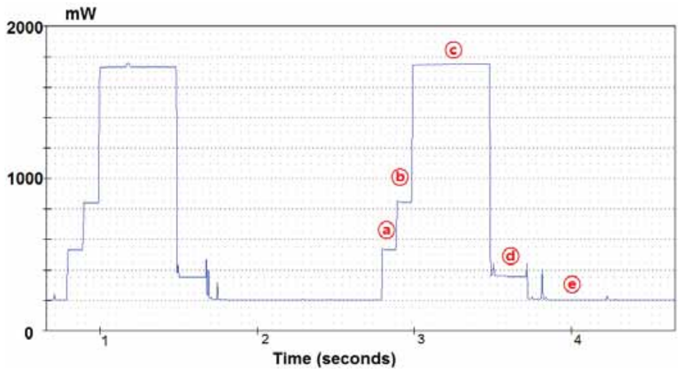


Figure 5. Power Consumption without Scheduling

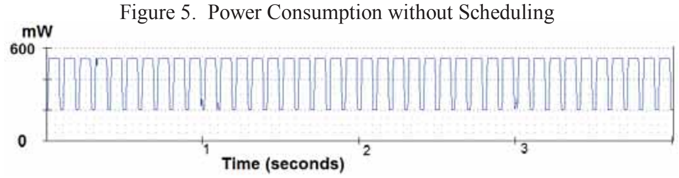


Figure 6. Power Consumption of Scheduling with100 ms Unit Time

Figure 5 shows the power consumption without scheduling along with time. The workload is set to be 30% of the maximum capacity. During the interval (a), the lowest voltage is maintained and then the OS raised the voltage to the next level during the interval (b) because the utilization rate was higher than the 90% threshold during the interval (a). During the interval (c), the voltage is set to the processor's highest with the same reason. After finishing all workload, it becomes idle during the interval (d) and then the voltage is set to the lowest during the interval (e). The interval (d) consumes higher energy than the interval (e) due to the leakage current of the high supply voltage.

Figure 6 shows the power consumption of the scheduling algorithm along with time. The unit time of scheduling is 100ms and the workload is set to be 30% of the maximum capacity. The highest power consumption is around 550mW, implying that the operating frequency remains as 500MHz. The applied workload (30% of the 1.2GHz) is 360MHz which is 72% workload to 500MHz. By executing only 36M cycles every 100ms, the scheduling algorithm maintains the utilization rate less than 90% threshold and thus keeps the OS scheduler from raising the supply voltage.

V. Related Works

In the field of OS, a wide variety of DVFS scheduling algorithm has been proposed for decades. Task-based scheduling aims the real-time environment such as multimedia systems where arrival times and deadlines of task are provided in advance [10]. Those algorithms determine the lowest required supply voltage by estimating the workloads with stochastic models. With a stochastic model, the number of input data values can be forecast beforehand and our algorithm can benefit from them. Recent interval-based scheduling algorithms estimate the best suited voltage on the basis of task characterization by utilizing processors runtime statistics such as cache hit/miss ratio [11]. The effect of those algorithms can be combined with the effect of our algorithm as we intended to run our algorithm in conjunction with them.

Application-level scheduling of data stream processing has been researched in the area of data stream management system [12]. The scheduling algorithms of stream processing focused on ordering the execution of queries to improve metrics such as average response time or average slow down. Those selection algorithms can be employed as a selection algorithm.

VI. Conclusion And Future Work

This paper proposed a scheduling algorithm for applications to increase the effect of the interval-based DVFS scheduling algorithm of the OS. The algorithm purposely sleeps after executing minimum required cycles per unit time and thus reduces the power consumption of the CPU by maintaining low utilization rate. Our algorithm can be effective in the environments where the applications process data streams continuously with under-loaded workloads such as sensor data processing in smartphones. The experiment demonstrated that the power consumption can be saved up to 32% when the system workload is 30% of the maximum capacity. As a future work, we plan to incorporate the proposed algorithm into the data stream management system to provide systematical support for developers.

References

[1] A. Carrol and G. Heiser, "An Analysis of Power Consumption in a Smartphone," In *Proc. of the 2010 USENIX Technical Conference*, 2010.

[2] J. R. Lorch and A. J. Smith, "Task-Based Speed and Voltage Scheduling on Windows 2000," *Technical Report: CSD-02-1190, University of California at Berkeley*, 2002.

[3] W. Yuan and K. Nahrstedt, "Energy-efficient soft real-time CPU scheduling for mobile multimedia systems," In *ACM SOSP*, 2003.

[4] P. Leijdekkers, V. Gay, and E. Barin, "Trial Results of a Novel Cardiac Rhythm Management System Using Smart Phones and Wireless ECG Sensors," In *Proc. of the 7th Int. Con. on Smart Homes and Health Telematics*, 2009.

[5] Y. Wang, J. Lin, M. Annavaram, Q. A. Jacobson, J. Hong, B. Krishnamachari, and N. Sadeh, "A Framework of Energy Efficient Mobile Sensing for Automatic User State Recognition," In *Proc. of the 7th Int. Conf. on Mobile Systems, Applications, and Services*, 2009.

[6] Y. Kawahara, N. Ryu, and T. Asami, "Monitoring Daily Energy Expenditure using a 3-Axis Accelerometer with a Low-Power Microprocessor," *e-Minds: International Journal on Human-Computer Interaction*, vol. 1, no. 5, 2009.

[7] Monsoon Solutions Inc., *Power Monitor* [Online]. Available: http://www.msoon.com/LabEquipment/PowerMonitor/

[8] D. Brodowski, *Linux CPUFreq Governors* [Online]. Available: http://www.kernel.org/doc/Documentation/cpu-freq/governors.txt

[9] Thimmarayaswamy K, M. M. Dsouza, and G. Varaprasad, "Low power techniques for an android based phone," *ACM SIGARCH Computer Architecture News,* vol. 39, no. 2, pp. 26-35, 2011.

[10] Z. Cao, B. Foo, L. He, and M. van der Schaar, "Optimality and Improvement of Dynamic Voltage Scaling Algorithms for Multimedia Applications," *IEEE Trans. on Circuits and Syst. I: Regular Papers*, vol. 57, no. 3., 2010.

[11] Gaurav Dhiman and Tajana Simunic Rosing, "Dynamic voltage frequency scaling for multi-tasking systems using online learning," In *Proc. of the Int. Symp. on Low Power Electronics and Design*, 2007.

[12] M. A. Sharaf, Panos K. Chrysanthis, Alexandros Labrinidis, and Kirk Pruhs, "Algorithms and metrics for processing multiple heterogeneous continuous queries," *ACM Trans. on Database Syst.,* vol. 33, no. 1, 2008.

Virtual Bass System Based on a Multiband Harmonic Generation

Taegyu Lee[1], Seokjin Lee[2], Young-cheol Park[3], and Dae Hee Youn[1]
[1]School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
[2]LG Electronics, convergence R&D Lab.
[3]Computer & Telecommunication Engineering Division, Yonsei University, Wonju, Korea

***Abstract*— In this paper we propose a method of creating virtual bass based on a multiband harmonic generation. The proposed algorithm separately generates harmonics for each band whose bandwidth is adaptively adjusted according to the tonal distribution in the bass signal. Both even and odd harmonics are generated by combining two nonlinear devices. The proposed algorithm overcomes the intermodulation distortion and spectral smearing problem that are often encountered in the previous algorithms. Results of objective and subjective tests are presented to validate the proposed algorithm.**

I. NTRODUCTION

Loudspeakers of consumer devices have poor performances on reproducing low frequency component of the sound signal due to their sizes and cost constraints. Virtual bass system (VBS) enhances low frequency reproduction by using the psychoacoustic property of human auditory system. This property is known as “Missing Fundamentals” which suggests that pitch perception of a set of harmonics without fundamental frequency can result in a perception of the fundamental frequency [1].

Many efforts have been made to compensate for the low frequency capability using “Missing Fundamentals”. The nonlinear device (NLD) approach has advantages of structural simplicity and ease of transient handling [3]. However, the nonlinear processing generates intermodulation distortion (IMD) which causes unnatural artifacts [3]. As an alternative approach, a method of using the phase vocoder (PV) was suggested [3]. This method allows for precise control over the individual harmonic components [4]. However, the PV approach often exhibits smearing effect which generates buzzy artifacts [4]. To cope with drawbacks of both NLD and PV, a hybrid approach was suggested [4]. This approach combines advantages of NLD and PV approaches by using a transient content detector. However, this approach cannot completely remove the smearing effect of PV. In this paper, we propose a new virtual bass system that can overcome the problems of the previous methods.

II. PROPOSED ALGORITHM

A schematic diagram of the proposed VBS is shown in Fig. 1. To deal with the computational complexity issue, the system is implemented using the short-time Fourier transform (STFT) and the signals are low-pass filtered and down-sampled by a factor of D.

A. Harmonic and Percussive Component Separation (HPCS)

Percussive bass components such as kick drum sound are not harmonically related. To enhancing such percussive components in low frequency we first separate the percussive components from the bass input. We utilize the median filtering approach [5] which is known to be fast and effective. HPCS performs median filtering across successive frames and frequency bin. The two resulting median filtered spectrograms are used to generate masks which are applied to the original spectrogram to separate the harmonic and percussive parts of the signal [5]. Since “Missing Fundamental” cannot be applied to percussive components, we just apply a band boosting to the separated percussive components. Overtones based on Missing Fundamentals are generated only for the harmonic components.

B. Spectral Analysis

In order to minimize IMD, the proposed algorithm performs Spectral Analysis (SA) which splits the FFT bins into a group of spectral bands. Using a 1st order derivative of spectrum, local minimum is obtained and it is set as a partition frequency of the filter bank. As a result of band partition, each frequency band contains only one spectral peak [6]. After band partition, each band signal is transformed back to the time-domain via IFFT or sum-of-sinusoids method. Accordingly, dominated tonal components in the bass signal are separated. In fact, this approach can be considered as a variable-bandwidth filterbank.

C. Harmonics generation

According to pattern recognition model of pitch perception, human beings perceive pitch by using frequency differences in complex tones [1]. In order to perceive virtual pitch, at least three consecutive overtones are needed. To satisfy these conditions, we use half-wave rectifier (HWR) and clipper (CLP) in parallel. Since HWR creates even harmonics and CLP creates odd harmonics, we can achieve consecutive overtones by combining the both. Let $x_{VB}(n)$ is generated virtual bass signal and the subband signal $x_h(n)$ is separated harmonic component using HPCS. The closed-form nonlinear equation is defined,

$$f_{HWR}(x_h(n)) = 0.5\left(x_h(n)+\left|x_h(n)\right|\right) \qquad (2)$$

$$f_{CLP}(x_h(n))=\begin{cases} 0.5\operatorname{sgn}(x_h(n)) & \mathit{if}\ \left|x_h(n)\right|>0.5 \\ x_h(n) & \mathit{otherwise} \end{cases} \qquad (3)$$

$$x_{VB}(n) = f_{HWR}(x_h(n)) + f_{CLP}(x_h(n)) \qquad (4)$$

One problem associated with CLP is that it is amplitude sensitive. CLP doesn’t work properly for low-amplitude signals. To prevent this problem, we employ a normalization technique, in which the input of CLP is normalized with its energy estimate, and the output of CLP is de-normalized to

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

restore the original energy.

III. Performance Evaluations

In order to validate the proposed VBS, objective and subjective tests were carried out. First, to evaluate the harmonic generation, an input signal consisting of two sinusoids with 80Hz and 100Hz was processed by NLD, PV and proposed algorithms, respectively. Fig. 2 compares the output spectra. The results show that the NLD approach creates significant intermodulation distortion at 180Hz, 220Hz, 250Hz, 280Hz, 340Hz, and so on. In the case of the PV and proposed methods, harmonics of 80Hz and 100Hz are well generated and intermodulation distortion is negligible.

Secondly, Fig. 3 shows the spectrograms of the output signals for a pop music where impulsive bass occurs consecutively. The PV approach obviously suffers from the smearing effect especially at higher overtone while the proposed algorithm generates stable harmonics.

Finally, subjective listening test were conducted in which eight subjects were asked to grade two factors: (1) Bass quality (2) Distortion in terms of noise, artifacts and timbre. The proposed system is evaluated in comparison to NLD and PV systems. Fig. 4 shows the mean and 95% confidence intervals of bass quality (left axis) and distortion (right axis). It is clear from Fig. 4 that the proposed method provides significant enhancement of natural bass perception.

IV. Conclusion

In this paper we proposed a virtual bass system using a multiband harmonics generation. The proposed algorithm has strengths both transient handling capability of nonlinear device and stable harmonic generation of phase vocoder. The proposed algorithm is rated objectively and subjectively against two conventional system using NLD and PV. The result confirms good performance of the proposed method with respect to the compared algorithm.

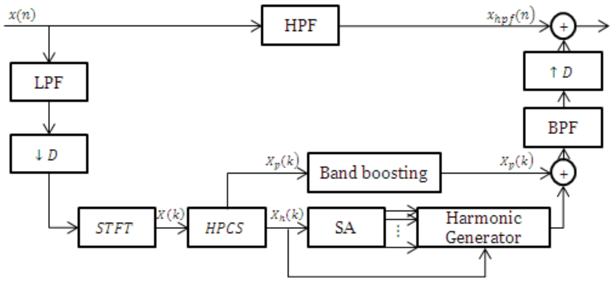


Fig. 1. Block diagram of proposed VBS

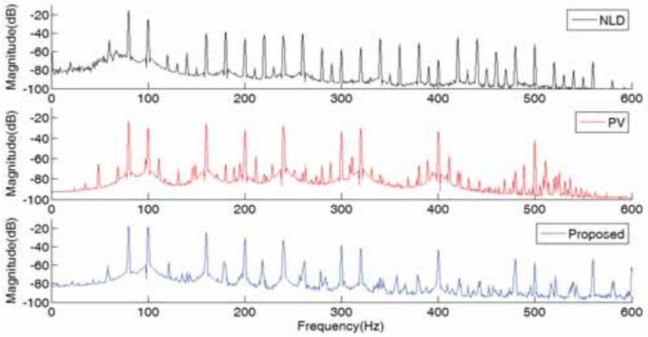


Fig. 2. Spectra of harmonics generated using NLD, PV and the proposed methods.

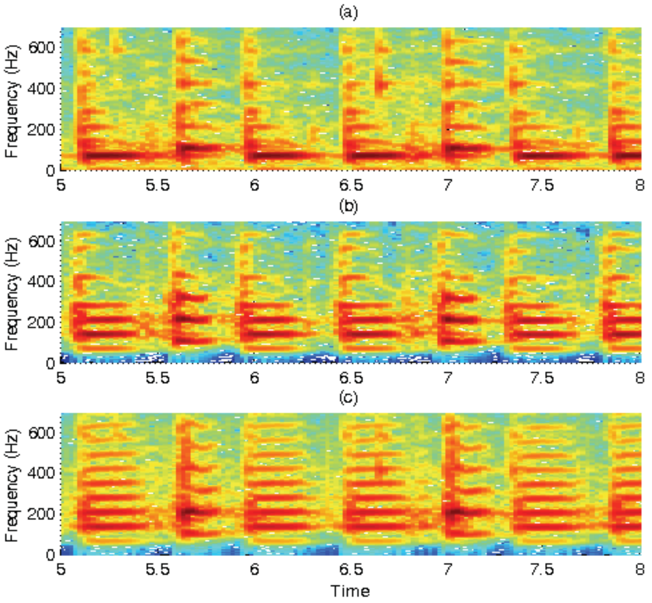


Fig. 3. Spectrograms of (a) the input signal, and the outputs of (b) the PV approach and (c) the proposed method.

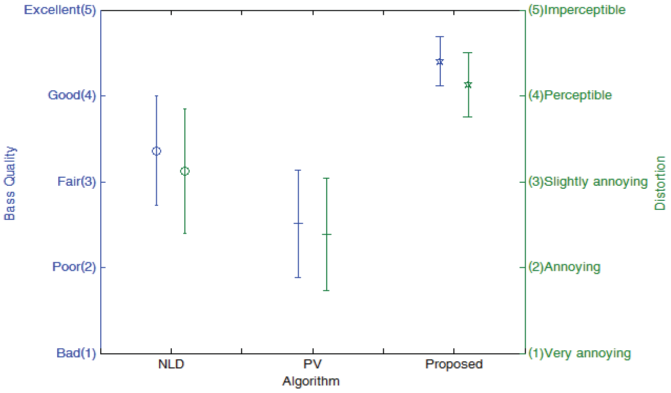


Fig. 4. Subjective listening test results

References

[1] B. C. J. Moore, *An Introduction to the Psychology of Hearing*: Academic Press, 2003.

[2] N. Oo, W. S. Gan, and M. O. J. Hawksford, "Perceptually-Motivated Objective Grading of Nonlinear Processing in Virtual-Bass Systems," *Journal of the Audio Engineering Society*, vol. 59, pp. 804-824, Nov 2011.

[3] M. R. Bai and W. C. Lin, "Synthesis and implementation of virtual bass system with a phase-vocoder approach," *Journal of the Audio Engineering Society*, vol. 54, pp. 1077-1091, Nov 2006.

[4] A. J. Hill and M. O. J. Hawksford, "A hybrid virtual bass system for optimized steady-state and transient performance," in *Computer Science and Electronic Engineering Conference (CEEC), 2010 2nd*, 2010, pp. 1-6.

[5] D. FitzGerald, "Harmonic/Percussive Separation Using Median Filtering," *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria , September 6-10, 2010*, 2010.

[6] J. O. Smith, "Audio FFT filter banks," in Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09), Como, Italy, September 1-4, *Sept. 2009*

Multi-Band Spectral Subtraction Based Zoom-Noise Suppression for Digital Cameras

Kwang Myung Jeon[*], *Student Member, IEEE*, Nam In Park[*], *Student Member, IEEE*,
Hong Kook Kim[*], *Senior Member, IEEE*, Myung Kyu Choi[**], and Kwang Il Hwang[**]
[*]Gwangju Institute of Science and Technology, Gwangju 500-712, Korea
[**]Digital Imaging Business, Samsung Electronics, Gyeonggi-do 443-742, Korea

***Abstract*--This paper proposes a new noise suppression method to reduce zoom noise generated when audio signals are recorded with a digital camera. The proposed method is based on multi-band spectral subtraction that can suppress spectral components of noise related to reference zoom-noise in the modified discrete cosine transform domain. In particular, in the proposed method, each frame is classified as either a noise frame or a non-noise frame, and depending on this classification, the reference zoom-noise is updated and the degree of suppression is controlled. It is shown from performance evaluation that noise due to a zooming operation of digital cameras is successfully suppressed while maintaining audio quality.**

I. INTRODUCTION

Today's digital cameras are increasingly used to record video and audio, diminishing the use of camcorders. One drawback to audio recorded by digital cameras is in that a significant level of mechanical noise is introduced by the camera's zoom operations. One intuitive solution is to limit the speed of the zoom motor [1]. However, such a solution decreases the zooming speed of digital cameras, making it difficult to capture fast moving objects. Therefore, additional effort should be needed to overcome the trade-off between the zoom speed and the noise level.

As an alternative to reducing the zoom-noise level without decreasing the zoom speed, a mechanical noise suppressor for digital cameras was proposed by adapting reference noise [1]. Zoom-noise reduction in this approach was carried out by assuming that a priori information on the intervals of zoom motor operation was precisely known and that only zoom noise existed during those intervals to update the reference noise. However, it is difficult to measure the exact timing of the zoom operation owing to unexpected time delay and/or jittering between the zoom-motor movement and its activation time. Moreover, audio signals and zoom noise are commonly mixed during the zoom-noise intervals. These factors cause performance degradation of zoom-noise reduction.

In order to address the aforementioned issues, we propose a zoom-noise suppression method by incorporating a zoom-noise detection algorithm. By doing this, information pertaining to zoom-noise operation is unnecessary. The proposed method is based on multi-band spectral subtraction (MBSS), which suppresses the spectral components of noise related to reference zoom-noise in the modified discrete cosine transform (MDCT) domain [2]. Moreover, for a given audio frame, the zoom-noise detection algorithm first estimates the sub-band signal-to-noise ratios (SNRs). Then, it controls the degree of suppression in the MBSS and determines whether the audio frame is a zoom-noise frame or not, according to the distribution of the sub-band SNRs over frequency. In other words, the reference zoom-noise is updated only if this audio frame is declared as a zoom-noise frame.

II. PROPOSED ZOOM-NOISE SUPPRESSION METHOD

Fig. 1 shows a flowchart of the proposed zoom-noise suppression method that operates in the MDCT domain because of higher performance of energy compaction and spectral resolution than in the Fourier transform domain [3]. First, the proposed method segments audio signals into a frame whose number of samples is 1024, corresponding to 32 ms at a sampling rate of 32 kHz. Next, it applies an MDCT to audio signals deteriorated by zoom noise and divides the MDCT coefficients into 49 sub-bands whose bandwidths are identical to those in MPEG advanced audio coding (AAC) [4]. After that, for a given l-th frame, the proposed method estimates SNR of the sub-bands, i.e., $SNR(l,k)$, $k = 0,\cdots,48$, by comparing the sub-band power of the audio signal and that of the reference zoom-noise. Note here that the reference zoom-noise is the zoom-noise signal recorded by a digital camera in a quiet environment. The estimated SNR for each sub-band is then used for zoom-noise detection. According to the result of zoom-noise detection, the reference zoom-noise is updated and the

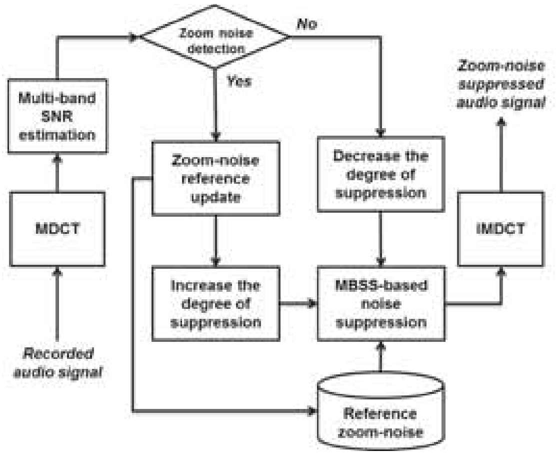

Zoom noise detection
No
Yes
Multi-band SNR estimation
Zoom-noise suppressed audio signal
Zoom-noise reference update
Decrease the degree of suppression
MDCT
IMDCT
Increase the degree of suppression
MBSS-based noise suppression
Recorded audio signal
Reference zoom-noise

Fig. 1. Flowchart of the proposed zoom-noise suppression method.

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-010636) and by Samsung Electronics.

*Corresponding Author: Hong Kook Kim (hongkook@gist.ac.kr)

degree of noise suppression is controlled for MBSS-based noise suppression. Finally, an inverse MDCT (IMDCT) is applied to obtain a zoom-noise suppressed version of the recorded audio signal.

As described above, the performance of the proposed method is highly dependent on that of the zoom-noise detection algorithm. The detection algorithm first counts the number of sub-bands whose SNR is below a predefined threshold, SNR_{thres}. That is,

$$N(l) = \frac{1}{49} \sum_{k=0}^{48} I(SNR(l,k), SNR_{thres}) \qquad (1)$$

where $I(x,y)=1$ if $x \le y$, otherwise $I(x,y)=0$. The l-th frame is declared as a zoom-noise frame if $N(l) \ge N_{thres}$. In this paper, the parameters are set as $SNR_{thres}=0$ and $N_{thres}=0.7$ from the exhaustive preliminary experiments.

If the current frame is a zoom-noise frame, the recorded audio signal is averaged with the reference zoom-noise. Subsequently, this averaged reference zoom-noise is used for the MBSS-based noise reduction. The degree of noise suppression in MBSS is controlled depending on whether the current frame is a zoom-noise frame or not. In other words, the suppression factor is increased for a zoom-noise frame, but deceased otherwise.

Finally, the MBSS-based noise suppression is performed using sub-band SNRs, a suppression factor, and updated reference zoom-noise. By taking an IMDCT, we obtain a zoom-noise suppressed version of the recorded audio signal, as shown in Fig. 1.

III. PERFORMANCE EVALUATION

In order to evaluate the performance of the proposed method, the method was implemented using a commercially available compact digital camera with a zoom function. The camera was equipped with two electret condenser microphones for audio recording. The initial reference zoom-noise was obtained by averaging the zoom noise recorded using five different cameras of the same model. Test audio signals were recorded in an office environment while performing zoom operations. Even though the method was applied to audio signals, the proposed method had short latency enough not to cause video and audio synchronization problem. In other words, the latency was 36.94 ms in total, which summed up an algorithmic delay of 32 ms by the MDCT/IMDCT operation and the processing delay of 4.94 ms measured in the digital camera.

Fig. 2 shows a comparison of the spectrograms of zoom noise recorded in a quiet environment, audio signals recorded without any zoom operation, audio signals recorded during a zoom operation and zoom-noise suppressed audio signals by the conventional method [1] and by the proposed method, respectively. Compared to the spectral components shown in Fig. 2(c), the spectral components of zoom noise shown in Fig. 2(e) were clearly suppressed while the other spectral components were preserved. Moreover, the performance of the proposed method was more preferred than the conventional method, in terms of spectral similarity with audio signal without zoom-noise, as shown in Fig. 2(b).

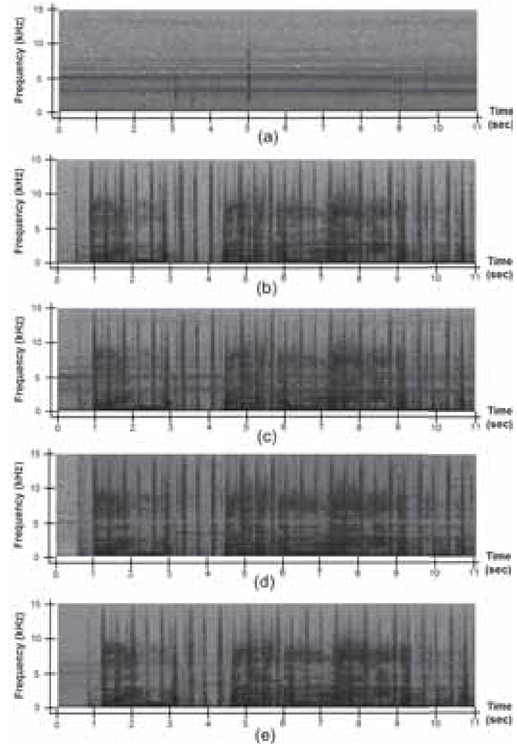


Fig. 2. Spectrogram comparison: (a) zoom noise, (b) audio signal without zoom-noise, (c) audio signal recorded during a zoom operation, (d) zoom-noise suppressed audio signal by the conventional method, and (e) zoom-noise suppressed audio signal by the proposed method.

IV. CONCLUSION

In this paper, a zoom-noise suppression method was proposed to reduce the mechanical noise generated by the zoom operation of digital cameras. The proposed method was performed by detecting zoom noise frames using sub-band SNRs, followed by updating the reference zoom-noise and controlling the degree of suppression. After applying the proposed method to audio signals recorded on a commercially available digital camera, it was shown that the proposed method could successfully reduce the zoom noise, resulting in better audio quality.

REFERENCES

[1] A. Sugiyama, T. Maeda, and K. Park, “A mechanical-noise suppressor based on *a priori* noise information for digital still cameras and camcorders,” in *Proc. of International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, pp. 415-416, Jan. 2011.

[2] K. M. Jeon, N. I. Park, H. K. Kim, M. K. Choi, L. C. Hwang, and S. R. Kim, “MDCT-domain noise reduction with block switching for the application to MPEG audio coding,” in *Proc. of International Conference on Advanced Signal Processing*, Seoul, Korea, p. 98, Mar. 2012.

[3] I. Y. Soon, S. N. Koh, and C. K. Yeo, “Noisy speech enhancement using discrete cosine transform,” *Speech Communication*, vol. 24, no. 3, pp. 249-257, June 1998.

[4] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, “ISO/IEC MPEG-2 advanced audio coding,” *Journal of the Audio Engineering Society*, vol. 45, no. 10, pp. 789-814, Oct. 1997.

A Zooming-Noise Suppressor with No *A Priori* Information for Digital Still Cameras

Akihiko Sugiyama and Ryoji Miyahara†
Information and Media Processing Research Laboratories, NEC Corporation
† Internet Terminal Division, NEC Engineering
1753, Shimonumabe, Nakahara-ku, Kawasakishi, Kanagawa 211–8666, JAPAN

Abstract–**This paper proposes a zooming-noise suppressor with no $a\ priori$ information for digital still cameras. Spectral peaks of the noisy speech are first detected as relevant information and preserved. Other components are suppressed to an estimated ambient-noise level. After these process of noise suppression in magnitude, the accompanying phase is randomized to make the phase-originating artifacts inaudible. Subjective evaluation results show that the proposed zooming-noise suppressor achieves a modified CCR score of 2.0.**

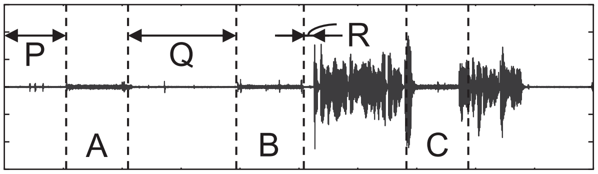


Fig. 1: Typical zooming noise with speech.

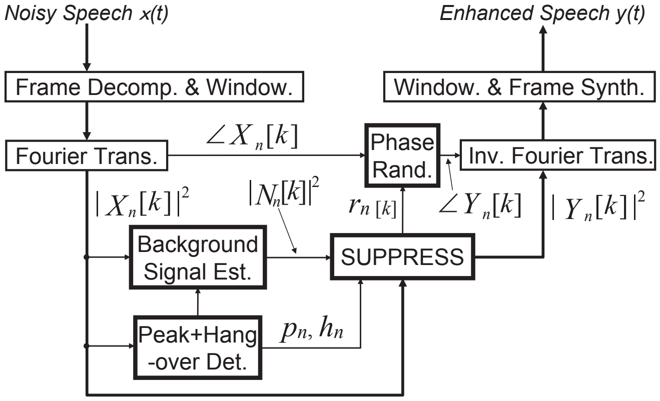


Fig. 2: Blockdiagram of the proposed zooming-noise suppressor.

I. Introduction

Video recording is a today's standard function of digital still cameras (DSCs). One of the most serious problems with DSCs in video recording is that zooming and auto-focusing (AF) noise generated by mechanical components and captured by microphones often contaminates the accompanying sound. In order to suppress the mechanical noise, a zooming-noise suppressor [1] and an AF-noise suppressor [2] have been proposed. In these noise suppressors, a noise reference that is specific to a product model is prepared in advance and used as a noise replica for subtraction from the noisy speech. Individual differences in the actual noise characteristics in each product are compensated for by a set of frequency-domain gain applied to the noise reference which is recursively adjusted based on the residual noise [1]. It is also possible, instead of this feedback approach, to adaptively combine multiple different noise references [2]. However, these noise suppressors need some *a priori* information about the mechanical noise.

This paper proposes a zooming-noise suppressor with no *a priori* information for digital still cameras. After suppressing all noisy speech components but detected peaks to an estimated ambient noise level, the noisy speech phase is randomized to make the residual noise inaudible.

II. Significance of the Phase

Figure 1 shows an example of a zooming noise mixed with speech. Zooming noise is intermittent and has clear noise sections like A, B, and C where zooming noise is suppressed. Please note that the ambient noise should be considered as a part of speech, because it is to be preserved for naturalness. This fact makes the SNRs in sections A and B negative. Sections P, Q, and R are ambient-noise sections with no speech nor zooming noise. In A and B, the enhanced speech level, *i.e.* the residual noise level, is adjusted to the ambient noise level for continuity. If there is any phase characteristics originating from the zooming noise in the enhanced speech, it is easily noticeable at the beginning and ending points of noise sections. This is because sections A and B have a negative SNR and phase of the zooming noise is dominating in the enhanced-speech phase. In order to make such an artifact inaudible even at boundaries of zooming-noise sections, phase randomization is effective.

III. Proposed Zooming-Noise Suppresor

One may be surprised to see that important components of the desired signal alone provides sufficiently good quality if there is some background signal. This is because artifacts caused by missing components of the desired signal are masked by the background signal. In the case of movie recording, background signal, that describes details of the captured scene, must be preserved and can be used for the above-mentioned masking.

Based on this observation, the proposed zooming-noise suppressor preserves important spectral components including formant frequencies and sets all other components to the ambient-noise level. To further reduce the residual noise, that is perceived behind the target signal due to the same phase characteristics as that of the noisy signal, phase randomization is applied to frequency components suppressed to the ambient-noise level.

Figure 2 illustrates a blockdiagram of the proposed

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

zooming-noise suppressor. The input noisy signal is decomposed into frames of L samples and applied an windowing function before it is converted to a frequency-domain signal by Fourier transform. Amplitude of the frequency-domain signal is provided to Background Signal Estimation, Peak and Hangover Detection (Pk+Ho Det.), and Suppression (SUPPRESS). Phase goes to Phase Randomization (Phase Rand.). Peaks are detected in the way described in [3]. Ambient noise is estimated in frequency bins that are not detected as peaks. The amplitude in peak frequency bins is sent to inverse Fourier transform (Inv. Fourier Trans.). Other frequency bins are considered as noise and its amplitude is suppressed to the estimated ambient-noise level. Hangover is detected in Peak+Hangover Det. and treated separately from peaks.

Hangover is determined when there is any peak in a past period to fill gaps in a speech section. A hangover index $h_n[k]$ is set as

$$h_n[k] = \begin{cases} 1 & \sum_{n-Q+1}^{n} p_n[k] > 0 \\ 0 & \text{otherwise} \end{cases}, \qquad (1)$$

where an integer Q is a hangover period. k and n represent the time and the frequency indexes, respectively.

An estimate of the background signal $\tilde{\lambda}_n[k]^2$ is updated based on a first-order leaky integration (recursive filter) with a leaky factor γ in non-peak frequency bins.

For a simple description, a suppression flag $f_n[k]$ that indicates detailed suppression is introduced. For peak bins and non-peak-non-hangover bins, $f_n[k]$ is defined by

$$f_n[k] = \begin{cases} 0 & p_n[k] = 1 \\ 2 & p_n[k] + h_n[k] = 0 \end{cases}, \qquad (2)$$

For non-peak-hangover bins,

$$f_n[k] = \begin{cases} 2 & |X_n[k]|^2 \geq |X_n[k-1]|^2 + \alpha\text{dB} \\ 0 & |X_n[k]|^2 < |X_n[k-1]|^2 \\ 1 & \text{otherwise} \end{cases}, \qquad (3)$$

Based on the suppression flag $f_n[k]$, amplitude of the noise suppressed signal $|Y_n[k]|^2$ is obtained by

$$|Y_n[k]|^2 = \begin{cases} |X_n[k]|^2 & f_n[k] = 0 \\ |X_{n-1}[k]|^2 & f_n[k] = 1 \\ \tilde{\lambda}_n^2[k] & f_n[k] = 2 \end{cases}, \qquad (4)$$

For $f_n[k] = 2$, a randomization index $r_n[k]$ is set to 1 and the phase is randomized. Otherwise, $r_n[k]$ is set to 0 to preserve the noisy-speech phase.

The input noisy signal phase $\angle X_n[k]$ is randomized based on $r_n[k]$ in Phase Rand. to obtain the enhanced signal phase $\angle Y_n[k]$ as

$$\angle Y_n[k] = \angle X_n[k] + r_n[k] \cdot \phi_n[k], \qquad (5)$$

where $\phi_n[k]$ is a random value between $\pm\pi$. $|Y_n[k]|^2$ and $\angle Y_n[k]$ are used to reconstruct the enhanced signal at the output.

IV. Evaluations

Evaluations were performed using a zooming noise signal recorded with a compact DSC that is available in the market. A single electret condenser microphone (ECM) was held as close as possible to the original microphone position from outside the DSC for recording. The recorded noise signal is shown in Fig. 3. Male and female speech with three different levels and sampled at 44.1kHz were mixed with the recorded zooming noise and a street or a office noise source. A total of 14 subjects were asked to give an integer score between ± 3 following a modified CCR [4][1].

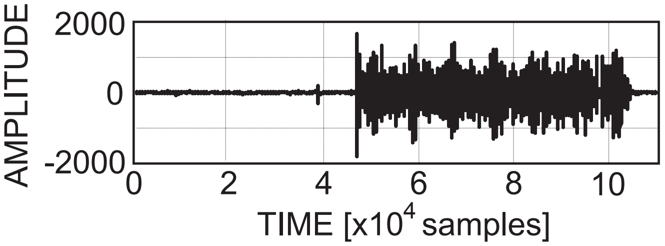


Fig. 3: Zooming-noise used for evaluations.

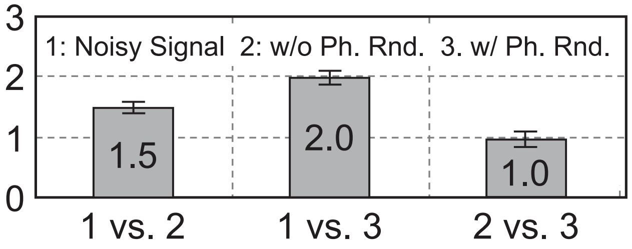


Fig. 4: Subjective evaluation result with modified CCR.

Figure 4 depicts the results in a bar chart with a 95% confidence interval. "1," "3," and "2," in the figure represents the noisy speech, the enhanced speech with and without phase randomization. A positive score means that "β" of "α vs. β" is superior to "α." It is clearly demonstrated that the proposed zooming-noise suppressor achieves scores of 2.0 (1 vs. 3) and 1.5 (1 vs. 2) with and without phase randomization. Because the lower limit of the 95% confidence interval, it is effective. The right-most bar (2 vs. 3) means that phase randomization in zooming noise suppression brings an additional subjective-quality improvement of 1.0.

V. Conclusion

A zooming-noise suppressor with no *a priori* information for digital still cameras has been proposed. It has been demonstrated by a 2.0 modified CCR score that important components of the desired signal alone provides sufficiently good quality if there is some background signal. Phase randomization of the noisy speech has been shown effective with a modified CCR score of 1.0.

References

[1] A. Sugiyama, T. Maeda, and K. Park, "A mechanical noise suppressor based on *a priori* information for digital still cameras and camcorders," Proc. of ICCE2011, pp. 426–427, Jan. 2011.

[2] A. Sugiyama and R. Miyahara, "An auto-focusing-noise suppressor for cellphone movies based on multiple noise references," Proc. of ICCE2012, pp. 45–46, Jan. 2012.

[3] ISO/IEC 11172-3:1993, Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 3 : Audio, Aug. 1993.

[4] "Minimum performance requirements for noise suppresser application to the AMR speech encoder," 3GPP TS 06.77 V8.1.1, Apr. 2001.

[1] Instead of the clean speech in CCR, noisy speech with no suppression is used as a reference.

UHDTV Transmission based on Broadcasting Channel Bonding

Woongshik You*, Joon-Young Jung*, DongJoon Choi*, O-Hyung Kwon*, and Oh-Seok Kwon**
*Electronics and Telecommunications Research Institute, Daejeon, KOREA
**Chungnam National University, Daejeon, KOREA

Abstract — This paper presents a UHDTV transmission scheme using broadcasting channel bonding which transmits a UHDTV program through multiple broadcasting channels. A signaling method is also introduced for providing service information about the UHDTV program transmitted over multiple broadcasting channels.[1]

Index Terms— **UHDTV, Broadcasting Channel Bonding, Service Information Signaling, Massive Broadcasting Contents, Broadcasting System**

I. INTRODUCTION

In recent years, realistic broadcasting services like 3DTV (3-Dimensional TV) have become very popular. In spite of the success of 3DTV, many experts anticipate that UHDTV (Ultra High Definition TV) will be a next generation broadcasting service after current HDTV service.

However, UHDTV has some problems to solve for being introduced as a commercial broadcasting service. One of the problems is huge data rate of UHDTV content. UHDTV content may not be transmitted via single broadcasting channel because of the data rate exceeding the transmission capacity of single channel.

This paper presents a transmission method based on broadcasting channel bonding for providing UHDTV service exceeding the single channel bandwidth. We also introduce a signaling method including a proposed UHDTV program descriptor for transmitting UHDTV content via multiple broadcasting channels.

II. BROADCASTING CHANNEL BONDING

There are two types of UHDTV system supporting different image resolutions as shown in Table 1 and defined in [1]. Because the resolutions of UHDTV 1 and 2 contents are at least 4 times bigger than that of full HDTV content, the capacity of single broadcasting channel may not be enough to transmit the UHDTV content. UHDTV content requires the transmission bandwidth of at least 36Mbps with the assumption that a full HD content is encoded with 9Mbps by H.264/AVC. Because the transmission capacity of single broadcasting channel using 8-VSB (Vestigial Side Band) modulation is about 19.3Mbps at the TS (Transport Stream) level, two terrestrial broadcasting channels are needed to transmit single UHDTV 1 content. Even though HEVC (High Efficient Video Coding) targeting 50% efficiency improvement in comparison with H.264/AVC becomes available, multiple broadcasting channels are still needed to transmit single UHDTV 2 content.

TABLE I. IMAGE SAMPLE STRUCTURES AND FRAME RATES OF UHDTV SYSTEMS

System	Pixels per line	Lines per frame	Frame rate (Hz)
UHDTV1	3840	2160	29.97/30/50/59.94/60
UHDTV2	7680	4320	29.97/30/50/59.94/60

A. TS Packets Distribution

Fig. 1 shows the procedure for transmitting UHDTV contents through broadcasting network. The difference with current HDTV transmission is the added channel bonding process which includes TS packets distribution and resequencing through multiple broadcasting channels to transmit UHDTV content.

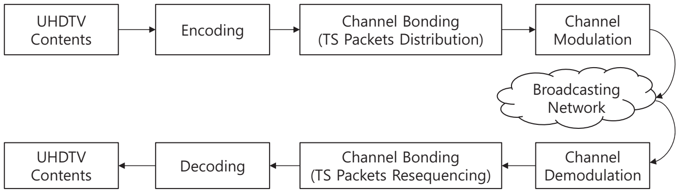


Fig. 1. Procedure of UHDTV content transmission using broadcasting channel bonding

Fig. 2 shows an example of TS packets distribution using multiple broadcasting channels. For combining multiple broadcasting channels, 4-bit CC (Continuity Counter) of TS packet header defined by MPEG is used [2]. During the packet distribution process, the inputted TS packets in order are tagged with CC per PID (Packet Identifier), and then are distributed on multiple broadcasting channels. TS packets have each PID value assigned for video, audio, and ancillary data consisting of the UHDTV content, and are distributed over multiple broadcasting channels to be transmitted.

[1] This work was supported by the IT R&D program of KCC. [10035709, Development of Next Generation Digital Cable Transmission Technology]

When being distributed over multiple channels, each TS packet having the same PID value is numbered by 4-bit CC. The CC per each PID is numbered from 0, increased by 1, and reset to 0 when reaching 15. TS packets assigned to each broadcasting channel are fed to the queue of each channel and transmitted in order.

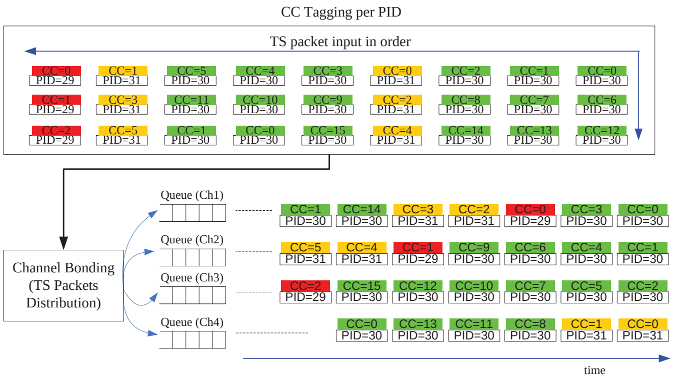


Fig. 2. Example of TS packets distribution on multiple broadcasting channels (4 channels in this example)

B. *TS Packets Resequencing*

At the UHDTV receiver, the TS packets received via multiple broadcasting channels are resequenced by using CC and PID values so that TS stream consisting of single UHDTV program is restored. Fig. 3 shows an example of TS packets resequencing process performed at UHDTV receiver.

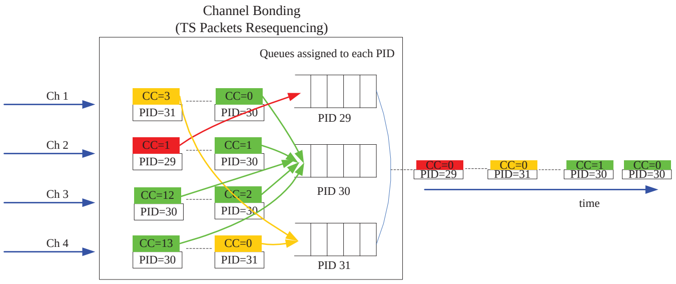


Fig. 3. Example of TS packets resequencing at receiver

III. Signaling UHDTV Service Information

In order to transmit and receive UHDTV program via multiple broadcasting channels, service information about the program needs to be signaled to the receivers. For signaling service information about UHDTV program, we define UHD program descriptor which is included in the PMT (Program Map Table). Table 2 shows the PMT structure including the UHD program descriptor.

Table II. PMT Structure Including UHD Program Descriptor

Syntax	bits
TS_program_map_section() {	
~~ *ellipsis* ~~	
program_info_length	12
for (i=0; i<N; i++) {	
descriptor() ← UHD_program_descriptor is inserted here	
}	
for (i=0; i<N1; i++) {	
~~ *ellipsis* ~~	
}	
}	
CRC32	32
}	

Table 3 shows the structure of UHD program descriptor signaling the transmission method for the corresponding UHDTV content. The UHD_type field indicates whether the UHDTV content is transmitted via single or multiple broadcasting channels. When the UHD_type is multi_stream, it means that TS packets consisting of the UHDTV program are transmitted over multiple broadcasting channels. The num_of_channel field indicates the number of channels used to transmit the UHDTV program.

When transmitting a UHDTV program over multiple broadcasting channels, the detailed information about transmission channel configurations also need to be signaled for UHDTV receiver to tune the channels. In order to signal the detailed information, we use a Multi-channel Descriptor defined in [3] and [4]. The Multi-channel Descriptor is included within VCT (Virtual Channel Table) signaling the whole configurations of the channel.

Table III. UHD Program Descriptor

Syntax	Bits
UHD_program_descriptor () {	
descriptor_tag	8
descriptor_length	8
Reserved	7
UHD_type	1
if (UHD_type=multi_stream) {	
reserved	3
multi_channel_present	1
num_of_channel	4
}	
}	

IV. Conclusion

This paper presents broadcasting channel bonding and signaling methods to transmit UHDTV program over multiple broadcasting channels. 4-bit CC in TS packet header is used to combine multiple broadcasting channels. For signaling service information about UHDTV program, we define the UHD program descriptor which is included in the PMT.

The proposed broadcasting channel bonding and signaling scheme can be adapted for UHDTV transmission system in the near future.

References

[1] SMPTE 2036-1-2009: Ultra High Definition Television - Image Parameter Values for Program Production (2009)

[2] ISO/IEC 13818-1: 2007 “Information technology - Generic coding of moving pictures and associated audio information: Systems”

[3] TTAK.KO-07.0092, “Transmission and Reception for Digital Cable 3D Broadcasting,” TTA, Sep. 2011.

[4] Joon-Young Jung, Dong-Joon Choi, Soo In Lee, and Jae-Min Ahn, “Signaling of Multi-channel for High Definition Dual-stream 3DTV Services,” in Proc. of ICCE 2012, Las Vegas, USA.

Minimizing the Bandwidth Consumption of the FlexRay Dynamic Segment

Minkoo Kang, *Student Member, IEEE,* Kiejin Park, *Member, IEEE,* Jinyoung Choi, and Man-sik Kong

***Abstract*--To minimize bandwidth consumption in the dynamic segment of the FlexRay communication systems, a frame packing algorithm that allows the packing of signals with different periods into a message frame is proposed. The performance of the proposed algorithms with existing algorithms is evaluated using the SAE benchmark data and GNU linear programming kit. The experimental results show that the bandwidth consumption of the proposed algorithms is less than that of existing frame packing algorithms.**

I. INTRODUCTION

Nowadays, automobiles are going to have all the capabilities that are typical of a computer or entertainment equipment. As automobiles get increasingly complex they become more like consumer electronics. As customer requirement with regard to safety, convenience, and environment protection in automobiles are increasing, the automotive industry is introducing and installing many electronic devices and software in automobiles. Because the high amounts of signal data from many electronics and software need to be properly managed, the design of an in-vehicle network is becoming increasingly important [1].

In-vehicle network protocols can be classified into two paradigms: time-triggered and event-triggered. Time-triggered communication provide a higher dependability and predictability. Event-triggered communication provide a higher flexibility and extensibility [2]. FlexRay protocol combining both the advantages of time-triggered and event-triggered paradigms was proposed. It has emerged as the *de facto* standard for automotive communication systems [3].

The FlexRay communication cycle consists of static segment (ST), dynamic segment (DYN), symbol window (SW) and network idle time (NIT). ST is defined for time-triggered communication and it consists of static slots (STS). DYN is defined for event-triggered communication, and it consists of minislots (MS). Message transmission in ST is based on time division multiple access (TDMA) and DYN is based on flexible TDMA (FTDMA). The FlexRay message frame consists of a header segment, a payload segment, and a trailer segment [4].

Automotive signals are transmitted over the in-vehicle network in the form of message frames. A message frame has signals as well as the fixed or variable size of the overhead, which includes information on the message frame itself and the transmission error check codes. Thus, the signals should be packed into the message frames, commonly called *frame packing*, for minimizing bandwidth consumption.

Several frame packing approaches for the FlexRay communication systems have been proposed. To reduce message response time of FlexRay, a scheduling algorithm that long static message is assigned to DYN was proposed [5]. However, neither a mathematical model nor time complexity of the algorithm were considered. In [6], to ensure reliability of frames in the presence of faults, a frame packing method that computes the required number of frame retransmissions has been proposed. Using an integer linear programming model, frame packing algorithm for minimizing bandwidth consumption of the FlexRay ST has been provided in [7]. However, the model has a drawback that signals having different periods cannot be packed into the same message frame.

In this paper, a frame packing algorithm to allow frame packing with signals having different periods and minimize bandwidth consumption of the FlexRay dynamic segment is proposed. Section II describes system model and the proposed algorithm. Next, the performance evaluation is carried out in Section III. Section IV presents the conclusions.

II. SYSTEM MODEL

The FlexRay communication system we consider in this paper consists of one FlexRay bus and N FlexRay electronic control units (ECUs). For this system model, we make the following three assumptions: 1) there are no failures and transmission errors in the FlexRay communication system; 2) the period, deadline, and size of all signals are known; 3) only signals from the same ECU are packed into the same message frame, and each signal period has to be an integer multiple of the period of the message frame.

A signal s_i is characterized by $\{e_i, p_i, d_i, b_i\}$, where e_i is the ECU, p_i is the period, d_i is the deadline, and b_i is the size in bits of a signal s_i, respectively. Similarly, a message frame m_j is characterized by $\{e_j, p_j, d_j, ms_j\}$, where e_j is the ECU, p_j is the period, d_j is the deadline, and ms_j is the size in minislots of a message m_j, respectively. Consider the following problem.

$$\min \; \textstyle\sum_j y_j \cdot p_j^{-1} \cdot N_{ms}^{-1} \cdot \left(ms_j + p_j - 1\right) \qquad (1)$$

$$\text{s.t.} \; \textstyle\sum_j x_{ij} = 1, \qquad \forall i = 1, \cdots, n_s, \qquad (2)$$

$$ms^L \le ms_j \le ms^U, \qquad \forall j = 1, \cdots, n_m, \qquad (3)$$

$$\textstyle\sum_i x_{ij} / n_s \le y_j \le \sum_i x_{ij}, \qquad \forall j = 1, \cdots, n_m, \qquad (4)$$

$$x_{ij} \in \{0,1\}, \qquad \forall i = 1, \cdots, n_s, \quad \forall j = 1, \cdots, n_m, \qquad (5)$$

$$y_j \in \{0,1\}, \qquad \forall j = 1, \cdots, n_m, \qquad (6)$$

This study was supported by the R&D Center for Valuable Recycling (Global-Top Environmental Technology Development Program) funded by the Ministry of Environment (Project No.:12-A32-MD).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$y_j \cdot p_j^{-1} \cdot N_{ms}^{-1} \cdot (ms_j + p_j - 1)$ represent the bandwidth consumption for a transmitting message frame m_j where N_{ms} is the number of minislots in the DYN. ms_j in minislots can be calculated by the following equation.

$$ms_j = \left\lceil \left(20 \cdot \left\lceil \sum_i x_{ij} b_i / 16 \right\rceil + O_F\right) gdBit / T_{ms} \right\rceil \qquad (7)$$

The binary variables $x_{ij} = 1$ means that the signal s_i is packed into the message frame m_j, otherwise $x_{ij} = 0$. That is,

$$x_{ij} = \begin{cases} 1, & \text{if } e_i = e_j, \gcd(p_i, p_j) = p_j, \text{and the signal } s_i \\ & \text{is packed into the message frame } m_j \\ 0, & \text{otherwise.} \end{cases} \qquad (8)$$

$\sum_j x_{ij} = 1$ implies that each signal has to be packed into only one message frame. n_s and n_m represent the number of signals and message frames, respectively. ms^L and ms^U represent the lower and upper bounds of ms_j, respectively. Using the binary variables x_{ij}, we can count the number of all message frames. Hence, a binary variable y_j is introduced where $y_j = 1$ means that at least one signal is packed into the message frame m_j, otherwise $y_j = 0$. The equation (4) is given to calculate the value of y_j using the values of x_{ij}.

III. Performance Evaluation

To evaluate the performance of the proposed algorithm proposed in this paper, the Society of Automotive Engineers (SAE) benchmark data [8] are used. The SAE benchmark data have 31 sporadic signals used for class C application in automotive communication systems. The signals have information on the transmission periods, deadlines, transmitting and receiving ECUs, and data size in bits as shown in Table I.

TABLE I
Sporadic Messages in the SAE Benchmark

ECU	Period/Deadline (ms)	Deadline (ms)	Size (bits)	# Signals
1	50	5	4	1
1	50	20	1	4
2	20	20	1	1
3	50	20	1	5
3	50	20	2	5
3	50	20	3	5
4	50	20	1	4
4	50	20	2	1
4	50	20	8	1
5	50	20	1	8
5	50	20	2	1
5	50	20	7	1
5	50	20	8	1

In recent years, more than 70 ECUs exchange around 2500 signals in luxury cars [1]. Hence, the signal set has to be extended to a relevant size. Our approach to extend the set of signals involves randomly choosing signals. In addition, the GNU linear programming kit [8] is applied to solve the ILP problems in this experiment.

Figure 1 shows the change in bandwidth consumption during signal demand $SD = 0.02$ to $SD = 0.7$. The results show that bandwidth consumption of the proposed algorithm ($BC_{proposed}$) is less than bandwidth consumptions of the algorithm in [7] ($BC_{Schmidt}$) and the one signal per frame (OSpF) approach (BC_{OSpF}.) because the proposed algorithm allows the frame packing of signals having different periods.

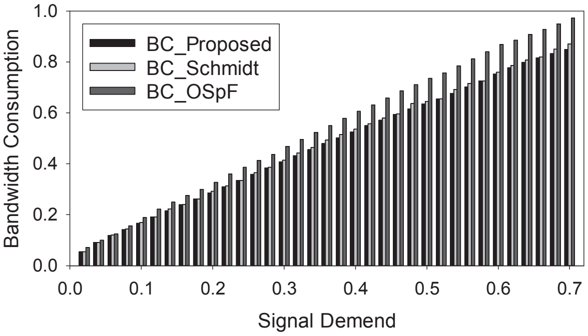


Fig. 1. Comparison of Bandwidth Consumption with respect to Signal Demand

IV. Conclusion

In this paper, we presented the frame packing algorithm by using a new ILP formulation for minimizing the bandwidth consumption in the dynamic segment of the FlexRay communication systems. The proposed algorithm allows the frame packing of signals having different periods. To evaluate the performance of the proposed algorithms, SAE benchmark data are used. The experimental results show that the bandwidth consumption of the proposed algorithm is considerably reduced compared to existing frame packing. In future work, we may need to extend our model to cases of ECU failures or transmission errors. Furthermore, we will conduct research on a message scheduling algorithm for FlexRay dynamic segment.

Reference

[1] N. Navet, Y. Song, F. Simonot-Lion, and C. Wilwert, "Trends in Automotive Communication Systems," *Proceeding of the IEEE,* Vol. 93, No. 6, pp. 1204-1223, June 2005.

[2] R. Obermaisser, *Event-Triggered and Time-Triggered Control Paradigms,* Sptinger-Verlag, Dec. 2004.

[3] R. Shaw and B. Jackman, "An Introduction to FlexRay as an Industrial Network," *International Symposium on Industrial Electronics (ISIE 2008),* July 2008, pp. 1849-1854.

[4] FlexRay. Flexray Communications System Protocol Specification, ver. 2.1, revision a. [Online]. Available: http://www.flexray.com

[5] Kiejin Park, Minkoo Kang, and Bongjun Kim, "A Scheduling Algorithm for Reducing FlexRay Message Response Time Using Empty Minislots in Dynamic Segment," ICCE 2010.

[6] B. Tanasa, U.D. Bordoloi, P. Eles, Z. Peng, "Reliability-Aware Frame Packing for the Static Segment of FlexRay," Proceedings of the ninth ACM international conference on Embedded software, pp. 175-184, Oct. 2011.

[7] K. Schmidt and E. G. Schmidt, "Message Scheduling for the FlexRay Protocol: The Static Segment," IEEE Transactions on Vehicular Technology, Vol. 58, pp. 2170-2179, 2009.

[8] GNU Linear Programming Kit, [Online]. Available: http://www.gnu.org/software/glpk

Self-mixed Interference Cancellation Method in Direct Conversion Receivers

Moonchang Choi and Sooyong Choi
School of Electrical & Electronic Engineering, Yonsei University, Seoul, Korea

***Abstract*—A novel method for cancelling the time-varying offset due to self-mixed interferences in direct conversion receivers is proposed. The proposed method is based on the adaptive interference cancellation using the least mean square adaptation algorithm. Some simulation results show that the proposed method cancel the self-mixed interference almost perfectly.**

I. Introduction

Though the direct conversion (also known as *zero-intermediate frequency (IF)* or *homodyne*) receiver is a low-cost and low-power radio receiver architecture compared with the super-heterodyne receiver which is more commonly used in communication systems, it still has some critical drawbacks [1]. Among them, the self-mixing problem which is generated in the mixer is one of the important practical problems in the direct conversion receiver (DCR) [2].

The offset signal caused by the self-mixing problem can be classified into two types [2]. One is the time-invariant offset (also well-known as *direct current (DC) offset*) caused by local oscillator (LO) leakages and the other one is the time-varying offset caused by interference leakages.

There have been several methods to mitigate offset signals caused by self-mixing problems [3],[4]. The cancellation method in [3] can be effective in mitigating the time-invariant offset while it cannot eliminate the time-varying offset which is varying fast over time [4]. In [4], the authors proposed a cancellation method to mitigate the time-varying offset signals. However, the cancellation methods in [4] can mitigate the time-varying offset only when the offset is the periodic signal.

In this paper, we propose a novel method for cancelling the time-varying offset due to self-mixed interferences in DCRs, regardless of the periodicity of the time-varying offset.

II. Self-mixed Interference cancellation

A. Self-mixing problem caused by the interference leakage

Consider the received radio frequency (RF) signal, $x_{RF}(t)$, given in [4] as

$$x_{RF}(t) = \text{Re}[(s(t) + i(t)e^{j2\pi f_o t})e^{j2\pi f_c t}], \tag{1}$$

where $s(t)$ is the baseband equivalent desired signal and $i(t)$ is the baseband equivalent interference signal which is located f_o away from the desired signal.

For ideal DCRs which have not any interference leakages, the down-converted and ideal low pass filtered baseband signal can be written in [4] as $x_{base}(t) = A_{LNA}s(t)$ where A_{LNA} is the gain of the low noise amplifier (LNA) in front of the mixer.

In practice, however, there are some interference leakages in DCRs. For practical DCRs which suffer from interference leakages, the baseband signal can be written in [4] as

$$x_{base}(t) = A_{LNA}s(t) + {A_{LNA}}^2(K_I + K_Q)(|s(t)|^2 + |i(t)|^2), \tag{2}$$

where K_I and K_Q are the attenuation factors of interference leakages for inphase and quadrature channels, respectively [4].

It can be known from (2) that the interference leakage generates time-varying offset on the desired signal band. Thus, the desired signal $s(t)$ is severely distorted by the squared-envelope of the interference.

B. Proposed interference cancellation method

The proposed receiver structure to remove the self-mixed interference is illustrated in Fig. 1. The configurations of low pass filters (LPFs) which are used in the proposed receiver structure are presented in Fig. 2.

In order to remove the self-mixed interference, we need a reference signal which is correlated with the self-mixed interference and is not correlated with the desired signal. In the following, we present the procedure of obtaining the reference signal from the received signal itself. As shown in Fig. 2(a), the received signal after passing through the LPF 1 is composed of the desired baseband signal and the interference signal which has a carrier frequency offset of f_o. From the discrete complex signal after passing through the analog-to-digital converter (ADC), we can obtain the discrete signal $d(n)$ using the LPF 2 configured in Fig. 2(b). Also, we can obtain another discrete signal $v(n)$ by taking the LPF 3 configured in Fig. 2(c) and the square of the absolute value.

Finally, in the proposed receiver structure, two signals which are needed for adaptive interference cancellation (IC), $d(n)$ and $v(n)$ can be obtained as

$$d(n) = K_1 s(n) + K_2|i(n)|^2 + i_d(n), \tag{3}$$

$$v(n) = K_3|i(n)|^2 + i_v(n), \tag{4}$$

, respectively, where

$$K_1 = A_{LNA}, K_2 = (K_I + K_Q)A_{LNA}^2, K_3 = A_{LNA}^2, \tag{5}$$

$$i_d(n) = (K_I + K_Q)A_{LNA}^2|s(t)|^2, \tag{6}$$

$$\begin{gathered} i_v(n) = 4A_{LNA}^4(K_I^2 + K_Q^2)(\text{Re}\{i(t)s^*(t)\})^2 \\ +4A_{LNA}^3\text{Re}\{i(t)s^*(t)\}(K_I\text{Re}\{i(t)\} + K_Q\text{Im}\{i(t)\}). \end{gathered} \tag{7}$$

The reference signal $v(n)$ is highly correlated with the interference component of $d(n)$. As shown in the adaptive

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

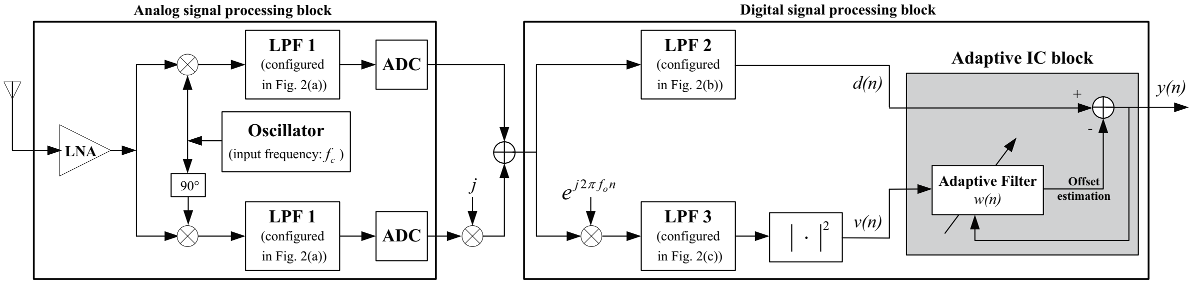

Analog signal processing block
Digital signal processing block
LNA
LPF 1 (configured in Fig. 2(a))
ADC
Oscillator (input frequency: f_c)
90°
LPF 1 (configured in Fig. 2(a))
ADC
j
LPF 2 (configured in Fig. 2(b))
$e^{j2\pi f_o n}$
LPF 3 (configured in Fig. 2(c))
$|\cdot|^2$
$v(n)$
$d(n)$
Adaptive IC block
Adaptive Filter $w(n)$
Offset estimation
+
-
$y(n)$

Fig. 1. Proposed receiver structure

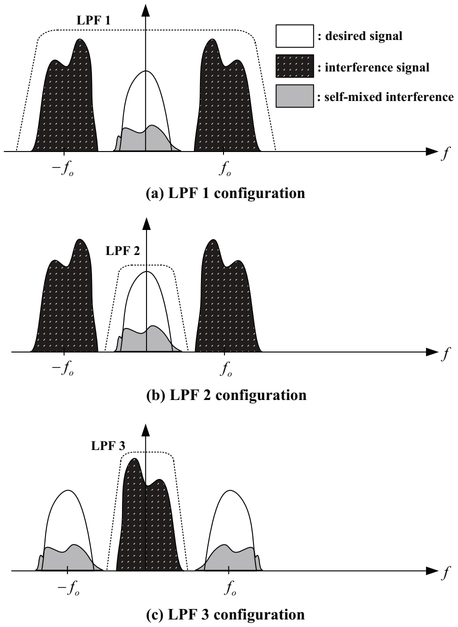


Fig. 2. Filter configuration of proposed receiver structure

IC block in Fig. 1, the adaptive IC method subtracts the interference estimate $w^*(n)v(n)$ from $d(n)$, which contains a desired signal component and an interference component [5]. The output of adaptive IC block can be written as $y(n) = d(n) - w^*(n)v(n)$. The adaptive filter coefficient $w(n)$ is updated according to the error signal $y(n)$, which is also the output of the whole IC structure. The filter adaptation can be done with any algorithms, such as the well-known least mean square (LMS) algorithm. In this paper, we use the LMS adaptation algorithm which is given by $w(n+1) = w(n) + \mu v(n)y^*(n)$ where μ is the step size for the LMS algorithm.

III. Numerical Results

Fig. 3 shows the bit error rate (BER) performance of the proposed method when the interference leakage is 30dB smaller than the desired signal. In this simulation, the interference leakage factors (K_I, K_Q) are 10^{-4} and the LNA gain (A_{LNA}) is 10^{-2}. For both the desired and interference signals, quadrature phase shift keying (QPSK) modulation is used. In order to evaluate the BER performance of the proposed method, the BER performances for no interference case and no interference cancellation case are also represented. The BER performance of the proposed method is almost the same as that of no interference case. It is shown that the proposed method can remove the self-mixed interference almost perfectly.

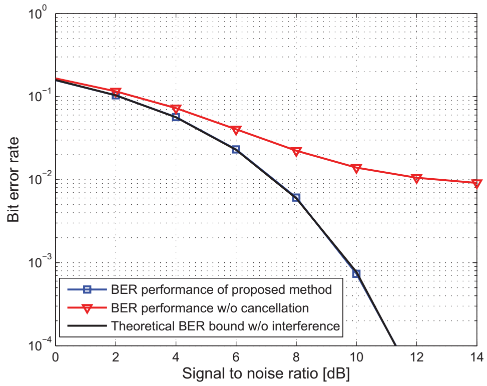


Fig. 3. BER performance of proposed method

IV. Conclusions

In this paper, we proposed the novel method for cancelling the self-mixed interference in DCRs. The proposed method cancel the self-mixed interference almost perfectly, regardless of the periodicity of the self-mixed interference.

Acknowledgment

This research was supported by the KCC(Korea Communications Commission), Korea, under the R&D program supervised by the KCA(Korea Communications Agency) (KCA-2012-(11-911-04-004)).

References

[1] B. Razavi, RF Microelectronics. Upper Saddle River, NJ: Prentice Hall PTR, 1998.

[2] B. Razavi, "Design considerations for direct-conversion receivers," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 428-435, June 1997.

[3] R. Magoon, A. Molnar, J. Zachan, G. Hatcher, and W. Rhee, "A single-chip quad-band direct conversion GSM/GPRS RF transceiver with integrated VCOs and fractional-N synthesizer," *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1710-1720, 2002.

[4] S. B Park and M. Ismail, "DC offset in direct conversion multistandard wireless receivers: Modeling and cancellation," *Analog Integrated Circuits and Signal Processing*, vol.49, no. 2, pp. 123-130, 2006.

[5] S. Haykin, Adaptive filter theory, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

A Modified MIMO with High Data Rates for Set-Top Box in Single Carrier System

Bong Gyun, Jo, *Student Member, IEEE* and Dong Seog Han, *Member, IEEE*

***Abstract*— In this paper, we propose a modified MIMO transmission system structure with high data rates for next generation television set-top box in single carrier system.**

I. Introduction

The next-generation full 3DTV and UHDTV needs sophisticated transmission technologies with the maximum data rate of about 30~60 Mbps. So, it is difficult to support the next generation broadcasting system using (ASTC) standard or DVB-T2 with the transmission rate of 19.39 Mbps. And DVB-T2 system uses OFDM that has PAPR problem. It reduces the bit error rate (BER) performance due to nonlinear distortion of high power amplifier (HPA). But the deployment of the 3DTV or UHDTV service using terrestrial transmission may be possible thanks to recent communication techniques such as multi-input multi-output (MIMO) and low density low-density parity-check code (LDPC) channel codes. And we consider single carrier system for solving the problem of PAPR.

In this paper, we propose a MIMO transmission structure for the next generation broadcasting system through terrestrial transmission in single carrier system. Practical modulation schemes for MIMO systems are decoupled into two areas known as diversity and multiplexing. Diversity modulation or space time coding (STC) uses specially designed codewords that maximize the diversity gain at the expense of a loss in capacity. On the other hand, spatial multiplexing or Bell labs layered space-time (BLAST) system transmits independent data streams to each transmitting antenna. The spatial multiplexing allows the capacity to be improved at the expense of the loss of diversity gain. The most typical STC method, space time block code (STBC), transmits data by utilizing orthogonality between signals. The received data can be decoupled and then the transmitted symbols are obtained by using the maximum likelihood (ML) detection. On the other hand, the linear dispersion code (LDC) method is that the basis matrices are chosen such that the resulting codes maximize the capacity of the MIMO system. These basic matrices are multiplied to the transmission matrix, then the transmitted symbol having information of all data in the transmit matrix is detected by ML detection at the receiver [1][2][3].

In this paper, LDC is considered to achieve high data rates for the next generation broadcasting system. And we propose the structure that changes the LDC scheme in time domain to frequency domain for applying single carrier system.

II. Conventional MIMO system

Spatial multiplexing architecture breaks the original data stream into sub-streams that are transmitted on individual antennas. So, the data rate is increased in proportion to the number of transmit antennas. However, when received signal is detected by ZF and the minimum mean square error (MMSE) algorithm, the BER performance is decreased unlike the coded STC. The LDC scheme is proposed to overcome a defect of spatial multiplexing method.

The modulated symbols s_q, $q = 1, \ldots, Q$, are modified to a block S by linear dispersion matrices A_q and B_q as [3]

$$s_q = \alpha_q + j\beta_q,\ q = 1, \ldots, Q \tag{1}$$

$$S = \sum_{q=1}^{Q} (\alpha_q A_q + j\beta_q B_q) \tag{2}$$

$$Q = \min(M, N) \times T \tag{3}$$

There are Q different symbols in a signal transmission matrix S. T is the number of columns in S. Linear dispersion matrices are defined as

$$A_{M(k-1)+l} = B_{M(k-1)+l} = \frac{1}{\sqrt{M}} D^{k-1} \Pi^{l-1} \qquad k = 1, \ldots, M,\ l = 1, \ldots, M \tag{4}$$

where

$$D = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & e^{j\frac{2\pi}{M}} & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & \ldots & e^{j\frac{2\pi(M-1)}{M}} \end{bmatrix} \text{and}\ \Pi = \begin{bmatrix} 0 & \ldots & & 0 & 1 \\ 1 & 0 & \ldots & & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \ldots & & \ldots & & \ldots \\ 0 & \ldots & 0 & 1 & 0 \end{bmatrix}$$

If we assume that the symbols, s_k, s_{k+1}, s_{k+2}, and s_{k+3} are transmitted by applying the linear dispersion matrix for M=2and Q=4, the LDC coded symbols can be expressed as

$$\begin{aligned} S = ((&\alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3 + \alpha_4 A_4) \\ &+ j(\beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + \beta_4 A_4)) \end{aligned} \tag{5}$$

The received signals from the channel can be expressed as

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

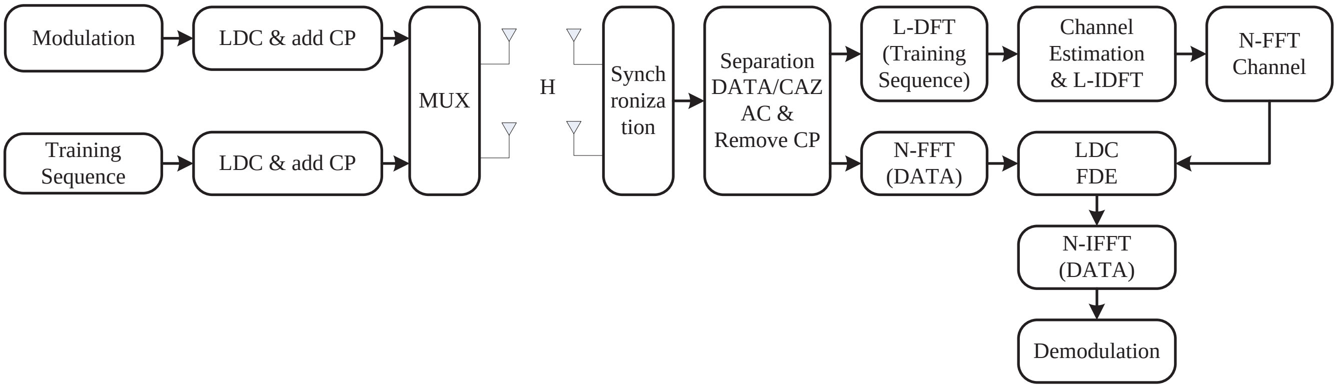


Fig. 1. Proposed modified MIMO structure in single carrier system

$$\begin{bmatrix} y_{R,1} \\ y_{I,1} \\ \cdots \\ y_{R,N} \\ y_{I,N} \end{bmatrix} = \sqrt{\frac{\rho}{M}} H \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \cdots \\ \alpha_Q \\ \beta_Q \end{bmatrix} + \begin{bmatrix} n_{R,1} \\ n_{I,1} \\ \cdots \\ n_{R,N} \\ n_{I,N} \end{bmatrix} \tag{6}$$

where y_R and y_I are real and image components of received signal, n_R and n_I those of white noise, ρ is the signal-to-noise ratio (SNR) and H is a changed channel matrix. Then, the transmitted signals can be detected by ML detection with the received signals. We apply this LDC algorithm to single carrier system.

III. Proposed Positioning Scheme

It is difficult to use MIMO algorithm in single carrier system, because of estimation or equalization in time domain. So, we transmit the signals in time domain and estimate the channel in frequency domain. The proposed modified MIMO transmission structure is illustrated in Fig. 1. We use LDC algorithm and CAZAC sequence for channel estimation. In the receiver, we can estimate the channel using L-length CAZAC sequence, and then we detect N signals in frequency domain.

CAZAC sequence maintains good performance in time and frequency domain. For detecting the LDC signals in frequency domain, we need some algorithm that converts time domain LDC signals to frequency domain signals. So, we consider Fast Fourier Transform (FFT) characteristic for converting the LDC signals between time and frequency domain.

IV. Conclusion

In this paper, a modified structure of MIMO algorithm is proposed in single carrier system. It is a way to solve PAPR of OFDM system and improve the data rates for next generation broadcasting system. And it is important to find the training sequence that maintains good performance in time and frequency domain.

Acknowledgment

This research was supported by the KCC (Korea Communications Commission), Korea, under the "Development of highly efficient transmission technology for the next generation terrestrial 3D HDTV" support program supervised by the KCA (Korea Communications Agency)" (KCA-2012-10912-02002).

References

[1] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun*., vol. 17, pp. 1451-1458, Oct. 1998.

[2] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, “VBLAST: An Architecture for Realizing Very High Data Rates Over the Rich-Scattering Wireless Channel,” *in Proc. ISSE*, Pisa, Italy, September 1998.

[3] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1804-1824, Jul. 2002.

Protection Ratios for Interference Between ATSC Signals (Invited)

Charles Rhodes

The minimum usable received signal power depends on the robustness to undesired signals by deployed receivers. Tests of 26 NTIA approved "converter boxes" show that these are subject to de-sensitization by a signal offset from the desired signal by 1 to 15 channels.

Dynamic Voltage and Frequency Scaling Over Delay Constrained Mobile Multimedia Service

Jihyeok Yun(1), Kyungmo Park(2) and Doug Young Suh(1)
(1)Kyunghee University – Yongin, Korea, (2)SAMSUNG Electronics Co., Ltd – Seoul, Korea

***Abstract*—This paper proposes an application of dynamic voltage and frequency scaling (DVFS) which is used in the power saving technique to process video decoding which requires much energy consumption in handheld devices having power limitations. The proposal of this paper is a DVFS application method to prevent an execution delay while applying DVFS and the complexity estimation method. The proposed method can achieve energy gains in power-limited and quality-sensitive video decoding settings by providing a computationally inexpensive power management technique.**

I. INTRODUCTION

The requirement for video quality of video service in handheld devices has increased significantly. As a result, there has been a problem in maintaining the quality of video service in a way that satisfies the consumer requirements. This paper aims to provide an energy saving method that can reduce power consumption in power-limited handheld devices without increasing the delay or complexity associated with power savings. Dynamic Voltage and Frequency Scaling (DVFS) is employed for this purpose. DVFS is a method of power consumption reduction for processors that adjust applied voltage to processors dynamically.

According to [1] and [2], the power consumption of a processor is proportional to the square of supply voltage, and supply voltage is proportional to frequency. Based on these relationships, power consumption can be reduced by adjusting voltage and frequency accordingly. After an estimation of complexity—which is required for decoding—is done, voltage and frequency can be applied to the estimated complexity accordingly.

Though [1] proposed an appropriate complexity model by modeling the complexity of video frames (which added a complexity profiler to the video decoder), this model actually increased complexity due to the additional profiler, and did not consider a frame drop phenomenon that could be caused by an estimation error.

[3] proposed an algorithm of finding the optimum combination of frequency and voltage in which the decoding slack time is 0 while decoding time information is stored by the frequency and voltage applied to a processor. However, this algorithm has to store decoding information between certain periods, and cannot prevent a frame drop phenomenon since an estimation of the next frame is calculated after taking into account the overhead due to an estimation error.

Unlike the methods suggested in [1] and [3], the method proposed in this paper does not require a profiler or a computing unit for decoding complexity estimation. The proposed method introduces a simple but effective scheme to estimate the decoding complexity (or the decoding time) of video frames. In addition, the proposed method employs a delay (jitter) threshold for frequency scaling to prevent frame drops occurring in DVFS.

Our proposed estimation method does complexity estimation without an extra profiler or calculation by using the characteristic of multimedia contents, which requires repeatedly processing similar calculations. For video content, since coded frame type is the same and frames are closer in terms of time, the similarity becomes greater. Therefore, it is effective to perform voltage and frequency scaling by taking advantage of the complex information of frames that are decoded most recently, and the same coded frame types requiring no extra calculation processing. In addition, our proposed method sets the limited delay bound of the delays caused by an estimation error during estimation, and if the limited delay bound is exceeded, the corresponding frame is decoded with the maximum frequency of a processor. As such, if one frame is decoded with maximum frequency, the delay problem is solved but overhead still exists in terms of power. However, since the number of estimation errors is small because of the offset of adding and subtracting repeatedly, and decoding time is reduced significantly when decoded with the maximum frequency, the related overhead is very minimal. In this way, the proposed DVFS method can prevent frame drops caused by estimation errors in DVFS and decrease decoding energy consumption in real-time multimedia services where delays are not allowed.

Section II contains the proposed method for complexity estimation and delay control.

Section III contains the experimental data of the proposed method.

II. PROPOSED METHOD

A. Complexity estimation

The complexity estimation method proposed in this paper is performed with regard to frame types constituting video such as Intra Frame (I-frame), Unidirectionally Predicted Frames (P-frame), Bidirectionally Predictive Frames (B-frame) respectively, as shown in Fig. 1. It refers the same type of complexity information that is decoded most recently but without extra modeling or estimation algorithm.

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2012- H0301-12-1006) supervised by the NIPA (National IT Industry Promotion Agency).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

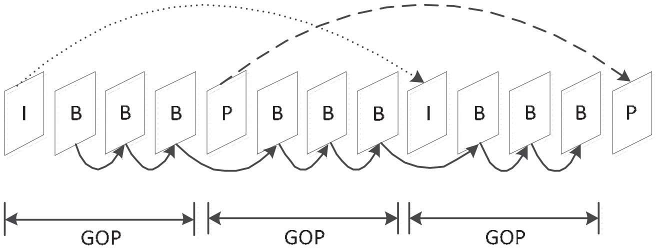


Fig. 1. The concept of proposed complexity estimation.

Since our proposed method targets video service, it takes advantage of video's characteristics that have the most similarities between frames and within close times. In our proposed method, referenced frames for complexity can be searched by using the size of Group of Pictures (GOP) and Intra period, which are parameters of video coder according to [6] and [7].

If GOP size is s and Intra period is p, then the nth expected complexity $\hat{c}[n]$ can be calculated by using (1).

$$\begin{aligned} I, P - \mathit{frame} \quad &: \hat{c}[n] = c[n-2s] \quad \mathit{for} \quad (n \% s) = 0 \\ B - \mathit{frame} \quad &: \hat{c}[n] = \begin{cases} c[n-2] & \mathit{for} \quad (n \% s) = 1 \\ c[n-1] & \mathit{Otherwise} \end{cases} \end{aligned} \tag{1}$$

In formula (1), (n % s) means (n module s).

In the estimation method proposed – when there is no anchor for estimation, such as time of decoding start or change of channel – decoding is performed by applying maximum frequency to a processor. For each frame type, after one frame is performed by performance decoding, the previous frame can act as anchor so that estimation can be carried out.

Below, Fig. 2 shows a scatter graph of actual complexity against estimated complexity for total frames (167,857 frames) of [8] which uses our proposed method.

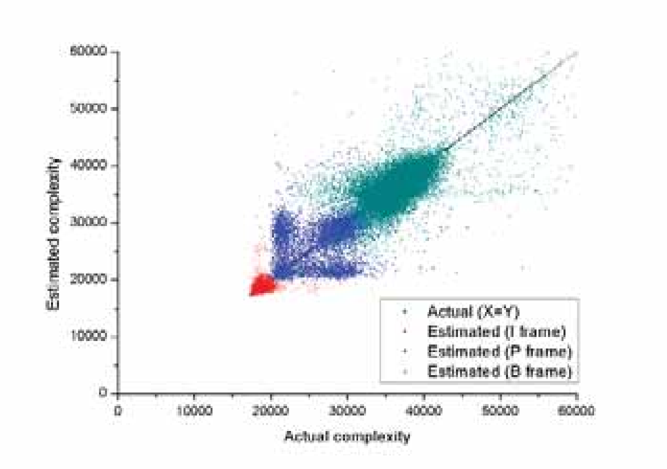


Fig. 2. The variance of error of proposed complexity estimation.

Fig. 3 shows a histogram graph of estimation error. Even though there are many scene changes generated in 167,857 frames, we can verify from Fig. 2 and 3 that neighboring frames have high similarity.

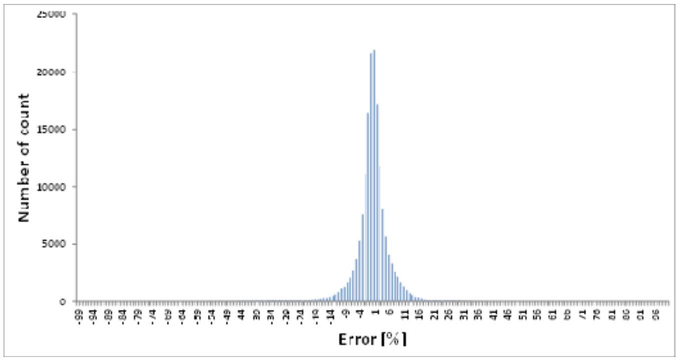


Fig. 3. The histogram of error of proposed complexity estimation.

If there are many scene changes occurred, the proposed complexity estimation method generates an estimation error the similarity between neighboring frames decreases. If the value of estimation error is larger than the actual value, it generates energy wasting, and if the estimation value is smaller than the actual value, it generates delay. As shown in Fig. 2, since estimation error shows symmetrical forms between right and left sides, estimation errors are offset. As a result, we can see that delay and power saving efficiency have a tradeoff relationship with one another.

B. *Delay control*

Regarding delay and power wasting, which are generated despite offset estimation errors while using the proposed method, we compare the two by considering the accepted delay bound according to the characteristics of the buffer and the service. Our proposed methods for delay control is shown as Fig. 4 performing delay control by calculate expected decoding time $\hat{t}[n]$ by using (2).

$$\hat{t}[n] = t_{slack} - D[n-1] + D_{th} \tag{2}$$

t_{slack}, $D[n-1]$ and D_{th} is frame slack time which is determined by a frame rate of video, accumulated delay of (n-1)th and delay threshold. Accumulated delay is calculated by using (3).

$$D[n-1] = \sum_{i=1}^{n-1} \frac{c[i]}{f[i]} - (n-1)t_{slack} \tag{3}$$

c[i] and f[i] is complexity and frequency of ith frame.

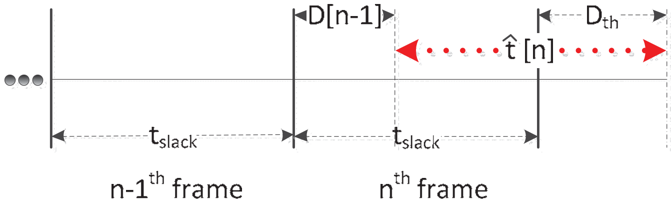


Fig. 4. The concept of proposed delay control.

Scaling frequency f_s for DVFS is calculated by using (4).

$$f_s[n] = \begin{cases} \dfrac{\hat{c}[n]}{\hat{t}[n]} & for \quad \hat{D}[n] \le D_{th} \\ f_{\max} & for \quad \hat{D}[n] > D_{th} \end{cases} \tag{4}$$

$\hat{c}[n]$, $\hat{t}[n]$, $\hat{D}[n]$ and $f_{\max}$ is expected complexity, expected decoding time, expected accumulated delay of nth frame and maximum frequency. Expected accumulated delay of nth frame is calculated by using (5).

$$\hat{D}[n] = D[n-1] + (\hat{t}[n] - t_{slack}) \tag{5}$$

If $\hat{D}[n]$ is exceeding D_{th}, proposal performs decoding with max frequency of a processor. In case of decoding with max frequency, actual delay D[n] is not exceeding D_{th} since decoding time is very short and provides a little time of application for DVFS for the next frame, overhead in terms of energy is very small.

When $f_s[n]$ is determined by (5), decoding energy according to $f_s[n]$ can be calculated by using the underlying platform coefficients of a processor as shown in [2] and equations commonly used in [1], [4] and [5]. According to [1], [4] and [5], decoding energy is calculated by supply voltage and supply voltage is calculated by frequency with underlying coefficients. Therefore we can calculate decoding energy that uses the proposed method by scaling frequency $f_s[n]$.

III. Simulation

In this paper, we use the underlying coefficients of Intel Pentium Mobile Processor 1.6 GHz, and 167,857 frames of video [8] (DVD ver. 720 by 480 pixels quality) decoded by H.264/AVC reference software version 18.3.

To display a raw file that is the output of the decoder, it is necessary to convert the raw file to an RGB file. In the conversion, pixel-based real-valued calculations are performed. In a general scenario where successive frames have the same resolution, the complexity of frames with regard to the conversion does not vary.

Even though the complexity of frames associated with the conversion is constant, the conversion is influenced by processor frequency scaling in DVFS. That is, the energy consumed in the conversion of different frames can vary. This characteristic was taken into account in the experiments.

A. Complexity estimation

The first simulation compares decoding energy consumption between a method not using DVFS, [1] using complexity modeling, and our proposed method. This simulation uses a science fiction action movie [8], which is the worst simulation environment for our proposed method. We assume that the estimation error of the comparison counterpart method [1] will be 0%, which is the best estimation.

As shown in Fig. 5, the proposed method saved 73% of the energy used for decoding compared to a conventional non-DVFS method. This performance was almost the same as those of previous methods [1] whose control algorithms were much more sophisticated than ours.

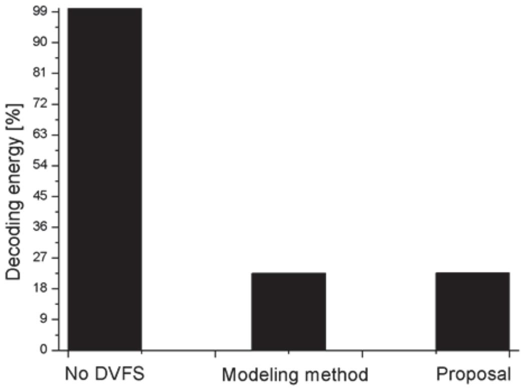


Fig. 5. Performance comparison between a modeling method and the proposed method.

B. Delay control

The second simulation compares energy consumption while a method for the prevention of frame drop—which is generated by the estimation error of DVFS to support delay constrained service— is used. In the case of [1], since it does not consider delay, in order to overcome a 3% estimation error, DVFS shall be performed with a margin of 3% complexity estimation. Our proposed method overcame frame drop by setting the delay threshold as a buffer. Threshold values for delay, D_{th} are set as three values, such as 0.01s (=10ms), 0.1s (=100ms), and 1s (=1000ms).

The result of the second simulation shows that energy efficiency improves 1.94% in the case of threshold being 0.01 second; 1.96% in the case of 0.1 second; and 2% in the case of 1 second, respectively, compared to one using [1], as shown in Fig. 6.

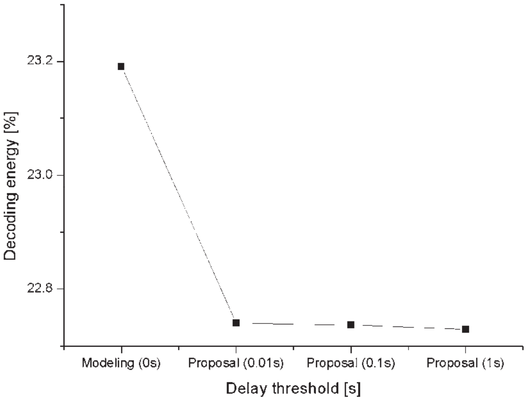


Fig. 6. Performance comparison between modeling method and the proposed method (0.01s, 0.1s, 1s delay threshold) with overhead of estimation error.

IV. Conclusion

This paper proposes an estimation method that takes advantage of similarities between video frames. It is in contrast with other methods that use estimation by calculation.

This method should have larger estimation errors compared to the ones using existing methods. However, the proposed DVFS method yields energy savings as good as conventional DVFS methods that have more sophisticated, and thus more computationally expensive, complexity estimation algorithms. The proposed DVFS method with a lightweight complexity estimation scheme and a delay threshold is suitable for use in a delay-constrained, real-time multimedia service in which computationally expensive classical DVFS methods are not applicable.

In this paper, we have discussed DVFS, which only considers video decoding. However, our future research may develop further energy saving methods that can be used in the overall video service system such as video data receiving, memory access, video decoding, and video display, in terms of client points of view.

REFERENCES

[1] Z. Ma, H. Hu, Y. Wang, “On complexity modeling of H.264/AVC video decoding and its application for energy efficient decoding.” IEEE Transactions on Multimedia, vol. 13, pp. 1240-1255, December 2011.

[2] Intel Pentium Mobile Processor. [Online]. Available: http://www.intel.com/design/intarch/pentiumm/pentiumm.htm.

[3] Y. Seo, K. Park, “A Window-Based DVS Algorithm for MPEG Player”, Journal of KIISE: Computer systems and Theory, vol. 24, no. 11, December 2008.

[4] R. Jejurikar, C. Pereira, “Leakage aware dynamic voltage scaling for real-time embedded systems”, Proceeding of the Design Automation Conference, June 2004.

[5] S. Martin, K. Flautner, “Combined Dynamic Voltage Scaling and Adaptive Body Biasing for Optimal Power Consumption in Microprocessors under dynamic Workloads.” Proceedings of the International Conference on Computer Aided Design, November 2002.

[6] T. Wiegand, G. J. Sulivan, “Overview of the H.264/AVC video coding standard”, IEEE Transactions on Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560-576, July 2003.

[7] A. M. Tourapis, “H.264/MPEG-4 AVC Reference Software Manual”, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-X072, Geneva, Switzerland, July 2007.

[8] Warner bros. entertainment, “The Matrix”, Harrison & Company, 1999.

Salient Object Detection based on Spatiotemporal Attention Models

Ruxandra TAPU and Titus ZAHARIA

***Abstract*- In this paper we propose a method for automatic detection of salient objects in video streams. The movie is firstly segmented into shots based on a scale space filtering graph partition method. Next, we introduced a combined spatial and temporal video attention model. The proposed approach combines a region-based contrast saliency measure with a novel temporal attention model. The camera/background motion is determined using a set of homographic transforms, estimated by recursively applying the RANSAC algorithm on the SIFT interest point correspondence, while other types of movements are identified using agglomerative clustering and temporal region consistency. A decision is taken based on the combined spatial and temporal attention models. Finally, we demonstrate how the extracted saliency map can be used to create segmentation masks. The experimental results validate the proposed framework and demonstrate that our approach is effective for various types of videos, including noisy and low resolution data.**

I. INTRODUCTION

Visual attention techniques provide a methodology for semantic context understanding in both images and videos. Extracting the informational content included in images/videos can be used in various computer vision applications such as: image retrieval, automatic cropping, object tracking/detection, segmentation, compression…

One basic principle in the human visual system is to suppress the response to frequently occurring input patterns, while at the same time being sensitive to novel features. Image/video information in this case can be regarded as a redundancy plus a sparse contribution. Inspired by this insight, also from the perspective of cognitive science, the informational content can be decomposed into two parts [1]: redundancy which denotes the information with high regularities of the visual inputs and saliency which represent the novel part.

In this context, we propose a framework for salient object detection in video streams. First the video is temporally segmented into shots. For each detected shot, a set of representative keyframes is determined. Then, for each keyframe the salient regions are obtained by combining spatial and motion information. The main contribution introduced in this paper concerns a novel bottom-up approach for modeling the spatiotemporal attention in movies. The spatial model is developed starting from a region-based contrast measure associated to individual keyframes. The temporal model relies on interest points correspondence, geometric transforms (*i.e.* homographic motion model), motion classes estimation (using agglomerative clustering) and regions temporal consistency. Finally, the interest object is extracted with the help of GrabCut segmentation [2], which takes as input to the saliency map previously determined.

The rest of this paper is organized as follows. Section II presents and analyzes the related work for both spatial and temporal attention models. Section III introduces the novel spatiotemporal attention method proposed and details the main steps involved. The experimental results obtained are presented and discussed in details in Section IV. Finally, Section V concludes the paper and opens perspectives of future work.

II. RELATED WORK

Methods for annotating the image/video content have attracted a great deal of attention in the last few years. One of the first techniques proposed in the literature [3] uses several feature attributes such as color, intensity and orientation. The model has been shown to be successful in predicting the human focus of attention. In addition, it can be further enhanced to completely detect image/video objects. However, the associated objective function is not clearly specified and the method parameters need to be tuned by the user. A modified version of this technique is presented in [4]. Here, a graph-based principle is applied in order to highlight the regions of interest.

A technique based on a so-called spectral residual is introduced in [5]. The spectral residuum is simply defined based on the difference between the log spectrum of the given image and the averaged log spectrum of the considered image. However, the role of spectral residual in finding the salient region is not clearly defined. Moreover, in [6] authors suggest that the phase spectrum, and not the amplitude spectrum is more appropriate to determine the location of the relevant regions. A closely-related approach is introduced in [7]. Here, Gestalt features are included in order to capture the object details.The algorithms presented in [8], [9], are based mostly on Gabor or DoG filter responses. They involve numerous parameters to be tuned, such as the number and type of filters, choice of the nonlinearities and a proper normalization scheme. Such methods tend to emphasize textured areas as being salient regardless of their context.

Current methods of saliency detection generate regions that have low resolution [3], poorly defined borders [4] or are expensive to compute [5], [6]. In addition, some methods [8], [9] produce high saliency values at object edges instead of generating maps that uniformly cover the whole objects, since they fail to exploit all the spatial frequency content of the original image. Salient motion models combined with bottom-up and top-down cues can lead to an efficient visual saliency

Part of this work has been supported by the French FUI7 project 3D-LIVE.

model. In [10], authors use both low-level (such as skin color) and semantic features (*e.g.*, captions), to develop a visual attention model.

The salient motion is defined in [11] as the motion from a typical surveillance target (person of vehicle) that is opposed to other type of movements (*e.g.* camera displacements). Based on this observation, the detection algorithm is able to identify objects with a consistent motion over the time.

In [12], the objective is to extract all objects present in the video flow that respect a set of heuristic principles. The visual saliency detection algorithm introduced in [13] incorporates the motion trajectory in order to identify relevant objects.

In [14] the authors introduced a hybrid algorithm that includes stationary saliency models based on top-down and a bottom-up visual cue combined with motion information and prediction.

However, in practice the video motion can be caused by the salient objects, but also by background objects or camera movement. In this case, different types of motion need to be analyzed and appropriately taken into account.

III. SPATIOTEMPORAL ATTENTION MODEL

We start our analysis by using our previous work presented in [15] that introduces an enhanced shot boundary detection algorithm, based on the graph partition model combined with a non-linear scale-space filtering and a leap keyframe extraction procedure.

The spatiotemporal saliency model introduced in this paper is based on the stationary saliency technique so-called region-based contrast (RC) [16]. The saliency value of a region (r_k) is defined based on the color contrast to all other regions in the image (Fig. 1c):

$$S(r_k) = \sum_{r \neq r_k} w(r_i) \cdot d_r(r_k, r_i) \ , \tag{1}$$

where $w(r_i)$ is the weight of region r_i, computed as the total number of pixels included in the region while $d_r(,)$ is the color distance metric between regions defined as:

$$d_r(r_1, r_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} p(c_{1,i}) \cdot p(c_{2,j}) \cdot \delta(c_{1,i}, c_{2,j}\ ) \ , \tag{2}$$

where $p(c_{k,i})$ is the probability of the *i*-th color $c_{k,i}$ among all n_k colors in the *k* region ($k = 1{,}2$).

The salient motion is the movement that attracts the attention of a human subject. Most of the previously developed methods [11], [14] are based only on the temporal difference of adjacent frames and cannot effectively identify the salient motion. In this paper, we propose a novel temporal attention technique that combines the above-described spatial (stationary) saliency model, on the interest point (I_p) correspondences between successive video keyframes. The algorithm consists of the following steps:

Step 1*: Interest point detection and matching* – The image is segmented into regions based on the mean shift technique [17]. The Scale Invariant Feature Transform (SIFT) [18] is applied on two successive frames (by taking as starting frame each detected keyframe). The correspondence between the interest points is established using KD-tree matching technique [19] (Fig. 1b).

Let $p_{1i}(x_{1i}, y_{1i})$ be the *i*-th key point in the first image and $p_{2i}(x_{2i}, y_{2i})$ be the correspondence in the second image. The associated motion vectors (v_{ix}, v_{iy}), magnitude ($D_{i(1,2)}$) and angle of motion ($\theta_{i(1,2)}$) are also computed in this step:

$$v_{ix} = x_{2i} - \ x_{1i}\ ; v_{iy} = y_{2i} - \ y_{1i}, \tag{3}$$

$$D_{i(1,2)} = \ \sqrt{v_{ix}^2 + v_{iy}^2}\ \ , \ i = \overline{1, n}, \tag{4}$$

$$\theta_{i(1,2)} = acos\frac{v_{ix}}{D_{i(1,2)}}, \theta \in [0{,}2\pi] \tag{5}$$

where *n* is the total number of correspondences.

Step 2*: Interest points saliency initialization* –The interest point's spatial saliency values are determined based on the region based contrast (Fig. 1c).

Step 3*: Background / Camera motion detection* –We start our analysis by identifying a subset of *m* keypoints located in the background (Fig. 1c) and selected based on their saliency value. More precisely, an interest point $p_{1,i}$ is defined as a background point if:

$$Sal(p_{1,i}) \leq T_h \ , \tag{6}$$

where $Sal(p_{1,i})$ is the saliency value of point $p_{1,i}$ while T_h is the average saliency value over the considered key-frame.

The subset of *m* background interest points is used to determine the geometric transformation between the selected images, by considering a homographic motion model, determined with the help of a RANSAC (*Random Sample Consensus*) [20] algorithm. Based on the optimal homographic matrix ***H*****,** for a current point $p_{1i} = [x_{1i}, y_{1i}, 1]^T$ expressed in homogeneous coordinates, its estimated correspondence position $p_{2i}^{est} = [x_{2i}^{est}, y_{2i}^{est}, 1]^T$ is determined as:

$$\begin{bmatrix} x_{2i}^{est} \\ y_{2i}^{est} \\ w \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \cdot \begin{bmatrix} x_{1i} \\ y_{1i} \\ 1 \end{bmatrix}, \tag{7}$$

where:

$$w = \ 1/(h_{20} \cdot x_{2i}^{est} + h_{21} \cdot y_{2i}^{est} + h_{22}). \tag{8}$$

The estimation error is defined as the difference between the estimated and actual position of the considered interest point, as described in equation (9):

$$\epsilon(p_{1i}, \boldsymbol{H}) = \|p_{2i}^{est} - p_{2i}\| \ . \tag{9}$$

Ideally $p_{2i}^{est} = [x_{2i}^{est}, y_{2i}^{est}, 1]^T$should be as close as possible to $p_{2i} = [x_{2i}, y_{2i}, 1]^T$.

In the case where the estimation error $\epsilon(p_{1i}, \boldsymbol{H})$ is inferior to a predefined threshold *E*, the corresponding pixels are marked as belonging to background. The outliers, *i.e.* pixels with estimation error $\epsilon(p_{1i}, \boldsymbol{H})$ exceeding the considered threshold, are considered to belong to foreground objects.

In our experiments, the background/foreground separation threshold *E* has been set to 5 pixels. An example of obtained results is illustrated in Fig. 1d.

Step 4: *Motion classes estimation* - In practice, multiple moving objects are present in the scene. In this case, we determine a new subset of points formed by all the outliers and all the points not considered in previous step (obtained after

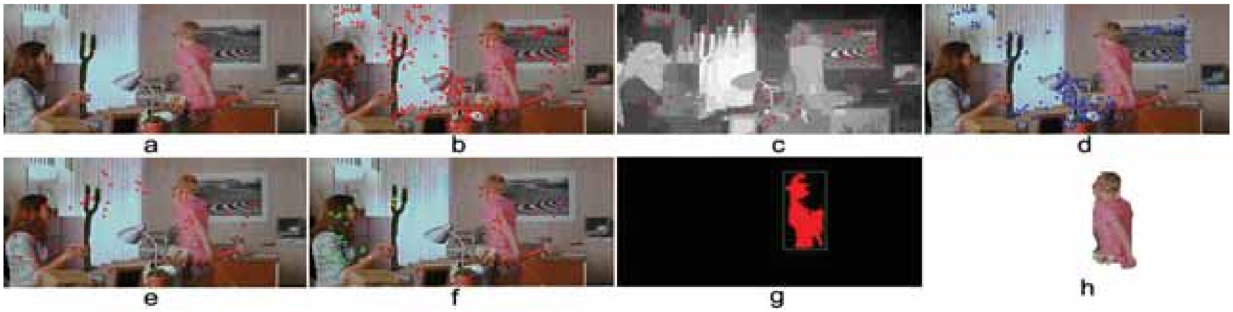


Figure 1.Object detection framework. (a) Selected keyframe; (b) Initial interest points; (c) Subset of keypoints used for camera/background motion estimation; (d) Estimated camera / background movement; (e) Outlier interest points; (f) Motion classes; (g) Salient regions; (h) Segmented object.

subtracting from all the interest points the subset of m background points previously determined) (Fig. 1e)

Then, the considered points are agglomerative clustered into classes. The basic principle behind agglomerative techniques is to consider each individual point as a cluster and then successively reduce the number of classes by merging the two closest clusters until all points are assigned to a category [21]. The key operation of the proposed algorithm is the proximity computation between two interest points that are classified into clusters based on the following steps:

- *Phase I* - The motion vectors are sorted in descending order based on the number of occurrences of the corresponding motion vector angle. For the first interest point, of the current list, a new cluster is formed (MC_i) having as centroid its motion vector angular value (θ_c);

- *Phase II* - For all the other interest points not assigned to any motion class, we compute the angular deviation.

If $Dev(\theta_i, \theta_c)$ is beyond a predefined threshold Th_θ and the motion magnitude is equal with the cluster centroid then the current point will be grouped into MC_i cluster. For the remaining outliers, the process is applied recursively until all points belong to a motion class (Fig. 1f).

In our work the threshold Th_θ has been set to 15 degrees.

Step 5*: Interest point refinement* – For all the interest points included in motion classes we applied the *k-NN* algorithm in order to check that their assignment to the current class is not caused by an error. For the current point we determine its k nearest neighbors based on Euclidian distance. If at least half of the detected points do not belong to the same motion class then this point is removed from cluster. In our experiments k has been set to 5.

Step 6: *Salient motion detection* – For all the motion classes determined at *Step 5* we compute their saliency values as:

$$SalClass(M_i) = \frac{\sum_{j=1}^{m_i} Sal(p_{1,i})}{m_i}, i = \overline{1,N}, \qquad (10)$$

where m_i is the total number of points included in motion class M_i, N is the total number of classes and $Sal(p_{1,i})$ is the value of an interest point $p_{1,i}$ in the spatial saliency map. In this case, the salient motion is determined as:

$$SalientMotion = \max_{i=1,N}\{SalClass(M_i)\}. \qquad (11)$$

Using the salient motion cluster we determine next the relevant regions. A region is considered as salient if it contains an interest point belonging to the salient motion class (Fig 1g).

Step 7: *Object temporal consistency* – Based on the assumption that a salient region should smoothly vary over time, we propose considering the sequence of salient regions as a three-dimensional function r_{crt}*(x, y, t)* with *(x, y)* being the spatial coordinates and *t* being the temporal coordinate. In this step we search for a solution, between successive frames, that preserves the region area as much as possible. A region is tracked over a number of W consecutive frames (using a simple, template matching techniques) if the following condition is satisfied:

$$area(r_{crt}) \bigcap area(r_{ant}) = area(r_{crt}) \qquad (12)$$

where $area(r_{crt})$ denotes the current area of the salient region, $area(r_{ant})$ is the area of the salient region from the previous frame. The result of the tracking provides the regions' object support over time. In the case where the tracking is successful (in the sense of equation 12), the steps 1 to 4 are no longer applied to the corresponding frames, which makes it possible to speed up the detection process.

Step 8: *Object detection* – The final object mask is extracted based on the GrabCut algorithm [2], which is automatically initialized with the ternary saliency map detected at *step 6* (Fig. 1h).

IV. EXPERIMENTAL RESULTS

We tested the proposed methodology on a set of 20 general purpose videos with a resolution of 640 x 264 pixels. The videos may include multiple objects, with various semantics and appearances, including humans performing various activities, animals in the wild, ground and air vehicle.... The first 8 videos are mostly documentaries, noisy, and vary in style and date of production. The next 6 contain important camera and object movement, but with smooth background or without excessive texture, while the final 6 include dark, cluttered and highly dynamic scenes which make them very challenging for an automatic object extraction system. In addition, various types of both camera and multiple object motions are present.

Some object detection results are presented in Fig. 2. Let us first note that for videos with rich texture or including multiple objects, the result of a spatial attention model is, in most of the cases, unrepresentative. However, after incorporating the information associated to the temporal attention model, the method successfully detects the relevant moving regions.

The proposed spatiotemporal visual saliency (STVS) method is compared with the state of the art graph-based visual saliency (GBVS) technique [4]. As it can be noticed from Fig. 2, the GBVS method is not able to correctly identify salient objects neither to strictly localize only the salient object in textured movies. On the contrary, the proposed STVS method successfully recovers the objects of interest.

Regarding the computational complexity, the algorithm was run on a Pentium IV machine with 3.4 GHz and 2 Go RAM, under a Windows XP SP3 platform and the average time required for object detection and segmentation is 3 seconds.

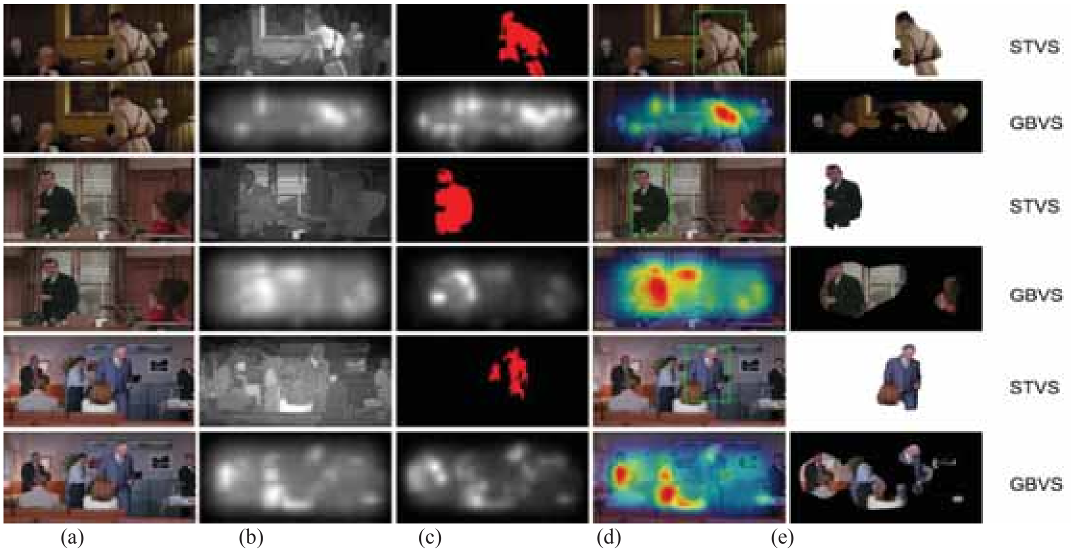


Figure 2. Salient map extraction process. (a) Keyframe selected from a video shot; (b) Spatial saliency map; (c) Temporal saliency map; (d) Candidate regions selection; (e) Detected object.

V. CONCLUSIONS

In this paper we have introduced an automatic salient object extraction system based on a spatiotemporal attention detection framework. The model combines a spatial attention model, based on a region-oriented contrast measure applied with a temporal model derived with the help of interest points correspondence and geometric transforms between keyframes.

The technique is robust to complex background distracting motions and does not require any initial knowledge about the object size or shape. The various experimental results and comparisons with existent methods demonstrate the effectiveness of the proposed technique.

In our future work, we plan to extend the proposed method by taking into account not merely successive frames, but the whole content of a video shot in order to (1) increase the robustness of the algorithm and (2) to incorporate additional information as disparity maps in order to apply it on 3D videos.

REFERENCES

[1] Junchi Yan, Jian Liu, Yin Li, Zhibin Niu, and Yuncai Liu, "Visual saliency detection via rank-sparsity decomposition," in ICIP, pp. 1089–1092, 2010.

[2] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts", Proc. ACM SIGGRAPH, pp. 309-314, 2004

[3] L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE TPAMI*, 20(11), pp. 1254–1259, 1998.

[4] J. Harel, C. Koch, P. Perona. "Graph-based visual saliency", *Advances in Neural Information Processing Systems*, pp. 545- 554, 2007.

[5] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach", IEEE CVPR, 2007.

[6] C. Guo, L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," In IEEE Trans. Image Processing, Vol. 19(1), pp. 185-198, 2010.

[7] Z. Wang, B. Li, "A two-stage approach to saliency detection in images", IEEE Conference on Acoustics, Speech and Signal Processing, 2008.

[8] D. Gao, V. Mahadevan, N. Vasconcelos, "On the plausibility of the discriminant center-surround hypothesis for visual saliency", *Journal of Vision*, vol. 8(7), pp. 1–18, 2008.

[9] A. Oliva, A. Torralba, M. Castelhano, J. Henderson, "Top-down control of visual attention in object detection", In ICIP, pp. 253–256, 2003.

[10] G. Zhai, Q. Chen, X. Yang, W. Zhang, "Scalable Visual Sensitivity Profile Estimation," IEEE International Conference on Acoustics, Speech, and Signal Processing 2008, pp. 873-876, March, 2008.

[11] Tian, Ying-Li; Hampapur, Arun; , "Robust Salient Motion Detection with Complex Background for Real-Time Video Surveillance," *Application of Computer Vision WACV/MOTIONS Volume 1. Seventh IEEE Workshops*, vol.2, pp.30-35, 2005.

[12] Tarkan Sevilmis, Muhammet Bastan: "Automatic detection of salient objects and spatial relations in videos for a video database system", Image and Vision Computing, Volume 26, pp. 1384–1396, 2008.

[13] Lin Ma; Songnan Li; Ngan, K.N, "Motion trajectory based visual saliency for video quality assessment", ICIP, pp.233-236, 2011.

[14] F.F.E. Guraya, F.A. Cheikh, "Predictive Visual Saliency Model for Surveillance Video", EUSIPCO, Barcelona Spain, pp. 554-558, 2011.

[15] R. Tapu and T. Zaharia, "High Level Video Temporal Segmentation", 7th International Symposium on Visual Computing, ISVC-2011, Part I, LNCS 6938, pp. 226–237, Las Vegas, Nevada, USA, 2011.

[16] M. Cheng, N. Zhang, G.and Mitra, X. Huang, S. Hu, "Global contrast based salient region detection," in *IEEE CVPR*, pp. 409–416, 2011.

[17] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603-619, May 2002.

[18] Lowe, D., "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, pp. 1-28, 2004.

[19] R. Panigrahy, "An improved algorithm finding nearest neighbor using kd-trees", In Proceedings of the 8th Latin American conference on Theoretical informatics, LATIN'08, pp. 387–398, 2008.

[20] J. J. Lee and G. Y. Kim. "Robust estimation of camera homography using fuzzy RANSAC", ICCSIA, 2007.

[21] P. Cimiano, A. Hotho, S. Staab, "*Comparing conceptual, divisive and agglomerative clustering for learning taxonomies from text"*, In European Conference on Artificial Intelligence, pp. 435–439, 2004.

Error Resilient Reference Selection for H.264/AVC Streaming Video Over Erroneous Network

Shaikhul Islam Chowdhury*, Jeng-Neng Hwang**, Po-Han Wu**, Goo-Rak Kwon*, and Jae-Young Pyun*+
Dept. of Information and Communication Engineering, Chosun University, Gwangju, Korea*
Dept. of Electrical Engineering, University of Washington, Seattle, USA**

Abstract-- H.264/AVC video delivered through erroneous network has been reported weaker to errors when intra-refresh is used in multiple reference frame structure. This paper proposes a new reference selection method to improve this inefficient error resilience.

I. Introduction

As an error resilient method, typical video codecs including H.264/AVC, employ intra-refresh coding to avoid error propagation on a distorted video sequence in an erroneous network [1]. The refresh order of intra-refresh coding is not simply the raster scan order, but a randomly defined order: once before the start of encoding and a cyclic intra refresh procedure afterwards [2]. Moiron et al. [3] described the limitations of using multiple reference frame (MRF) with cyclic intra refresh (CIR) in an erroneous network, even though increasing the number of reference frames is better than using single reference frames in terms of the pure coding efficiency. In this study, an experiment was performed using JM 17.2 H.264/AVC encoding standard QCIF sequences (tennis and stefan) based on a MRF with and without CIR. Table I lists the corresponding PSNR values in the H.264/AVC reference software. The GOP coding structure consists of an initial intra-coded IDR frame and inter-coded P-frames, i.e. IPPP.., which is suitable for mobile applications. It is shown that the quality of a H.264 video sequence is slightly improved with MRF under an error-free network (as in the case of 0% PLR), but the video quality deteriorates in the presence of CIR in an erroneous network (as in the case of 5% PLR). This occurs because of a typical rate-distortion (RD) based reference selection on MRF, called the RDRS in this paper. The RDRS calculates the RD cost of each coding block mode and chooses the optimal coding block as well as its mode with the minimum RD cost for each macroblock (MB) without considering the possible block damage on the erroneous network. Fig. 1 (a) shows that CIR is not used effectively to refresh the damaged current frame, when multiple frames are buffered and used for a reference prediction in the step of motion estimation (ME).

TABLE I
Mean PSNR values of H.264 video sequences encoded with and without a CIR (packet loss occurred from the 50th frame and PSNR was observed through in all frames)

No. of reference frames	CIR	PSNR (dB) for Tennis		PSNR (dB) for Stefan	
		0% PLR	5% PLR	0% PLR	5% PLR
1	X	34.35	32.09	34.72	23.28
7	X	34.39	32.05	34.74	23.30
1	○	34.37	32.12	34.73	25.01
7	○	34.43	29.45	34.71	23.97

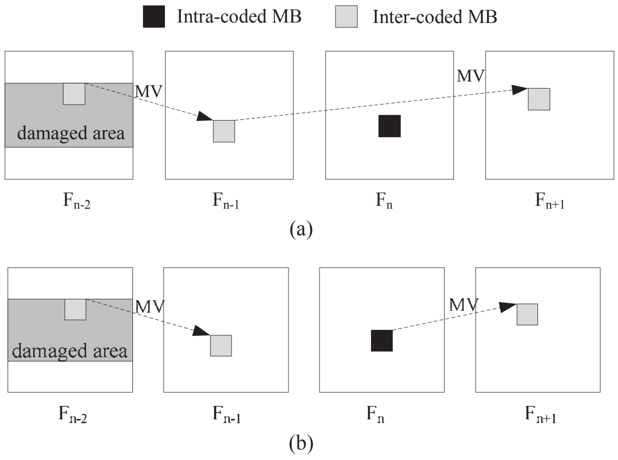


Fig. 1. Effects of multiple reference frames with intra refresh, (a) limitation of MRF, (b) proposed error resilient reference selection method.

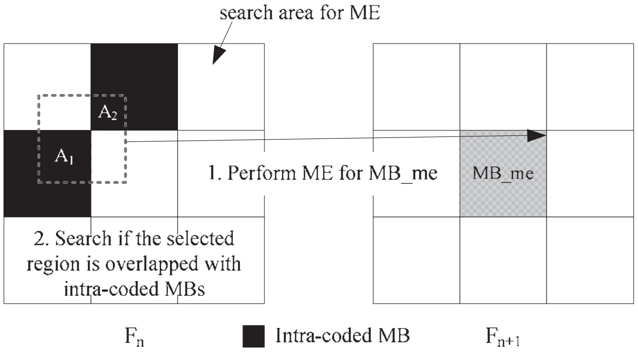


Fig. 2 MV search procedure of the proposed error resilient reference selection method.

II. Proposed Error Resilient Reference Selection

An intra-refresh based reference selection (IRRS) scheme is proposed for the higher error robustness of MRF-based videos transferred over the erroneous network. The IRRS method attempts to choose a reference block that was intra-coded on MRF during ME as shown in Fig. 1 (b). The IRRS searches the intra-coded blocks among the candidate reference blocks

+Jae-Young Pyun (jypyun@chosun.ac.kr) is a corresponding author.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0023389)

in the previous multiple reference frames (F_n, F_{n-1}, and F_{n-2} in this example), and selects a block with the largest region refreshed by intra-coding. Fig. 1 (b) shows that the selected block in F_n is fully or partially refreshed with intra-coded MBs. Therefore, the error propagation is limited on the erroneous network. The minimum requirement of IRRS is that it should recognize the coding modes of all MBs in the ME step to search for the error-resilient and best matched block among the reference frames.

Fig. 2 shows the visualized MV search procedure of the proposed reference selection method. The dotted box in frame F_n is the block matched region that was chosen for the current MB depicted as a grey box in frame F_{n+1}. The matched region is fully or partially overlapped with the intra-coded MBs, which are shown as black boxes in frame F_n. A refresh ratio RR_i in IRRS is defined to select the most refreshed block among the best matched blocks in the reference frames as follows:

$$RR_i = \frac{\sum_{w} A_w^i}{\text{\# of pixels in MB}}, \qquad (1)$$

where A_w^i is w^{th} intra-coded area size (pixels) of block matched MB in i^{th} reference frame.

III. Simulation Results

The IRRS scheme was implemented using a JM encoder. Standard test video sequences (tennis, soccer, football, and stefan) were encoded at 30 frames per second using seven reference frames. The H.264/AVC bit-stream was packetized in real time protocol (RTP) mode. A random packet loss model was used to simulate the erroneous transmission links. 0% to 10% of the packets were dropped randomly from the RTP bit-stream. Error concealment at the decoder was set as the frame copy. Simulation results are shown in Fig 3, 4, and Table II.

Fig. 3 presents a frame by frame PSNR comparison for a tennis QCIF sequence, where 5% video packets are dropped from the network. The IRRS performs better than the RDRS in terms of the video quality when an error occurs on the video streams encoded with a MRF-based inter-prediction. The IRRS reduces the error propagation by choosing more refresh MBs in the H.264/AVC videos using both MRF and CIR, simultaneously. Fig. 4 shows the quality of a tennis video sequence against different PLR, and Table II lists the quality enhancement of IRRS compared to the RDRS under various PLR conditions (the positive gain in Table II implies that the PSNR of the IRRS is larger than the RDRS). This proposed error resilient method can be used in the most of the streaming video applications developed on consumer electronics and mobile devices.

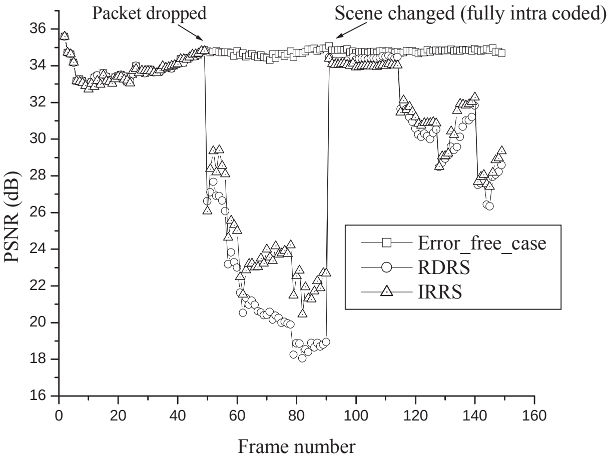


Fig.3 Quality of a tennis sequence with either RDRS or IRRS in an erroneous network (PLR 5%).

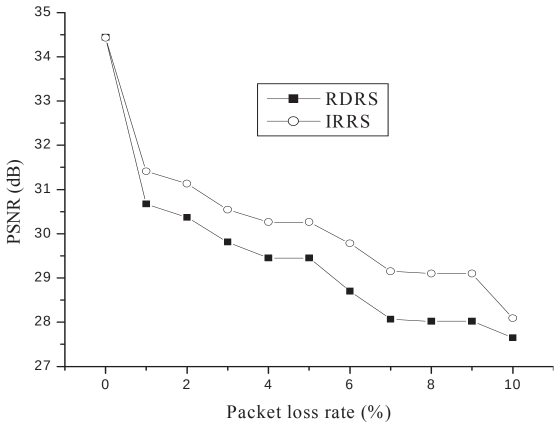


Fig. 4 Quality of a tennis sequence with either the RDRS or IRRS under different PLRs.

TABLE II
Quality difference between the IRRS and RDRS method against the packet loss rate for various standard sequences

PLR (%)	Quality difference between IRRS and RDRS			
	Tennis	Soccer	Football	Stefan
0	- 0.01	+ 0.02	+ 0.04	- 0.005
2	+ 0.76	+ 0.52	+ 0.20	+ 0.11
4	+ 0.81	+ 0.50	+ 0.03	+ 0.16
6	+ 1.08	+ 0.69	+ 0.45	+ 0.18
8	+ 1.08	+ 0.62	+ 0.43	+ 0.23
10	+ 0.44	+ 0.69	+ 1.14	+ 0.22

References

[1] S. Wenger, "H.264/AVC over IP", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, Jul. 2003.

[2] Paulo Nunes, Luis Ducla Soares, and Fernando Pereira, "Error resilient macroblock rate control for H.264/AVC", Proc. of IEEE International Conference on Image Processing, pp. 2132-2135, Oct. 2008

[3] S. Moiron, I. Ali, M. Ghanbari and M. Fleury, "Limitations of multiple reference frames with cyclic intra-refresh line for H.264/AVC", *ELECTRONICS LETTERS*, vol. 47, no. 2, Jan. 2011.

Efficient Video Transmission Using Network Coding over WLAN

Kyu-Sung Hwang
Kyungil University
Gyeongsan, Gyeongbuk, 712-701 Korea

Ronny Yongho Kim
Korea National University of Transportation
Uiwang, Gyeongki, 437-763 Korea

***Abstract*— In order to effectively transmit multimedia traffic in error prone wireless LAN networks, a novel transmission scheme using soft decision values and symbol-level random network coding over wireless LAN is proposed in this paper. A simple combining technique using soft decision values can improve video quality substantially which is shown with simulations. The proposed WLAN video packet transmission scheme can enhance packet transmission efficiency substantially. The proposed scheme can be used for unequal error protection of video frames for robust video transmission over wireless networks.**

I. INTRODUCTION

With the emergence of mobile multimedia devices, such as smart phones, multimedia services over wireless networks are receiving considerable interest. Wireless local area network (WLAN) based on IEEE 802.11 standards [1] is getting more popular than ever, thanks to its simple and esay deployment and excellent packet processing performce. IEEE 802.11 air interface is based on carrier sensing and multiple access with collision avoidance (CSMA/CA) and employs a stop and wait automatic repeat request (ARQ) error control scheme. Advanced video coding standard, H.264 [2] is a widely used video coding scheme for multimedia video. In video coding, a group of pictures (GOP) specifies the order in which intra- and inter-frames are arranged. The GOP is a group of successive pictures within a coded video stream. Each coded video stream consists of successive GOPs. A GOP can contain I-frame (intra coded picture), P-frame (predictive coded picture) and B-frame (bidirectionally predictive coded picture). Due to their interdependency, each type of frame has a different importance. Bit errors in P- and B-frame and their effect on video quality is demonstrated in Fig. 1. In order to provide better protection to more important video frame type, unequal error protection (UEP) schemes have been extensively studied [3].

With respect to the objective of maximizing the throughput, network coding had been originally proposed in information theory [4] and has since emerged as one of the most promising information theoretic approaches to improve throughput performance. Authors in [5] proposed a cooperative symbol-level network coded packet transmission scheme, referred to as Drizzle. Drizzle exploits soft decision values in the operation of random network coding and provides efficient packet transmission.

Why would the conventional ARQ not be good enough for efficient video transmission over WLAN ? One of reasons is because no partial recovery is possible with the conventional WLAN ARQ. If there exists even a single error in a received packet, that packet has to be dropped. In this paper, in order to overcome the conventional ARQ's efficiency, a novel transmission scheme using soft decision values and symbol-level random network coding is proposed.

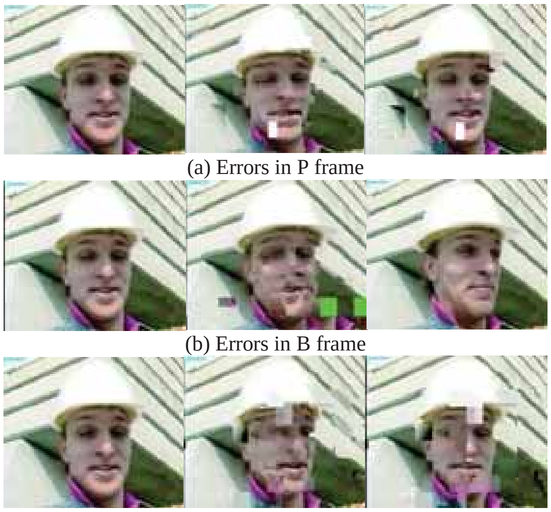

(a) Errors in P frame
(b) Errors in B frame
(c) Errors in P and B frames

Fig. 1. Errors in specific video encoding frames and their impact on the video quality.

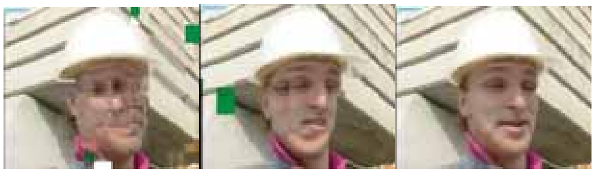

(a) No Combine (b) 1 Combine (c) 2 Combine

Fig. 2. Packet combining using soft decision values improves the video quality

II. PROPOSED SCHEME

In order to understand how soft decision values can be used in partial error recovery, SOFT [6] is implemented in simulation and how it can improve video quality is examined. SOFT works by combining the soft decision values of physical layer across multiple faulty receptions to recover a clean packet. SOFT's simple combining technique using soft decision values can improve video quality substantially. Our simulation result in Fig. 2 shows the benefits of the combining technique using soft decision values. Since the combining technique requires processing power, it can be selectively used to protect important video frames, e.g., I-frame, P-frame for UEP.

Fig 3. shows the encoding process of random network coding and WLAN packet formation with the random network coded blocks. The same symbol-level random network coding schemes as in [5] are employed and a detailed explanation on the symbol-level random network coding is omitted in this paper due to page limitation. In order to integrate the symbol-

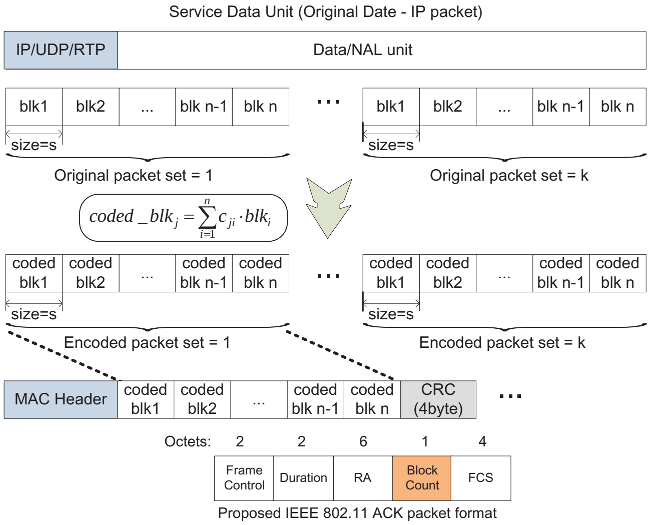


Fig. 3. An example of video data integration (NAL units) in the IEEE 802.11 framework with random network coding and the proposed ACK packet format.

level random network coding with soft decision values and inform the transmitter of the required number of retransmission coded blocks, an additional 1 byte of feedback information is added in the proposed ACK packet. Since random network coding can provide favorable rateless and randomness properties, the receiver does not need to specify the location of errors and only needs to feedback the number of errored blocks in the previous transmission. Once the transmitter receives the required number of blocks, it generates different versions of the random network coded blocks using different coding coefficients as many as the receiver requests. As studied in [5], in the decoding process, soft decision values are utilized to check which blocks are in error without additional parity check bits in each coded block. An error checking step using soft decision values can be omitted if the payload is error free by checking cyclic redundancy check (CRC) provided in the WLAN MAC frame.

In order to see the performance enhancement of the proposed scheme, packet delivery efficiency is defined as follows:

$$Delivery_{eff} = \frac{D_{success}}{D_{tx}} = \frac{D_{success}}{D_{data} + D_{overhead} + D_{ack}} \qquad (1)$$

where, $D_{success}$ is the size of successfully transmitted data without error, D_{data} is the size of data to transmit, $D_{overhead}$ is the WLAN MAC overhead, and D_{ack} the size of the ACK packet. D_{ack} of the proposed scheme is 1 byte larger than the conventional WLAN.

The simulation result in Fig. 4 shows that the proposed scheme substantially enhances packet delivery efficiency for various packet sizes. The proposed scheme can improve the packet transmission efficiency up to 45dB at a SNR of 10 dB compared to the conventional scheme. The proposed scheme's efficient partial recovery capability shows more gain as the packet size gets larger. Since typically, I-frame and P-frame sizes are much larger than B-frame, the proposed scheme can be used to provide UEP for large size I-frame and P-frame at the cost of random network coding processing.

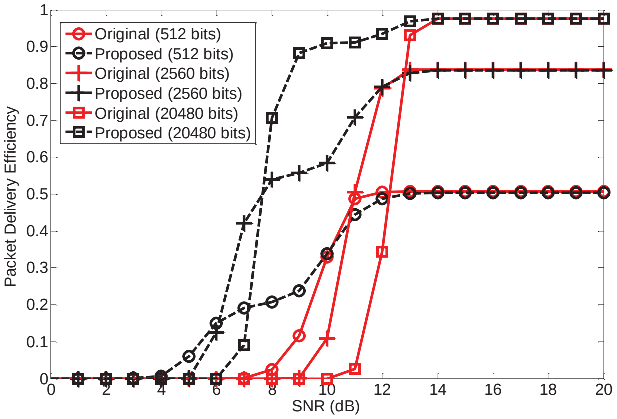


Fig. 4. A packet delivery efficiency in the original transmission compared with the proposed scheme. Numbers in the parenthesis are packet sizes in the payload.

III. Conclusion

An efficient video transmission scheme using symbol-level random network coding is proposed in this paper. We have first examined how bit errors in different video frames impact on video quality and studied the improvement of video quality with partial recovery using soft decision values. The proposed scheme can provide a strong partial recovery capability using both soft decision values and random network coding. We can see that the proposed scheme can enhance packet delivery efficiency substantially. Thanks to the favorable randomness and rateless properties of random network coding, the proposed scheme can provide robust and efficient video transmission over WLAN.

IV. Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2012R1A1A1014610). R. Y. Kim is the corresponding author.

Reference

[1] "IEEE Standard for Local and Metropolitan Area Networks - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications ," IEEE Std 802.11-2007, June 18 2007

[2] J.V.T.J. of ISO/IEC MPEG and I.-T. VCEG, *ITU-T H.264 — Series H: Audiovisual and Multimedia Systems—Advanced Video Coding for Generic Audiovisual Services*, Mar. 2005.

[3] Y. C. Chang, S. W. Lee, and R. Komyia, "A fast forward error correction allocation algorithm for unequal error protection of video transmission over wireless channels," *IEEE Transactions on Consumer Electronics*, vol 54, no. 3, pp 1066-1073, 2008.

[4] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network Information Flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000

[5] R. Y. Kim, J. Jin, and B. Li, "Drizzle: Cooperative symbol-level network coding in multi-channel wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp 1415-1432, Mar. 2010.

[6] G. R. Woo, P. Kheradpour, D. Shen, and D. Katabi, "Beyond the Bits: Cooperative packet recovery using physical layer information," in *Proc. of ACM MobiCom*, Sept., 2007

Implementation of a Seamless Uncompressed Video Transmission System in 60GHz Bands

Jong Hwa Choi[1], Hyongjin Kwon[1], Jinkyeong Kim[1],
Woo-Yong Lee[1], and Younggap You[2]
[1]Wireless Telecommunication Research Department, ETRI, Korea
[2]Department of Information and Communications, Chungbuk National University, Korea

Abstract—A ***60GHz communication can be a best solution for high capacity wireless data communications such as uncompressed HD video streaming. However, due to highly directional characteristics of 60GHz frequency band, the line-of-sight (LOS) path blockage should be avoided to maintain the peer-to-peer connectivity. In this paper, we propose a novel 60GHz relay system for blocking avoidance, and show the experimental results of the prototype system employing proposed relay scheme which outperforms the conventional LOS communication system, and also present that video streaming is seamless when LOS is blocked with a prototype system.***

I. INTRODUCTION

Over the past few years, many efforts have been made to realize high quality video services in High Definition (HD) TV with video streaming technology employing either IEEE 802.11n or Ultra-Wide Band (UWB) wireless links which can support less than 1Gbps data rate [1]. However, full HD videos having 1080p video frames with each pixel having RGB components, and a frame rate of 60GHz require a channel bandwidth of about 3Gbps to support video data only [2]. In this case, due to the limitations of the available data rates, HD video streaming using current wireless technologies has always been accompanied by a video compression technique like a MPEG2 [3]. These compression schemes must be used for the delivery of video data due to the insufficient throughput of the 802.11n and UWB wireless links, thus resulting in unsatisfactory delivery latency and video quality degradation.

For this reason, new wireless communication standards are being developed to support uncompressed full HD video streaming services. One of the most promising wireless solutions to achieve a multi-gigabit transmission is to use the 60GHz band. The attractiveness of the 60GHz band for short-range high-rate communications is the short wavelength at these frequencies. It makes the band ideally suite for short range communication and dense deployment scenarios such as indoor environments. Although the 60GHz communication may have a multi-gigabit transmission rate, the device working at the 60GHz band suffer from high path loss and high penetration loss by human or wall due to the inherent 60GHz band characteristic. In addition, since 60GHz band has highly directive characteristics, the peer-to-peer connectivity is not guaranteed without the line-of-sight (LOS) link. To solve these problems, several researches are being carried out [2, 4, 5]. Authors in [2] resolved the uncompressed HD video wireless delivery and loss problem with efficient error protection and concealment schemes that exploit the unequal error resilience properties of uncompressed video. But, this research only focused on the performance analysis of the uncompressed video wireless delivery problem with or without blocking. When the LOS path is blocked by obstacles, such as human or other movements, the peer-to-peer communication is impossible. For this reason, a real-time audio/video (A/V) streaming service to provide seamless A/V could be adversely affected. Therefore, we suggest a relay system to solve this problem.

II. DESIGN OF THE VIDEO TRANSMISSION SYSTEM INCLUDING RELAY

A block diagram of the 60GHz video transmission system using relay is shown in Figure 1. The system structure comprises a source, a destination and a relay. The block diagram of the Wireless Video Transceiver (WVT) is shown in Figure 2. It consists of several blocks including PHY, MAC and Video PAL (VPAL). Relaying system consist of three devices: a source device, a relay device and a destination device. All three devices are assumed to be capable of directional peer communication, which means that each of them has multiple directional antennas or beam formable array antennas, so that they can send data in the direction pointing to one another, not omni-directionally.

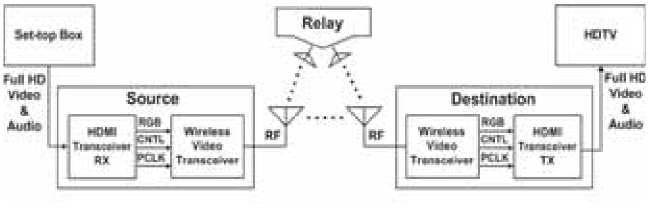


Fig. 1 60GHz wireless transmission system Block Diagram.

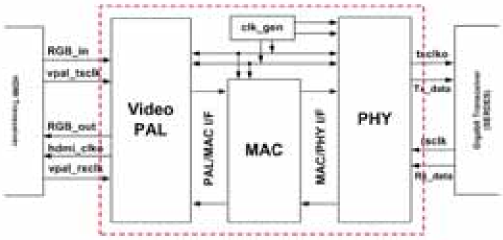


Fig. 2 Wireless Video Transceiver Block Diagram.

This research was supported by the KCC (Korea Communications Commission), Korea, under the Short Range Wireless Transmission Technology with Robustness to Radio Interferences supervised by the KCA (Korea Communications Agency) (KCA 12-921-01-301)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

These high-gain directional antennas make it possible to overcome the severe channel condition and to extend the service coverage in the 60GHz frequency environment.

III. IMPLEMENTATION AND EXPERIMENTAL DEMONSTRATION

This section presents the performance of the 60GHz wireless transmission system and the demonstration of uncompressed full HD video streaming using our system. We implemented a wireless video transmission board using an FPGA board and HDMI transceiver board as shown in Figure 3. The FPGA prototype implementation uses Altera Stratix-II-GX EP2SGX 130GF1508C4. All the blocks of the WVT are soft cores described in Verilog HDL.

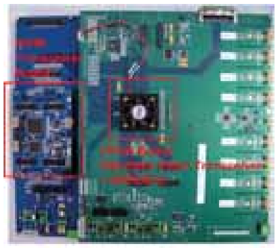


Fig. 3 Wireless Video Transmission Board.

With the WVT and a relay, we built a test platform for seamless uncompressed full HD video TRX systems, as shown in Figure 4. The test platform is connected with the 60 GHz RF module developed in collaboration with Comotech Co., Ltd in Korea. As shown in Figure 4, a relay device is used for seamless data transmission. When the direct link between two devices (a source and a destination) communicating with each other is blocked, another link via a relay device is used to continue data transmission and reception. In this case, after relaying a frame transmitted from the source, a relay device switches its antenna mode based on the ACK policy, which enables its antenna switching before the destination sends an ACK frame. To support this functionality, the relay device should have at least two RF modules capable of antenna training.

Figure 5 shows an example of the video transmission that does not use a relay system. The transmitter transmits the video data to the receiver through the direct link. As shown in the figure, when the direct link is blocked by obstacles, the video image is not displayed. In other words, if a direct link is blocked, the receiver cannot receive data. Figure 6 shows the video transmission using the relay system. The transmitter transmits the video data to the receiver through the direct link and the relay link. The receiver displays the video image on a screen even if a direct link is blocked by obstacles. Receiving data through the relay link is possible even if a direct link is blocked.

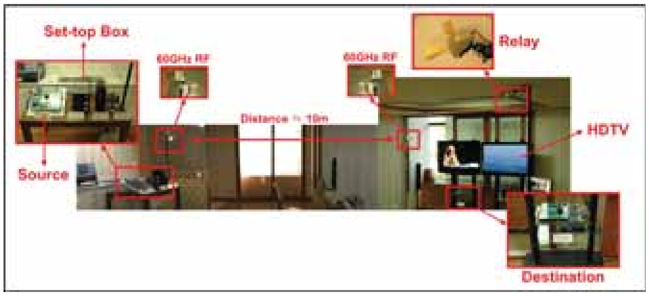


Fig. 4 Implemented 60GHz wireless video transmission system and demonstration of uncompressed full HD video transmitting.

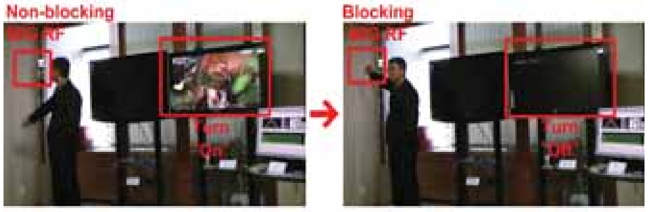


Fig. 5 Video transmission that does not use relay system.

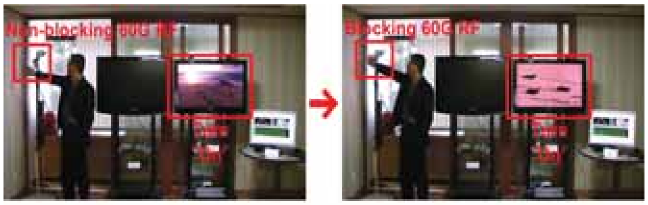


Fig. 6 Video transmission using the relay system.

IV. CONCLUSION

We demonstrated a 60GHz video transmission system which provides another communication link when LOS is blocked by typical obstacles, including human movements. Also, through a prototype system implementation, the proposed concept of blocking avoidance using a relay for video streaming is shown to work well and provides seamless video streaming services.

REFERENCES

[1] D. Porcino, B. V.D. Wal, and Y. Zhao, “HDTV over UWB: Wireless Video Streaming Trials and Quality of Service Analysis,” JPEG2000 Technical article, Analog Devices, Inc., 2006.

[2] H. Singh, O. Jisung, K. Changyeul, Q. Xiangping, S. Huai-Rong, N. Chiu, “A 60 GHz wireless network for enabling uncompressed video communication”, *IEEE Communications Magazine*, vol. 46, issue 12, pp. 71-78, Dec. 2008.

[3] Youngae Jeon, Sangjae Lee, Seonghee Lee, Sangsung Choi, Dae Young Kim, “High definition video transmission using bluetooth over UWB,” *IEEE Trans. Consumer Electronics*, vol. 56, No.1, pp. 27-33, Feb. 2010.

[4] I. Lakkis *et al*, "IEEE 802.15.3c Beamforming Overview*," IEEE 802.11-09/0355r0*, Mar. 2009.

[5] Wonjin Lee, Kwangseok Noh, Saejoon Kim and Jun Heo, "Efficient Cooperative Transmission for Wireless 3D HD Video Transmission in 60GHz Channel," *IEEE Trans. Consumer Electronics,* vol. 56, no. 4, pp. 2481-2488, Nov. 2010.

Development of HTTP-based Multivision Video Streaming Server and Benchmark Evaluation

Yuuki WAKISAKA, and Hiroyuki KASAI, The University of Electro-Communications, JAPAN

Abstract – **High functionality and interactivity of video services are strongly desired. The authors have presented a proposal of an encoding scheme and a fast joining scheme for interactive multi-vision systems, where joining is the combination of multiple streams into a single video stream. This paper describes implementation of an actual HTTP-based streaming server by embedding the stream joiner module into a widely available HTTP server. The contribution of this paper is that the joining speed of the server surpasses the necessary speed attained with practical network equipment connecting to the server.**

I. INTRODUCTION

Multi-vision systems are expected to form the nexus of a new paradigm of video delivery services. Earlier reports [1, 2] described implementation of a system that enables viewing of multivision video at any view area at multiple resolutions. The system is achieved by dynamically joining multiple video streams that are encoded from an entire video area or a small partitioned area of a video. Its greatest benefit is that it enables a client to play a stream simply using a generic video decoder. This structure obviates complexity at the client terminal because the client does not handle multiple decoders, synchronous rendering, and multiple sessions with a server. However, because of the high load of the joining process of the tile streams on the server side, an earlier reported method achieved much faster stream joining. This paper describes stress performance testing of the joiner embedded in a real server-client system under a present practical environment. Results show that the joining speed of the streaming server surpasses the necessary speed achieved under the latest practical network equipment connecting to the server.

II. MULTIVISION VIDEO SYSTEM BASED ON STREAM JOINER

The server joins pre-encoded multiple streams into one bitstream according to a user request. Then it transmits it to many clients simultaneously. Tile stream joining is performed by joining two MB-lines of two vertically and horizontally adjacent tile streams. The special features are a new encoding scheme for a lightweight and fast stream joiner consisting of the prediction restriction scheme, the MB-line size insertion scheme, and the MB-line byte-alignment scheme. CAVLC decoding, which requires a dominant processing load in the conventional joiner, is avoidable. This configuration achieved about 40–80 times faster joining than the conventional scheme.

III. MULTIVISION HTTP-BASED VIDEO STREAMING SERVER

A. Background and Basic Architecture

Recently, HTTP streaming has been gaining wide support in the video streaming field because it requires only a standard HTTP Server and no specific addition to the system. Furthermore, HTTP causes no firewall issues related to the streaming-specific protocol: RTP. Accordingly, MPEG started to specify system architecture, protocols [3]. Therefore, we implemented our stream joiner onto a streaming server using Apache Server [4] because of its popularity. We implemented the stream joiner module to be compliant to the server specifications, and embedded the joiner module onto the server. Only the H.264/AVC joiner module was implemented, excluding a multiplexer such MP4 to evaluate the stream joiner purely. Our current implementation adopted a prefork (Multi-Processing Modules: MPM)) rather than a worker MPM because the former presents little risk of thread safety and a limited number of file descriptors.

B. System Details and Process Flow

The Apache root process creates a new child process, called a Prefork, to accept a user request before a practical request. At this stage, the developed joiner module is loaded in a dynamic link or static link manner (Fig. 1:(i)). Apache Core calls the *fvjoiner_register_hooks* in the joiner module, which tells a request handler function when requested to Apache Core. In general, Apache httpd dispatches a request to the registered handlers in series. Selection of a handler process of a request is configured by SetHandler (ii). Then, this process waits for an incoming request (iii). Once a request comes, a handler is set is for the joiner (v) when its requested URL (iv). From here, the creation of its response starts. First, Content-Type like "video/mp4" is set. Second, the main routine of the joiner module is executed by starting initialization of the joiner process liketile stream file openings and buffer allocations (vi). The main joining process operates in this step, and sends an output joined stream into Output Filer of Apache. After a finalization process such as tile file closing, buffer releases, and a connect closing, they might be performed at this stage (vii) if other optional Output Filer modules are inserted into Apache. Finally, a response message is transmitted to the request. Once the transmission of the response is completed, the prefork process reverts to the waiting status, or terminates when expired (viii).

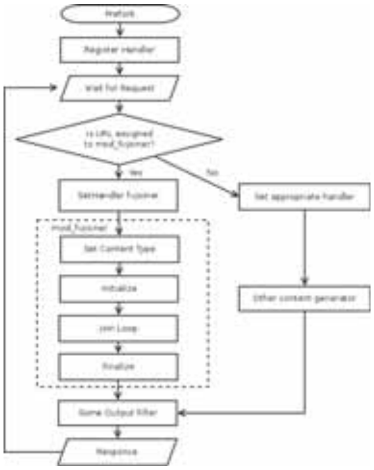


Fig. 1. HTTP Server and Joiner Flowchart.

IV. PERFORMANCE EVALUATION

A. Overhead Evaluation of the Embedded Joiner Module

Overhead of the embedded joiner module for the server was measured by comparison with a native-based executable joiner engine. The overhead might be attributable to handling

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

of the request-message, entailing additional Prefork processes. The results shown in Fig. 2 depict that the bitrate of the embedded joiner is about 35% lower than the native bitrate.

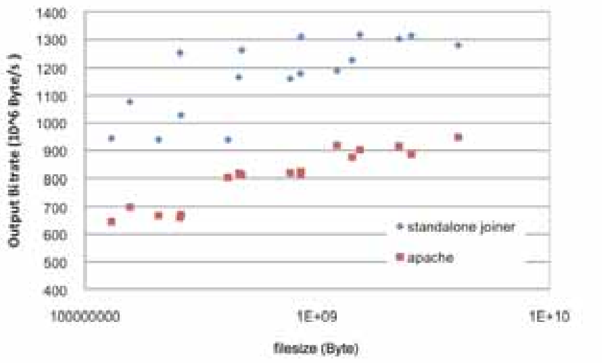


Fig. 2. Overhead Evaluation of Embedded Joiner Module.

B. *Transmission Speed with Parallel Streams Streaming*

A comparative evaluation was conducted using the conventional method, which sends streams in parallel that are pre-stored inside the server. For this evaluation, we used Apache Benchmark (ab), a benchmark tool distributed with Apache HTTP Server. The server has 2.4 [GHz] dual CPU with eight cores, DDR3-1066 Dual Channel 12 [GB] Memory, Serial-ATA HDD with 7200 [rpm], and 32 [MB] cache. The Apache version is 2.2.15. The tile stream resolution was 160 × 96 [pixels], the concurrent connections were 1, 5, 10, 20, and 40. The numbers of streams to be joined were 16, 36, and 64. The numbers of frames were 900, 2700, and 5000.

1) *Stress Test on Local Loopback Access*

Download times were measured by connecting multiple sessions to the server via local loopback. The results depicted in Table I show that the parallel transmission method is superior to the joining method in terms of the transmission speed. However, the parallel transmission method cannot accommodate many connections, although the joiner can maintain its serving capability under the same circumstances. Meanwhile, the overhead of the joiner was distinctly apparent when the number of tile streams and the concurrent connections were lower. Consequently, the joiner is better suited for accepting more requests. Because a video streaming in general requires longer sessions, and because this forces the streaming server to accommodate more multiple requests, this feature, which can reduce connects, can work well.

2) *Stress Test on External Network Access*

Realizing the joiner's disadvantage related to the transmission speed, we measured it under a practical ideal environment, where the benchmark client accesses via an external network with 1000Base-T NIC and Category 5 Ethernet. Results shown in Table I revealed that both methods indicated similar bitrates, which are expected to be the limitation of 1000Base-T NIC. Consequently, if the number of parallel accesses is one, then the transmission speed provided by the joining method reaches the network speed. Thereby, its speed is not regarded as a bottleneck. Furthermore, even given multiple accesses, the joiner speed itself does not produce a bottleneck because all sessions share this network interface.

Table I Transmission Speed with Parallel Transmission (10^6byte/s).

Concurrent Connections	Local Access		External Access	
	Joiner	Parallel	Joiner	Parallel
1	886.232	10411.555	111.964	126.511
2	642.451	3662.362	56.145	61.613
5	217.638	1707.007	22.566	25.414
10	104.247	1223.134	11.331	Timeout
20	50.058	Timeout	5.706	Timeout
40	24.542	Timeout	2.908	Timeout

C. *Evaluations of Joining Bitrate*

Results presented in the preceding subsection revealed that a bandwidth of the network can be a bottleneck of the practical server–client system. Therefore, if the output bitrate of the stream joiner exceeds the bandwidth, then the joining speed presents no critical problem. This subsection describes that the implemented multivision server with the joiner can achieve a required minimum bit rate under any parameter combinations of the number of simultaneous access clients, number of tile streams to be joined, and total frames of the joined stream. The joining bitrate per connection under concurrent connections and the total joining bitrate are portrayed in Fig. 3. As shown in Fig. 3, the bitrate tends to depend only on the number of concurrent connections where the number of connections is higher. Fig. 3 also shows that the total joining bitrate is nearly constant at 1000 [Mbytes/s] over all connections. This overall performance on the developed stream joiner can operate under the latest widely available server hardware and network infrastructure given any concurrent connections.

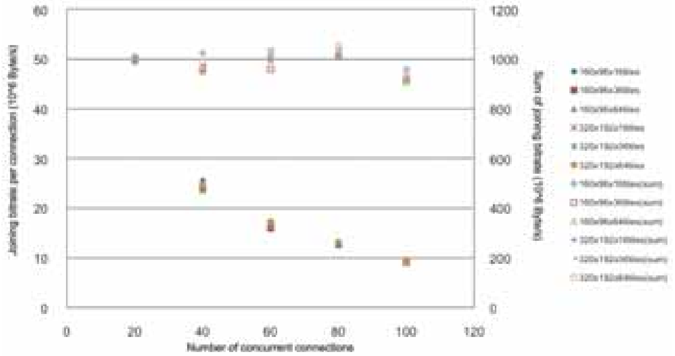


Fig. 3. Number of Concurrent Connections vs. Joining Bitrate.

V. Conclusions

We developed a multivision HTTP-based video steaming server based on a stream joiner. Results show that the joining process speed of the streaming server surpassed the necessary speed under the latest widely practical network equipment connecting to the server.

References

[1] N. Uchihara and H. Kasai, "Fast H.264/AVC stream joiner for interactive free view-area multivision video," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1311-1319, August 2011.

[2] N. Uchihara and H. Kasai, "H.264/AVC prediction restriction encoding control for fast multiple stream joiner," *IEEE International Conference on Consumer Electronics*, pp. 277-278, January 2012.

[3] ISO/IEC 23009-1:2012, Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats, 2012.

[4] Apache Software Foundation, http://www.apache.org/.

Accessible Display Design to Control Home Area Networks

Laisa C. P. Costa, *Member, IEEE*, Nicholas S. Almeida and Marcelo K. Zuffo, *Member, IEEE*
University of Sao Paulo / LSI-TEC

***Abstract*--Recently, the social inclusion and technical aid to assure autonomy to people with disabilities are getting attention all over the world. We present a display design for accessible interaction in home area networks. Based on a research on the accessible interfaces state of the art, an interface design was proposed. This interface was implemented over a Tablet that controls the domestic devices through a home network controller prototype. In order to evaluate the design, a research was conducted, interviewing people with disabilities in Brazil. This research consolidated a feasible accessible interface to control home area networks pointing out the main requirements for home area networks considering a diversified group of impairments.**

I. INTRODUCTION

Focusing on the use of home area networks to improve disabled people autonomy at home, this paper presents a display design for accessible home control.

In the past years, computational devices have turned faster, smaller, connected and cheaper. It brings the "intelligent house" vision, promised for decades, closer to reality. This pervasive, intelligent home, a luxury item to many people, could have a key role in assuring the autonomy of people with disabilities.

In Brazil, assistive resources and their use are relatively recent as compared to the United States, for example, where specific laws were established in 1988. In Brazil, similar regulations have existed since 2004 and establish general standards and basic criteria to promote accessibility. [1]

Thinking about users with disabilities, it is necessary to invest efforts in the research and development of accessible interfaces, through the perspective of a universal design that is easy to use and to learn how to use.

The design for all, also called universal design, began focusing on physical aspects (buildings, urban spaces, transport, health, leisure), and nowadays is extended to the digital world (computer networks and communication systems). In this perspective, accessibility is defined as "a condition for autonomous and safe use of space, furniture and urban facilities, buildings, transport services and devices, systems and media and information by people with disabilities or reduced mobility." [2]. It is worth stressing that accessibility is not the creation of exclusive spaces for people with disabilities, which could be a form of discrimination, but rather of thinking of systems and environments, which can be used by everyone.

The work was developed starting with an interface design proposal, based on the research on accessible interfaces state of the art. The interface was deployed to Android Operating Systems, targeting Tablets and Smart Phones interoperability. The interface was integrated to a home gateway prototype. In order to evaluate the design, ten interviews with people with disabilities were conducted in Brazil.

This research could consolidate a feasible interface to control home area networks pointing out the main requirements for home area networks considering a diversified group of impairments.

II. INTERFACE DESIGN

Our work considered target user people with visual, hearing, motor and cognitive disabilities. In order to develop a widespread and easy to use interface, we adopted a design approach based on quadrants and touch screen.

The touchscreen choice was made based on two factors: the widespread use of this technology on mobile devices and the touchscreen intuitiveness. Considering that people with disabilities have more locomotion difficulty, the possibility to have the home control interface on a portable device such as a smartphone or a tablet is a great advantage.

We adopted a design based on quadrants to achieve a universal user interface for home network control. The quadrant design had been previously used by Zhao et al. to deploy GUI to visually impaired users [3]. Although in this work the quadrants design showed not to be the best design solution for visually impaired persons, it proved to be suitable to them. In our work, we have the hypothesis that the quadrant approach is a good design considering a wider variety of impairments, allowing fast learning and intuitive use.

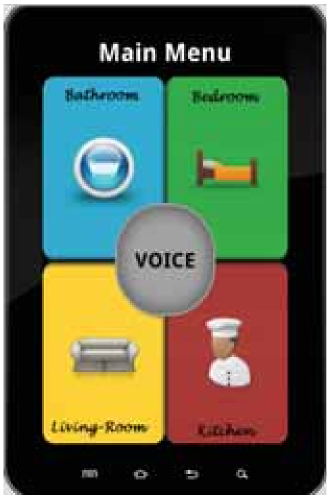


Fig. 1: Main menu screen. This Figure shows the main menu screen as the quadrant approach design example.

In our design, the quadrant approach was used, considering

This work was partially supported by FAPESP, Intel of Brazil, CAPES and FINEP.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

a layout with five buttons occupying the whole screen area. There is one button in each quadrant (four buttons) and one (the fifth) in the center of the screen. In order to improve the interface intuitiveness, a text label, an icon, and a color were associated to each button. An example of this approach can be seen in Fig. 1.

The interaction mechanism was based on "touch" and "hold pressed" events. The "touch" event selects the key that was pressed and a "hold pressed" event triggers the action related to the key. The action could be to send a command or to read the status of a device in the home, or to go to a next screen in the interface.

Both events generate visual and audible feedbacks. The "touch event" generates the synthetized locution of the touched key text label as audible feedback; the visual feedback is provided by the key enhancement by changing its brightness. The "hold pressed event" generates the synthetized locution of the name of the next screen as audible feedback, or the status of a device (if it is related to an action command); the visual feedback occurs by changing the screen to the next one or representing the new status of a device, if that is the case.

The proposed interface also has another interaction mechanism that uses speech control. The user can say the name of any screen or key in order to trigger an action. Although our proof of concept has only one speech option to each command, it is possible and necessary to register similar commands, in order to facilitate the voice interaction mechanism. Using the central button on the main menu screen activates the speech control mode.

The device that embeds the interface has keys that are used by the interface as well. These keys are the optional usage. The BACK key opens the main menu screen, no matter in which screen the interface is. The MENU key presents the interface configurations. It allows configurations related to the touch event and hold pressed event time threshold, to the speed and voice of the synthetized speech, etc.

III. IMPLEMENTATION

The interface was implemented over the Android Operating System 2.2. The interface was developed using the MOTODEV Studio and was deployed on a Tablet with 7-inch display, 512MB of memory and a 1GHz processor. The interface implementation consists of twenty screens with five devices and lighting being controlled.

The interface consistency was maintained, using the quadrant layout approach to every screen. In the first screen (main menu), the quadrant keys are used to select a place in the house and the center key is used to activate the voice command mode. In the next screens, the quadrant keys are used to select the devices that will be controlled and the center key to go back to the last screen. Selecting a device, the quadrant keys send actions to the device, and the center button goes back to the last screen. The MENU key has not been implemented yet, not allowing configurations by users.

IV. DESIGN EVALUATION

In order to have some feedback on our design, it was presented to 10 persons with different impairments – 3 blind, 1 deaf, 1 wheel chair user, 1 motor-impaired, 2 elderly, 2 cognitive.

We conducted individual interviews guided by a questionnaire lasting around 30 minutes each. The interviews were divided into two phases; in the first one, the concepts of home networking and home automation are presented and data about the individual is collected: name, age, information about their impairment, their informatics and technical knowledge level, their perception about the home network value and priorities. The second phase of the interview was directly related to the Interface evaluation. Firstly, we presented the interface and showed them how to use it. Secondly, we let them play with it for some minutes. Thirdly, we presented some challenges (turning off the kitchen lamp, for example). Finally, we followed the questionnaire, asking for the feedback concerning many aspects of the interface.

With the interviews, we concluded that the quadrant approach design could successfully lead to a universal design. The users evaluated the interface positively. In all the criteria (easy to use, layout, subservience) the interface got 80% of the interviewees' highest score. In addition, in our perception, most users could meet the proposed challenges after a short learning time (up to 5 minutes).

Regarding the perceived importance of the home networking deployment at home in order to achieve autonomy and comfort, all of the interviewees were noticed to consider safety aspects of high importance. As a second degree of importance, but still very desirable, is saving technological aids (energy and water).

V. CONCLUSIONS

Despite working with a considerably varied group of users, with different needs, we could achieve an interface suitable to them. Our interface integrates accessible interface ideas in a single portable interface that can contribute to people with disabilities autonomy at home.

ACKNOWLEDGMENT

We would like to thank "Instituto EFORT" for helping us with preparing and conducting the interviews. We would also like to thank to the volunteers and to Emerson Moretto.

REFERENCES

[1] Costa, L.C.P.; Ficheman, I.K.; Correa, A.G.D.; Lopes, R.D.; Zuffo, M.K. "Accessibility in digital television: designing remote controls," Consumer Electronics, IEEE Transactions on, vol.58, no.2, pp.605-611, May 2012.

[2] Brazil. "Federal Decree no 5296". December 12nd, 2004. Brazil.

[3] Zhao, Q.Y.; Xu, S.; Li, Z.Z.; Wang, L. "A comparative study of musical navigation methods for visually impaired users of GUI systems," Industrial Engineering and Engineering Management, 2007 IEEE International Conference on, vol., no., pp.446-450, 2-4 Dec. 2007.

[4] EBU Technical, "Report on Access Services" Information I44-2004.

[5] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

A Sketch-based Interface for Remote Robot Control on an Online Video Screen

Soon Mook Jung, Dongwon Choi, Keyho Kwon, and Jae Wook Jeon, *Member, IEEE*
College of Information and Communication Engineering, Sungkyunkwan University
kuni80@ece.skku.ac.kr, cdwtuna@naver.com, {jwjeon, khkwon}@skku.edu

***Abstract*--This research proposes an interface method to make a connection between a mobile device, such as a smart phone or a smart pad, and the real world we live in. The robot is only used as a medium between the mobile device and the real world. To make for intuitive and easy interaction, we designed a sketch-based interface, using a touch-panel mobile device.**

I. INTRODUCTION

Due to the emergence of high-end mobile devices, such as smart phones or smart pads, users have recently been provided with a wide range of services, beyond compare with the previous generation. In everyday life, in particular, as the proportion of these portable wireless devices increases, existing computer-based services have been transformed into a new form of mobile environment, so that these services are available anytime and anywhere. Mobile social networking is one of the most common services. Many people share each other's thoughts through this online service. Unfortunately, however, this is limited to the virtual world. There is no way to make direct contact with the real world, such as chat with colleagues at an office. Hence, a medium to be able to make connection with the real world is required. Existing computer-based telepresence have been developed for this purpose. However, there is a spatial constraint, since they are fixed in specific locations, such as conference rooms. We expect people to become more deeply involved in various social activities, without constraints of time and space, using mobile devices. A telepresence robot is suitable for this purpose. Using this robot, people can reflect their intention in the real world, even though they are not there. This is due to the mobility of the robot. Hence, we consider that the user interface is the most important element for remote robot control. Since the main users of the mobile device are ordinary people, including children and elderly people, who are not professional in their use of technology, the interface should be intuitive and easy. Various types of interface for remote robot control have been proposed in the field of human computer interaction. The most general interface is to control a remote robot using buttons, while the user looks at a video screen [1].

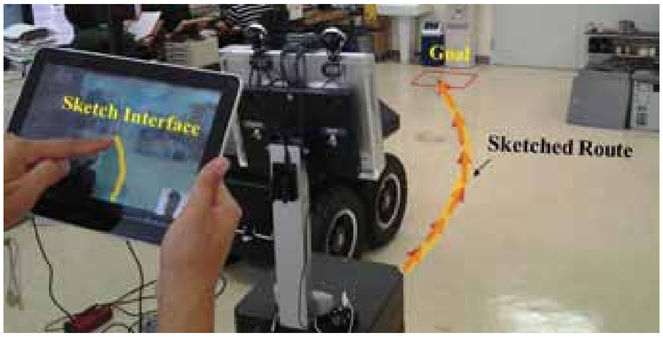


Fig. 1. A sketch-based interface using a touch-panel mobile device.

However, the narrow field of view makes it difficult to control the robot safely and quickly. So, more intuitive and easy interfaces are required. One of the most common methods is the sketch-based interface. The sketch-based interface, using a touch-panel device is proposed [2]. This allows the user to control a house-cleaning robot, by sketching its behavior on a top-down view from four ceiling cameras. However, there is a spatial constraint, since the robot is only controlled within the scope that the cameras are installed. Another method uses an environmental map [3]. The user sketches the moving path of the robot on the virtual map. Even though this interface is intuitive, and makes it easy to control the robot, the map should be pre-built. Our research proposes a sketch-based interface to be able to control the remote robot, without spatial constraint or pre-built map. The user sketches the route to move the remote robot, on the online video streaming from the robot. Other commands, in addition to the route generation, are also generated by the sketch method, and not by the button method. This interaction is only performed with one finger on the touch-panel screen. Fig. 1 shows our human to robot interaction using a touch-panel mobile device.

II. SYSTEM IMPLEMENTATION

A. Telepresence robot

The robot system is built to a height of 1.2m to talk with people, and it is equipped with a 19 inch LCD monitor, to show the life-size face of the remote robot pilot. Two low-cost web cameras are installed on the top of the monitor. Two images captured from these cameras are used to get the depth information of pixels constituting an image. Voice data is input via the microphone embedded in the webcam, and output via the mini speaker installed on the robot platform. The robot sends the image data (jpeg, 640×480), depth data of the image, and the voice data (wav, 16bps) to the mobile device. Fig. 2 shows the data transmission between the robot

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2012-(H0301-12-3001)) supervised by the NIPA (National IT Industry Promotion Agency), and by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0005861).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

and the device.

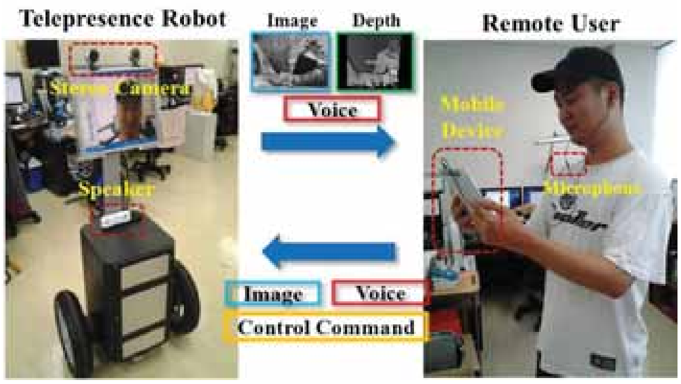


Fig. 2. The data transmission between the telepresence robot and the mobile device via a wireless module.

B. *Route generation based on the sketch interface*

The user sketches the moving route of the robot with one's finger on the touch-panel screen. This generated path is divided into sub-routes as shown in Fig. 3. Depth information of the pixels corresponding to the start point and the end point of the sub-route is used to calculate the actual length and direction of the sub-route. That is, the path generated on the touch screen is converted to the path of the real environment.

C. *Command generation based on the sketch interface*

The robot command is generated by the sketch method, using one finger. The robot commands are pre-defined as the gesture form. These commands are overlapped on the streaming video screen as shown in Fig. 4. Therefore, the user does not need to memorize any gesture for the robot control. To recognize these gestures, we adopt the Levenshtein edit distance algorithm. This algorithm is used to measure the similarity between two strings of different length. The similarity is calculated by counting the edit processes carried out to make different strings be the same. Hence, the gesture motion is converted to a string array consists of numbers between 0 and 7 according to its trajectory. The string array of the user gesture and the pre-defined gesture can be expressed as $x = \{x_1 x_2 x_3 \cdots x_k\}$ and $y = \{y_1 y_2 y_3 \cdots y_m\}$, respectively. The lengths of x and y are k and m, respectively.

TABLE I
LEVENSHTEIN EDIT DISTANCE

```
Input : User string array x and Pre-defined string array y
Output : Levenshtein Distance between x and y

Length of x : $k$,  Length of y : $m$
Create an Array $LD[0…m][0…k]$ of $(m+1)*(k+1)$
for (i = 0 to $k$)   $LD[0][i] = i$;            // First row initialization
for (j = 0 to $m$)   $LD[j][0] = j$;            // First column initialization

for (j = 1 to $m$) {
    for (i = 1 to $k$) {
        if($x_i = y_j$)   e-cost = 0;  else   e-cost = 1;     // Substitution cost
        $LD[j][i] = Min(\underbrace{LD[j-1][i]+\mathbf{1}}_{Insertion}, \underbrace{LD[j][i-1]+\mathbf{1}}_{Deletion}, \underbrace{LD[j-1][i-1]+\mathbf{e\text{-}cost}}_{Substitution})$;
    }
}
$LDist = LD[m][k]$;                  // Levenshtein Distance between x and y
```

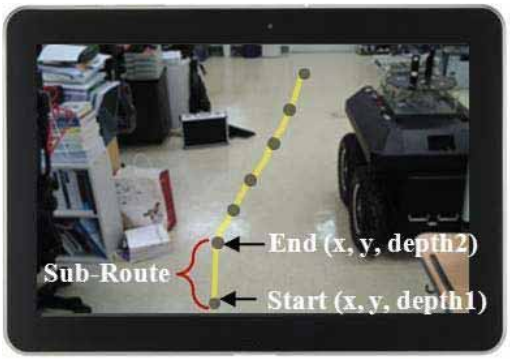


Fig. 3. A route generation using one finger sketch.

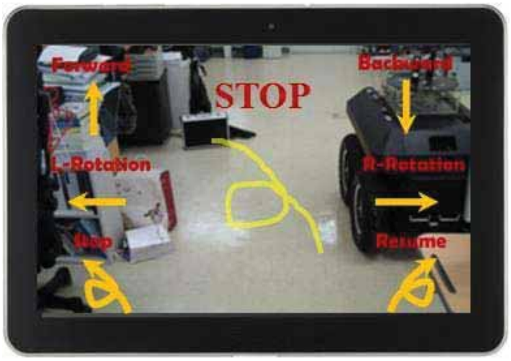


Fig. 4. Command generation using one finger sketch.

The Levenshtein distance algorithm is shown in Table I. Lastly, the Levenshtein distance between string array x and string array y is stored in the string array $LD[m][k]$. Using this method, the gesture motion corresponding to the string array which has the smallest Levenshtein distance is taken as the robot command.

III. CONCLUSION

In the near future, as robots will be more pervasive in our life, the interfaces for controlling them will be more important. And a convenient user interface for robot control also boosts the utilization of the robot. Hence, the proposed sketch-based interface for the remote robot control was made to be intuitive and simple, so this interface on a touch-panel mobile device can be used easily for everyone. However, the proposed method was implemented to target a static environment. The interface in a dynamic environment will be implemented with the additional sensor such as a laser scanner in the future work.

REFERENCE

[1] D. A. Lazewatsky, and W. D. Smart, "An inexpensive robot platform for teleoperation and experimentation," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1211-1216, May. 2011.

[2] D. Sakamoto, K. Honda, M. Inami, and T. Igarashi, "Sketch and run: a stroke-based interface for home robots," *CHI 09 Proceedings of the 27th international conference on Human factors in computing systems,* pp. 197-200, April. 2009.

[3] M. Negulescu, and T. Inamura, "Exploring sketching for robot collaboration," *HRI 11 Proceedings of the 6th international conference on Human-robot interaction,* pp. 211-212, March. 2011.

Pointing Gesture-based Large Display Interface with Automatic Display-Camera Calibration

Daehwan Kim and Ki-Hong Kim
Creative Content Research Laboratory, Electronics and Telecommunications Research Institute, South Korea

***Abstract*—We propose pointing gesture-based large display interaction using a depth camera. A user interacts with applications for large display by using pointing gestures with the barehand. The calibration between large display and depth camera can be automatically performed by using RGB-D camera. The experimental result shows that the pointing accuracy is 96.2%. We demonstrate the puzzle game application using the proposed system.**

I. Introduction

Currently most and popular interface devices are a mouse, keyboard, sensors, touch screens, and so on, but they are inconvenient and unnatural to control a large display[1]. As the size of the display could be as large as 100 inches, it encourages new natural interfaces[2]. The best way is to understand gestures or human motions because of its intuitiveness and naturalness. Especially, the pointing gesture is very powerful to control a large display.

We propose pointing gesture-based large display interface system, which naturally interacts with applications such as presentation, education, and entertainment. The proposed system calibrates between the large display and the depth camera by automatically searching them and also recognizes the pointing gesture directly. Fig. 1 shows the overall process of the proposed interface system.

In the paper, we describe our proposed system in detail. Then we present the experiment results of this system using 150 inch projection screen and a puzzle game application.

II. Pointing gesture-based large display interface system

There are two processes: calibration(off-line) and pointing gesture recognition(on-line). The calibration process is to estimate the 3D plane of the large display in the depth camera coordinate. The 3D plane is simply taken from searching 3D positions of 4 corner points of the large display.

The pointing gesture recognition process is to find the point of intersection between a pointing ray and the display plane. The point is calculated using two points(hand and shoulder) and the estimated display plane equation.

A. Automatic display-camera calibration

The proposed calibration consists of three steps:camera and 4 corner points detection, coordinate conversion, and display plane estimation.

First, we use another RGBD camera to detect the optical point $O_c = (x_c, y_c, z_c)$ of the camera and 4 corner points $C_1 = (x_1, y_1, z_1)$, $C_2 = (x_2, y_2, z_2)$, $C_3 = (x_3, y_3, z_3)$, and $C_4 = (x_4, y_4, z_4)$ of the large display, which are taken by using the Scale-invariant feature transform (SIFT) [3] and Harris corner detection algorithm in 2D color image, respectively.

Second, we convert the xyz coordinate for RGBD camera to the xyz coordinate for depth camera. The conversion is done by subtracting O_c from C_1, C_2, C_3, and C_3.

$$\begin{aligned}\vec{d_1} &= (dx_1, dy_1, dz_1) = (x_1 - x_c, y_1 - y_c, z_1 - z_c)\\ \vec{d_2} &= (dx_2, dy_2, dz_2) = (x_2 - x_c, y_2 - y_c, z_2 - z_c)\\ \vec{d_3} &= (dx_3, dy_3, dz_3) = (x_3 - x_c, y_3 - y_c, z_3 - z_c)\\ \vec{d_4} &= (dx_4, dy_4, dz_4) = (x_4 - x_c, y_4 - y_c, z_4 - z_c)\end{aligned} \tag{1}$$

Third, we take the 3D display plane equation using the converted 4 corner points. The linear homogeneous equation 2 is solved by the singular value decomposition (SVD).

$$\begin{bmatrix} dx_1 & dy_1 & dz_1 & 1\\ dx_2 & dy_2 & dz_2 & 1\\ dx_3 & dy_3 & dz_3 & 1\\ dx_4 & dy_4 & dz_4 & 1\end{bmatrix}\begin{bmatrix} a\\ b\\ c\\ d\end{bmatrix} = 0 \tag{2}$$

B. Pointing gesture recognition

The Pointing gesture recognition also consists of three steps: hand and shoulder detection, pointing ray estimation, and intersect point calculation. First, we take the hand point $H = (h_x, h_y, h_z)$ and shoulder point $S = (s_x, s_y, s_z)$ by using articulation detection modules[4]. Second, we calculate the pointing ray equation using the two points.

$$\frac{x - h_x}{s_x - h_x} = \frac{y - h_y}{s_y - h_y} = \frac{z - h_z}{s_z - h_z} \tag{3}$$

Third, we calculate an intersection point P by inserting the shoulder(or hand) point S (or H) and the slope $\vec{d} = (dx, dy, dz) = (s_x - h_x, s_y - h_y, s_z - h_z)$ to the line equation.

$$P(t) =< dx, dy, dz > t + (s_x, s_y, s_z) \tag{4}$$

If it inserts the P to the display plane equation,

$$aP(t)x + bP(t)y + cP(t)z + d = 0 \tag{5}$$

We get the parameter t from solving two equations Eq. (4) and Eq. (5).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

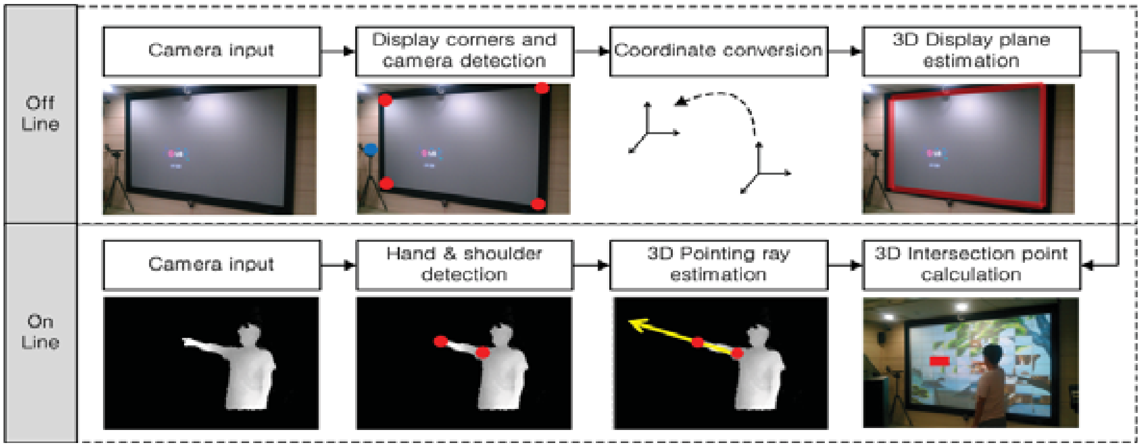

Off Line
Camera input
Display corners and camera detection
Coordinate conversion
3D Display plane estimation
On Line
Camera input
Hand & shoulder detection
3D Pointing ray estimation
3D Intersection point calculation

Fig. 1. Overall process of the large display interface system.

$$adxt + aA + bdyt + bB + cdzt + cC + d = 0 \tag{6}$$
$$t(adx + bdy + cdz) + aA + cB + cC + d = 0 \tag{7}$$
$$t = -(aA + bB + cC + d)/(adx + bdy + cdz) \tag{8}$$

Then, we take the intersection point P by inserting t to the line equation Eq. (4).

III. Experiment and Application

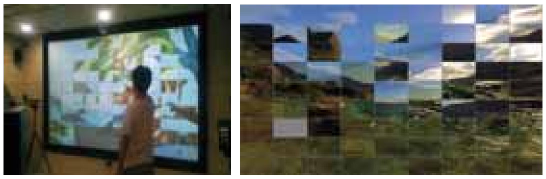
The proposed interface system was implemented on a Window PC platform with a 2.83 GHz Intel Core 2 Quad CPU and 6 GB RAM in the Microsoft Visual C++ environment. The 150 inch projection screen and TOF camera were used as the large display and the capture device for pointing gesture, respectively. The image size is 640(w)x480(h) and its frame rate is 30 fps.

We evaluated the accuracy of the proposed system by pointing a piece of screen with different division sizes: 3x3, 5x5, 7x7, and 9x9. Five humans joined the accuracy test. Table I shows the pointing accuracy, which was about 96.2%. The pointing interface system almost completely operated at the size of 3x3, 5x5, and 7x7, while it had some missed pointing results at the size of 9x9.

TABLE I
Pointing accuracy of different division sizes

Division size	Pointing accuracy				
	$H1$	$H2$	$H3$	$H4$	$H5$
3x3	9/9	9/9	9/9	9/9	9/9
5x5	25/25	25/25	25/25	25/25	25/25
7x7	48/49	48/49	47/49	48/49	47/49
9x9	75/81	77/81	77/81	76/81	76/81

We implemented a puzzle game application to demonstrate if the proposed system can be used or not. The game is to complete the puzzle by pointing pieces on the large screen. Fig. 2 shows the pointing gesture-based puzzle game. As a result, we demonstrated that our system can be applied some applications for large display.



(a) Interface system. (b) Puzzle game.

Fig. 2. Pointing gesture-based large display interface.

IV. Conclusion and Future Work

In this paper, we proposed a pointing gesture-based large display interface system using a depth camera. The proposed system automatically calibrated between the large display and the depth camera and recognized the pointing gesture directly. Our experimental results also showed that our proposed interface can use some applications for large display by implementing the puzzle game. From our experiments, we found that all partitions of the large display were not regularly pointed. We will solve the problem to improve the pointing accuracy.

Acknowledgments

This work was supported by the R&D program of Ministry of Culture, Sports and Tourism (MCST) and Korea Evaluation Institute of Industrial Technology (KEIT) (10039923, Development of Live4D contents platform technology based on expansion of realistic experiential space).

References

[1] W. Fikkert, P. Vet, H. Rauwerda, T. Breit, and A. Nijholt, *Gestures to intuitively control large displays*, Gesture-Based Human-Computer Interaction and Simulation, pp.199-204, 2009.

[2] C. Yeung, M. Lam, H. Chan, amd O. Au, *Vision-based hand gesture interactions for large LCD-TV display tabletop systems*, In Proc. 9th Pacific Rim Conference on Multimedia, pp.89-98, 2008.

[3] L. David, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, vol.60, no.2, pp.91-110, 2004.

[4] A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, *Real-time human pose recognition in parts from single depth images*, In Proc. of IEEE Compute Vision and Pattern Recognition, pp. 1297-1304, 2011.

Codebook based Stereo Matching for Natural User Interface

Sung-il Kang and Hyunki Hong, *Non-Member, IEEE*

***Abstract*-- This paper presents a stereo matching system using a codebook for natural user interface (NUI). Both color and disparity information in previous frames are stored and compared to deal with the occlusion problem in stereo matching.**

I. INTRODUCTION

Interactive user interface enabling the user to control the machine naturally has been one of the major topics in consumer electronics. The gesture based user interface is widely researched in smart home applications [1,2]. For example, home entertainment systems using gesture recognition is developed such as interactive smart TV, Nintendo Wii, Sony PlayStation3 Move, and Microsoft Kinect.

A Kinect sensor using an infrared band is able to capture precise range information, but it is mainly restricted to indoor application. Vision based systems, for typical passive sensing in indoor as well as outdoor environment, are widely classified into mono and stereo camera methods.

In the mono camera, image differencing over sequence and face detectors using skin color information are used to determine the areas where a person is, and the areas where the hands are [3]. However, color distributions are sensitive to changes in the lighting conditions and background, and human gestures are described in only 2D domain.

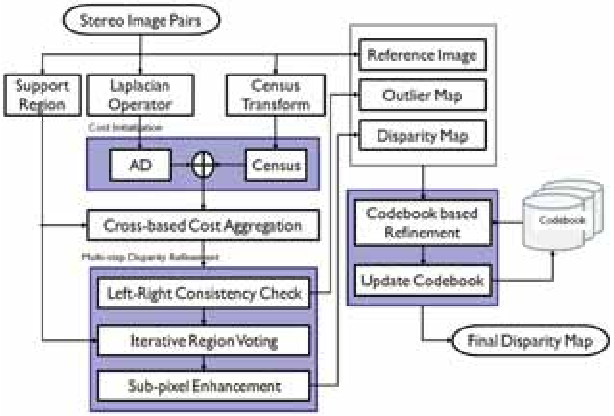


Fig. 1. Proposed block diagram.

Although the stereo system obtains the color and depth information about a human gesture, a high computation load is needed for a dense stereo matching [4]. Furthermore, there are many image ambiguities in stereo matching due to the occlusion problem. The proposed system is implemented on GPGPU for real-time performance, and we employ a codebook to solve the occlusion problem by the moving foreground object in dynamic situation.

Previous codebook extracts the background information without the foreground in advance, assuming the static situation that both the background and the camera are fixed [5]. However, the background changes and the noise often give many effects on the performance of the codebook. In order to deal with the background changes and the occlusion by the foreground, the proposed codebook stores both the color and the depth information for several frames. Our contribution is indicated by the blue-colored boxes in figure 1. The improved stereo system is applied to human gesture recognition in interactive multimedia application.

II. PROPOSED SYSTEM

In the stereo views, the left and right images are different from each other because of the effects by the different illumination condition and the surrounding environments. Using the longer baseline length allows us to handle a larger space, but the difference between two views is much increased.

An initial matching cost volume at each pixel and each disparity level is computed using AD (absolute differences)-Census in parallel, which combines the AD measure and the census transform [6]. Because the AD measure examines just the pixel intensity, it is much affected by the lighting changes. However, census encodes local image structures with relative orderings of the pixel intensities other than the intensity values themselves, and tolerates outliers due to radiometric changes and image noise. The proposed algorithm employs Laplace of Gaussian (LoG) filter as a pre-processing for AD-Census to alleviate the lighting effects.

The matching ambiguities and noise in the initial cost volume are reduced using the cross-based aggregation. Each pixel’s cost is aggregated over a support region with both the color similarity and the length constraint. Because there are still many mismatched regions by the occlusions, we examine the left-right consistency to detect the outliers by the occlusions. In the previous methods, the outliers are filled with reliable neighboring disparities in the segmented or the support region by the iterative region voting [6,7]. However, when the outlier region is too large or the depth is much different from that of the neighboring area, the iterative region voting would be unsuccessful.

The proposed codebook stores both the intensity values and the disparity as a codeword at pixel **p** before the occlusion by the foreground is occurred. Table 1 determines the refined disparity Δ^* at the current position **p** in the region taking account the color value **x** and the disparity Δ by the cross-based aggregation. Here, colordist(,) represents L-1 color

This research was supported by Basic Science Research Program (2010-0010398) and the second stage of the BK21 program through NRF funded by the Ministry of Education, Science and Technology.

distance measure of two pixels in the cost volume and ε_c is a color threshold.

TABLE I
CODEBOOK BASED DISPARITY REFINEMENT

```
I. The position p at current frame $t$, color information $\mathbf{x} = (R, G, B)$ and its depth $\Delta$.
II. $I \leftarrow \sqrt{R^2+G^2+B^2}$ , the final depth map is initialized ($\Delta^* \leftarrow 0$).
III. If occlusion is,      // $\Delta^*$is the matched codebook's depth as follows.
    (i) Find the codeword $\mathbf{c}_m$ in $F = \{\mathbf{c}_m \mid \mathbf{c}_m \in C\}$ satisfied with two conditions (a), (b), and minimized the condition (a).
        (a) colordist($\mathbf{x}$, $\mathbf{v}_m$) $\leq \varepsilon_c$
        (b) brightness($I$, $\langle I_{min}, I_{max}\rangle$) = true.
    (ii) If $F \neq \emptyset$, $\Delta^* \leftarrow \Delta_m$.
IV. Otherwise,
    (i) $\Delta^* \leftarrow \Delta$.
    (ii) Find the codeword $\mathbf{c}_m$ in $G = \{\mathbf{c}_m \mid \mathbf{c}_m \in C \wedge |\Delta - \Delta_m| < \varepsilon_\Delta\}$ satisfied with two conditions (a), (b), and minimized a condition (a).
        (a) colordist($\mathbf{x}$, $\mathbf{v}_m$) $\leq \varepsilon_c$
        (b) brightness($I$, $\langle I_{min}, I_{max}\rangle$) = true.
    (iii) If $G \neq \emptyset$,
        // update the satisfied codeword $\mathbf{c}_m$, consisting of $\mathbf{v}_m = (R_m, G_m, B_m)$
        - $\mathbf{v}_m \leftarrow (\frac{R_m+R}{2}, \frac{G_m+G}{2}, \frac{B_m+B}{2})$ , $\Delta_m \leftarrow (\frac{\Delta_m+\Delta}{2})$
        - $I_{min,m} \leftarrow \min\{I\times0.8, I_{min,m}\}$, $I_{max,m} \leftarrow \max\{I\times1.2, I_{max,m}\}$, $t_m \leftarrow t$
    (iv) Otherwise,                // there is no match,
        - $L \leftarrow L+1$            // Increase the codeword at p.
        - $\mathbf{v}_L \leftarrow (R, G, B)$    // Then add a new codeword $c_L$ as follows.
        - $\Delta_L \leftarrow \Delta$,   $I_{min,L} \leftarrow I\times0.8$,   $I_{max,L} \leftarrow I\times1.2$,   $t_L \leftarrow t$
```

A function brightness (I, $\langle I_{min}, I_{max}\rangle$) determines whether an input value I is between the minimum and the maximum of the codebook. The outlier pixels by the occlusions are filled with the disparity codeword of the codebook satisfying the condition III(i) and having the smallest color distance. In case of no occlusions, we determine that the reliable disparity is found. The codeword information satisfying both the color distance and the brightness range is averaged with the input values and then the codebook is updated. If there is no match, we add a new codeword including the color, depth, brightness range and frame number. The codeword that has been not matched for some period (10 frames) is deleted for memory efficiency.

In the next step, a sub-pixel enhancement process based on quadric polynomial interpolation is performed to reduce the errors by discrete disparity levels [7]. The final disparity results are obtained by smoothing the interpolated results with a 3×3 median filter.

The computational equipment includes an Intel Quad 2.66GHz with Nvidia GTX460. The stereo images are captured by a Bumblebee 3 from Point Grey Inc. Stereo matching is implemented on GPU and the codebook generation and its evaluation is on CPU. The total processing time is 80~110ms. In comparison results on synthetic sequences [9], the proposed system holds the second rank among competitors. Here the first two methods have non-realtime performances. However, when many occlusions by the foreground in interactive application are happened as figure 2, the proposed method outperforms the previous.

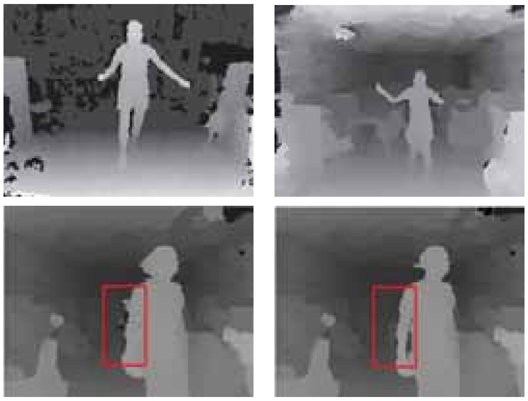


Fig. 2. Disparity results of Kinect, proposed system (upper); comparison of AD-Census [6] and proposed (lower)

REFERENCES

[1] M. Chen, L. Mummert, P. Pillai, A. Hauptmann, and R. Sukthankar, "Controlling your TV with gestures," *Proc. of Int'l. Conf. on Multimedia Information Retrieval*, 2010.

[2] S. Lin, Y. Lai, L. Chan, and Y. Hung, "Real-time 3D model-based gesture tracking for multimedia control," *Proc. Int'l. Conf. on Pattern Recognition*, 2010.

[3] P. Viola and M. Jones, "Robust real-time object detection," *Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.

[4] H. P. Jain and A. Subramanian, "Real-time upper-body human pose estimation using a depth camera," *Technical Report, HPL-2010-190*, HP Laboratories, 2010.

[5] K. Kim, Th. Chalidabhongse, and D. Harwood, "Real-time foreground background segmentation using codebook model," *Real-time imaging*, vol. 11, no.3, pp. 172-185, 2005.

[6] X. Mei, X. Sun, M. Zhou, H. Wang, and X. Zhang, "On building an accurate stereo matchng system on graphics hardware," *Proc. of GPUCV*, pp. 467-474, 2011.

[7] Q. Yang, C. Engels, R. Yang, H. Stewenius, and D. Nister, "Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 492-504, 2009.

[8] K. J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 650-656, 2005.

[9] C. Richardt, D Orr, I Davies, and A Criminisi, "Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid," *Proc. of ECCV*, 2010.

TABLE II
UNITS FOR MAGNETIC PROPERTIES

Algorithm	Book		Street		Tanks		Temple		Tunnel	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
Adapt weight[8]	84.2	1.24	56.1	2.67	87.7	2.01	72.8	1.80	58.4	11.7
Dichromatic DCB grid[9]	58.9	1.83	39.2	2.62	47.8	12.0	43.0	1.73	32.9	12.0
Temporal DCB grid[9]	44.0	2.02	25.9	2.00	31.4	6.06	31.7	1.82	36.4	7.88
Proposed approach	66.1	1.41	35.5	1.88	52.0	4.39	32.5	2.51	36.6	13.5

Cost Effective Smart Remote Controller Based on Invisible IR-LED Using Image Processing

Yunjung Park[1], and Minho Lee[2]
[1]Department of Robot Engineering, Kyungpook National University
[2]School of Electronics Engineering, Kyungpook National University

***Abstract*—We present a new cost effective smart remote controller using only a camera which may be installed or attached on digital appliances and infrared (IR) LED to control various electronic appliances, such as digital smart TV, air conditioner and so on. The users can easily operate the invisible IR LED to make the specific command by simply blinking the IR LED. Then, the proposed system can easily analyze the state of IR LED and understand the human intention by image processing using a built-in embedded processor within the digital appliances. Therefore, the proposed system can directly understand and generate specific command signal to control the digital appliances without any communication circuit between remote controller and digital devices unlike the previous remote controllers. Experimental results show that the proposed remote controller is easy to use and can successfully operate to execute human intent command.**

I. Introduction

Recently, smart appliances are being widely used and their control methods are getting complex. In other words, the remote controllers are becoming expensive with complicated functions. For instance, a smart TV remote controller should possess the functions of a computer 'MOUSE' for controlling GUI, in addition to the functions of a general TV remote, such as button input. Even though, the studies related to TV controlling using smart phone applications are in progress [1-2], the implementation of these functions in the remote controller becomes expensive and complex.

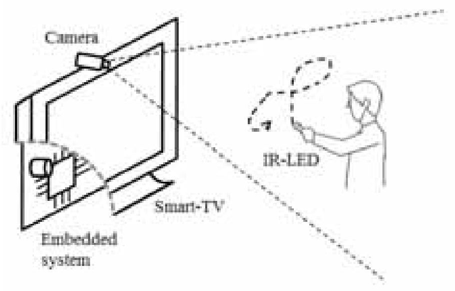


Fig 1. Concept of proposed system

In this paper, we propose a remote controller which has functions to move the cursor and select the menu in smart digital appliances such as TV systems that include GUI environment, similar to a computer mouse. The proposed system consists of an USB camera to receive the IR light from the source in the remote controller as shown in Fig. 1. The built-in camera and the embedded processor within the digital appliances were used to analyze the received image without extra devices such as communication circuits.

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology(2012K001342)

When the user controls the IR light of remote controller in viewing angle of cameras, the proposed system understands the user command by calculating the location of light source and the number of blinks during a certain number of successive frames. Thus the proposed system is able to transfer information of the button control without any special hardware circuitry. Since the proposed remote controller has simple control parts such as IR LED, switch and battery, it is very cost effective and easy to operate with simple design.

II. IR LED Based Remote Controller

As shown in Fig. 2, the proposed system consists of three parts: 1) preprocessing to gather the coordinates of IR LED based controller, 2) IR LED blinking recognition, and 3) IR LED controller's gesture recognition.

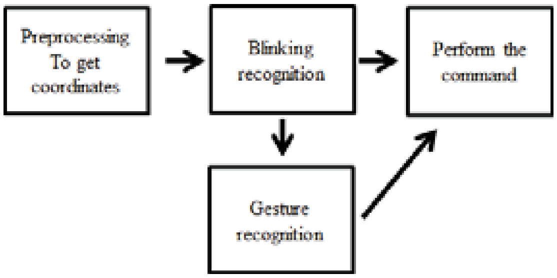


Fig 2. The flow chart of proposed HMI system

A. Preprocessing

In real-world environment containing noise and high intensity light sources preprocessing is necessary to successfully obtain the IR LED coordinates. The preprocessing includes two steps: 1) adjust the camera parameter settings and 2) continuously update the reference frame.

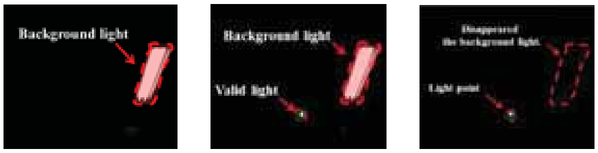
1) Camera parameter setting

The IR filter is removed to alleviate the noise and to easily detect the invisible IR LED. The exposure time of the camera is set to minimum by digital option to alleviate the effect of other light sources.

2) Continuously update the reference image

As the digital appliances are fixed on a specific region, we can easily analyze the IR LED coordinate by comparing the reference image with and without IR LEDs. Generally, within a camera image IR LED has highest intensity and a simple discrimination technique can be used to classify the IR LED or non-IR LED regions. After the discrimination process, the other light sources such as a fluorescent light can be removed by simple image subtraction as shown in Fig. 3. In order to obtain the accurate coordinates, we continuously update the reference image through the running time.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a) (b) (c)

Fig 3. (a) Reference (b) Input (c) Subtracted (b – a) Images

B. IR LED blinking recognition

Based on the blinking pattern through successive frames, there are 5 different controller modes such as 1) Drag, 2) Click, 3) Scroll, 4) Double click, and 5) Gesture mode as shown in Fig. 4. When the user firstly turns on the IR LED, the user is able to move the cursor. After that the user can select 5 different controller modes by changing the states of the IR LED, such as turn on or off, during successive frames as shown in Fig. 4.

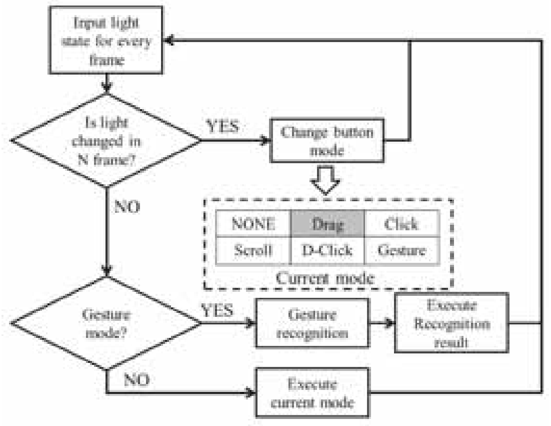


Fig 4. Flow chart for generating specific controller modes

During *N* successive frames if there exists a change of state of IR LED, the controller mode will be autonomously changed from one to the other mode. Depending on the number of changes in the states of IR LED, 5 different controller modes are sequentially changed from one to the other mode. Finally, if there exists no change in the states of IR LED, then the system operates in the current mode. In the case of gesture mode, the system can recognize the user friendly gesture by using neural networks to operate the specific functions [3].

C. Gesture recognition

The multi-layer perceptron (MLP) is used to recognize the moving patterns of the IR LED. The MLP is made up of 11 input nodes, including a fixed bias, 30 hidden nodes, and 8 output nodes. The MLP input consists of the directional components of an IR LED spot trace in collected images. Actually, in the gesture mode, the system collects the 11 coordinates for extracting 10 angular data between current and next points. The 11 coordinates of the detected IR LED spot are sampled through time and their directional information is used as the input to the MLP.

Its pattern information is compared with the pre-defined gesture information in the trained MLP. The MLP then generates a classification result for activating one of the 8 command signals [3].

III. EXPERIMENTAL RESULTS

In order to verify the proposed system, seven users were tested with the 5 different controller modes by using blinking pattern. As shown in Table I, most of the users can easily use the system even with less training time.

TABLE I
Mode selection recognition performance of

	Drag	Click	Scroll	D-click	Gesture	Total
Success ratio	100%	98.57%	95.71%	94.28%	92.85%	**96.48%**

Table II shows the recognition performance on each of the 8 different gestures. The overall recognition accuracy of the gesture recognition is 99.17%. In addition, the users can make their own functions by mapping between each gesture and desired functions.

TABLE II
Gesture recognition performance

Gesture	1 →	2 ←	3 ⊂	4 ⊃	5 ⤵	6 ↳	7 ⮍	8 ⮏	Total
success ratio	30/30 (100%)	30/30 (100%)	29/30 (96.67%)	30/30 (100%)	29/30 (96.67%)	30/30 (100%)	30/30 (100%)	30/30 (100%)	99.17%

IV. CONCLUSION

The proposed system can be used to accurately control the digital appliance such as smart TV without the consideration of external environment and distance between the system and the user. Contrary to general remote controllers which have various functions to control the general TV, it might be used for the restricted functions to simplify the usage method. But it is able to recognize clicks of the button and gestures of the user, so it is sufficient to be used as control device in smart TV and consists of low cost interface system. Also, users can easily implement their own functions, for examples, volume up or channel down as well as web surfing.

REFERENCE

[1] D. Zivkov, B. Majstorovic, T. Andjelic, M. Davidovic, and D. Simic, "Smart-Phone Application as TV Remote Controller,"Proceeding of the2011 IEEE International Conference on Consumer Electronics, pp. 431 - 432, Jan. 2012.

[2] Y. Hung, H. Chen, and S. Chu, "Content-aware Smart Remote C ontrol for Android-based TV," Proceeding of the2011 IEEE International Conference on Consumer Electronics, pp. 678 - 679, Jan. 2012.

[3] S. Jeong, C. Jung, C. Kim, J. Shim, and M. Lee, " Laser spot detection-based computer interface system using auto associative multilayer perceptron with input-to-output mapping-sensitive error back propagation learning algorithm," *Optical Engineering*, vol. 50, Aug. 2011.

Automatic Exercise Counter for Outdoor Exercise Equipment

Kyong Sik Choi, Yong Soo Joo, and Sang-Kyun Kim, *Member, IEEE*

***Abstract*-- In this paper, we present an algorithm that can count the number of trials on outdoor exercise equipment. Using sensed data from a three axis accelerometer, the proposed algorithm can reliably measure the number of trials on four different outdoor fitness machines. The experimental results prove that the proposed algorithm is robust to any deformation of the sensor location as well as the various types exercise patterns.**

I. INTRODUCTION

Recently many wellness related systems and services have been developed and released in the market not only for elderly people but also for general public. The systems and services mainly aim to prevent diseases and to maintain healthy life. USN-based Health park [1] is one of them. The park uses sensors attached onto exercise equipment to check the individual health status and measure impetuses.

One of the existing wellness related systems, which manage and guide a personal exercise program, is Fitlinxx [2]. Fitlinxx includes indoor fitness machines attached with sensors that can measure the number of trials. Fitlinxx, however, is expensive and designed only for indoor fitness activities so that it is not suitable for any outdoor applications. The literature [3] mentioned a little about measuring the number of trials from outdoor fitness machines using sensors, however, there were no details presented how to accomplish the task.

In this paper, we present an algorithm that can count the number of trials on outdoor fitness machines. We designed the algorithm under three requirements. The first requirement is that the algorithm shall be implemented in a rather inexpensive system. The second requirement is that the algorithm shall be robust to any deformation of the sensor location. The third requirement is that the algorithm shall be reliable upon the various types of difficulties such as inconsistent exercise speeds and ranges. We designed the proposed algorithm using a three axis accelerometer to fulfill the aforementioned requirements.

This paper is organized as follows. Section II explains the details of the proposed algorithm. Section III presents the experimental results. Finally, the paper is concluded in Section IV.

II. SENSING DATA FROM OUTDOOR-FITNESS EQUIPMENT

Figure 1 shows the sensed data from a three axis accelerometer. Main reasons to choose this sensor are that (1) an accelerometer can acquire acceleration data over time; (2) the sensor is relatively cheap and easy to install; (3) the data from a three axis accelerometer is robust to the deformation of the sensor location. In spite of any rotation of a sensor to any direction due to continuous impact over time, the three axis accelerometer sensor can produce consistent data.

Fig. 1. Sensed data from a three axis accelerometer.

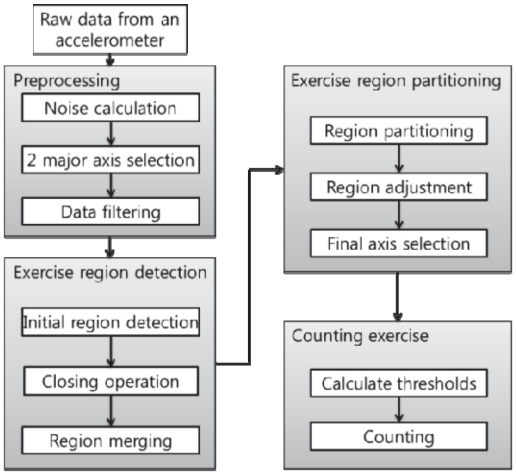


Fig. 2. Algorithm flow.

Figure 2 presents the overall algorithm flow. The preprocessing stage includes three sub steps. First of all, a threshold of noise level shall be estimated from the raw acceleration data. Noises can be occurred when a user stays still or makes a little movement on equipment. Those noises must be removed before counting. We estimated the noise level by accumulating and analyzing raw data of non-exercising conditions. One axis data with a minimum acceleration magnitude is then removed from a further calculation (e.g., a minimum among $X = \sum|x|, Y = \sum|y|, Z = \sum|z|$).

$$y = \frac{1}{n - (j \times 2)} \sum_{i=j}^{n-j} x_i \tag{1}$$

In order to exclude outliers and smooth the raw data, we used a filter combining median and average filters as shown in eq. (1). In eq. (1), *n* represents a mask size (e.g., 10); *j* is a median boundary (e.g., 2); the raw data x_i is sorted in an increasing order.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2010-0003212).

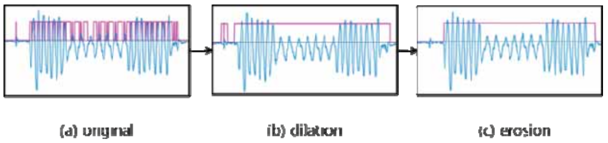


Fig. 3. Morphological close operation.

The pink lines in Figure 3 represent an actual exercising period. The initial exercise regions (Figure 3(a)) are detected by using the noise threshold calculated in the preprocessing stage. In order to figure out the actual exercise period, we merged the initial regions using a morphological close operation (Figure 3(b) and (c)). The magnitude of the dilation and the erosion is set to one second.

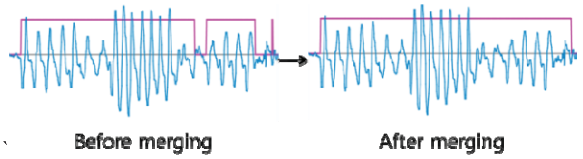
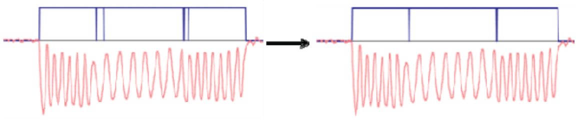


Fig. 4. Exercise region merging.

Even after the morphological close operation, there are still regions that are close enough to be regarded as a continuous exercise period. We merge the regions when an interval between two consecutive regions is less than or equal to two seconds. Figure 4 shows a result of the exercise region merging.

Finally, an exercise period obtained may contain several exercise patterns. For example, an exercise pace (e.g. speed) of a person would be fast at the starting point and gradually become slower. Some people change his/her pace intentionally during an exercise period. When the paces are changed, magnitudes of the acceleration are also changed. Therefore, we need to partition an exercise period in accordance with the exercise paces.



(a) (b)

Fig. 5. Exercise region partitioning and adjustment

Figure 5 presents an example of exercise region partitioning and adjustment. The partitioning occurs when a variation of exercise speeds exceeds 20%. The partitioning would generate exercise regions less than one second as shown in Figure 5(a). Figure 5(b) shows the adjustment of the short partitions. Each exercise region would include a consistent exercise pace. Then, a final axis is chosen among two axis data by eliminating the axis with a bigger frequency. We consider the axis with a bigger frequency contains more noises.

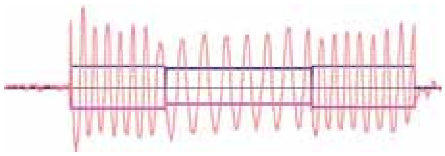
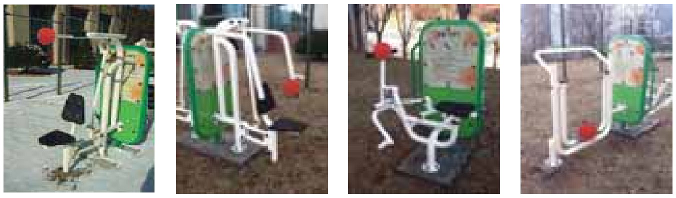


Fig. 6. Calculating thresholds for each exercise partition.

In order to distinguish explicit exercise movements, a threshold for each exercise region is then calculated using Otsu's threshold method [4]. Figure 6 shows the obtained thresholds using blue and pink lines. The acceleration data over thresholds becomes the data for counting the number of trials.

III. EXPERIMENTAL RESULTS



(a) (b) (c) (d)

Fig. 8. Outdoor fitness machines attached with an accelerometer; (a) a lat full down, (b) a chest press, (c) a rowing rider, (d) an air walker.

We tested the proposed algorithm to four outdoor fitness machines as shown in Figure 8. The accelerometer was attached to a place where the movement magnitude became maximized. The red dot in Figure 8 represents the place of the accelerometer attached. We tested total 192 exercise patterns which include various types of sensor locations, exercise speeds, ranges, and noises. The test results are shown in Table 1 where *x* is the actual number of trials and *y* is the number of trials calculated by the proposed algorithm.

Table 1. Test results.

$\lvert x-y \rvert$	Exercise Patterns	Rates
0	117	60.9375
1	67	34.89583
2	8	4.166667
Total	192	100

IV. CONCLUSION

In this paper, we presented an algorithm for converting low-level data from a three-axis accelerometer attached to outdoor exercise equipment to the number of workout trials. The algorithm was tested with various types of exercise patterns (192 patterns) with four outdoor fitness machines. The experimental results shows that the success rate of the proposed algorithm is over 95% of which the difference between the actual trial counting and the counting of the proposed algorithm is less than or equal to one.

REFERENCES

[1] Younghee Ro, Jaebom Joe, Hyunsun Ju, Hyunmin Park, and Jonghoon Chun, "Health-park: An RFID-based Exercise and Nutritional Management System for Ubiquitous Wellness Environment," In Proc. of 2011 Int. Conf. on Data Eng. and Internet Tech., Bali, Indonesia, March, 2011, pp. 834-837.

[2] Fitlinxx, http://www.fitlinxx.net/.

[3] Sang-Kyun Kim, Jonghoon Chun, Dongseop Kwon, Hyunmin Park, Younghee Ro, "Interfacing Sensors and Virtual World Health Avatar Application," NISS 2012, Macao, 2012.

[4] Nobuyuki Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*,V. SMC-9, NO. 1, 1979, pp. 62-66.

A Novel Iris Center Localization Method Based on the Spherical Eyeball Rotation Model for Human-Device Interaction

Kang-A Choi, Seung-Jin Baek, Chunfei Ma, Seung Park, and Sung-Jea Ko
School of Electrical Engineering, Korea University, Seoul, Korea

Abstract--**This paper presents a novel iris center (IC) localization algorithm for human-device interaction (HDI). In the proposed method, many different IC locations and their corresponding iris boundaries (IBs) are first registered as a database. Using the database, IB matching is performed to locate the IC. Experimental results show that the proposed algorithm outperforms the conventional ones especially when the iris is located at the corners of the eye.**

I. INTRODUCTION

The latest high-tech gadgets provide fresh perspectives on human-device interaction (HDI) allowing consumers to handle electronics in more intuitive ways. The HDI technology which uses the eye movements as the input information is attracting considerable attention since eyes and their movements are directly related to the user's desires and cognitive processes, as shown in Fig. 1.

Fig. 1. Example of the eye-controlled HDI.

In the eye-controlled HDI system, robust and elaborate iris center (IC) localization is the most important issue [1]. Several webcam-based IC localization methods have been reported. Valenti and Gevers employ isophote properties (i.e., curves connecting points of equal intensity) to locate the IC [2]. However, this approach often detects eyebrow or eye corner instead of the IC when the number of features within the eye region is insufficient. Based on the fact that the shape of the iris contour projected onto an image plane is an ellipse, the methods [3], [4] select two longest vertical edges (VEs) among all the edges of iris for ellipse fitting. After taking several steps to detect the edges of iris, two longest vertical ones are selected and exploited in the ellipse fitting process [5]. Then, the center of the resultant ellipse is regarded as that of the iris. The limitation of these methods is revealed when the user moves his/her eyeball to either side of the corner, since one of the two edges cannot be extracted due to the occlusion caused by eyelids or eye corners. This deteriorates the accuracy of IC localization occasionally.

To solve the aforementioned problem, we present a novel IC localization method for HDI. In the proposed method, many different IC locations and their corresponding iris boundaries (IBs) are first registered as a database. Using the database, IB matching is performed to locate the IC. Experimental results show that the proposed method can locate the IC robustly and precisely especially when the iris is located at the corners of the eye.

II. PROPOSED IC LOCALIZATION

Fig. 2 illustrates the overall procedure of the proposed IC localization algorithm. The proposed algorithm consists of a simple VE extraction method and spherical eyeball model (SEM) based elliptical IB matching. In this section, the proposed method is described in detail.

In general, the iris is the darkest region in an eye image. The proposed method finds a point minimizing the sum of the pixel values inside the window in the input eye image. In our implementation, the size of the window is set to one third that of the eye image. The detected point is regarded as the initial IC (IIC) point. From the grayscale eye image, a binary image is obtained using thresholding method. Then, VEs are extracted using the canny edge detector. Among several detected edges, the ones that are located within a pre-defined range from the IIC point are considered as the edges of interest. Two longest VEs from left and right sides of IIC are then chosen.

Fig. 3(a) shows the basic concept of the proposed SEM. Our SEM is based on the mathematical morphology [6] which provides how the 3-D rotation of sphere is represented on the 2-D plane. Let R_I and R_E, respectively, be the radius of iris and that of eyeball, and C_E be the 2-D projection of the eyeball

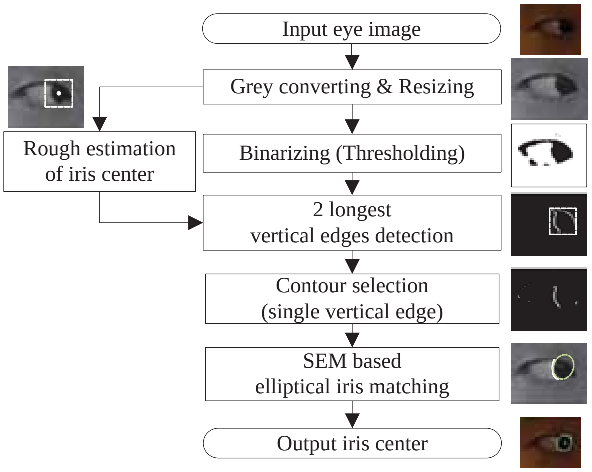

Input eye image
Grey converting & Resizing
Rough estimation of iris center
Binarizing (Thresholding)
2 longest vertical edges detection
Contour selection (single vertical edge)
SEM based elliptical iris matching
Output iris center

Fig. 2. Process flow of the proposed IC localization.

This work was supported by the Industrial Strategic Technology Development Program (10041656, Open Platform Development for See-Through Smart Glasses with Smart Vision) funded by the Ministry of Knowledge Economy(MKE, Korea)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

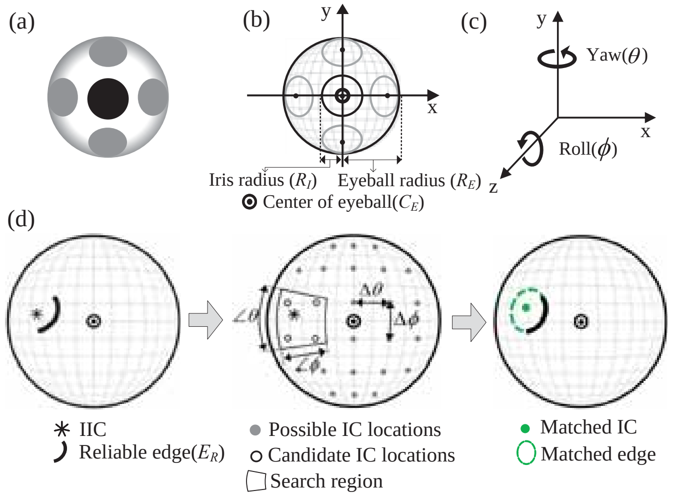


Fig. 3. IB matching process using SEM. (a) Variations in the shape of the iris due to the eyeball rotation in 3-D space, (b) different ICs and their corresponding IBs projected onto a 2-D image plane, (c) roll(ϕ) and yaw(θ) axes, and (d) an example of SEM based IB matching.

center as shown in Fig. 3(b). When the IC is located at C_E, its corresponding IB can be represented as a circle. Then, the points lying on the circle in rectangular coordinates (x, y) can be expressed in spherical coordinates as

$$(x,y)=(\delta\cos\alpha,\delta\sin\alpha),\ \ \alpha\in[-\pi,\pi], \tag{1}$$

where α stands for the central angle of the circle and δ is equal to R_I /R_E. As shown in Fig. 3(c), the points on the boundary can be transformed by performing yaw and roll rotations by θ and ϕ degrees respectively, and then projected onto the 2-D image plane as follows:

$$\begin{aligned}(x_r,y_r)=(&\delta\cos\theta\cos\alpha+\sqrt{1-\delta^2}\sin\theta)\cos\phi-\delta\sin\alpha\sin\phi,\\ &\delta\cos\theta\cos\alpha+\sqrt{1-\delta^2}\sin\theta)\sin\phi+\delta\sin\alpha\cos\phi),\end{aligned} \tag{2}$$

where x_r and y_r indicate the resultant x and y coordinates after the transformation, respectively. Note that the IC can be obtained when δ is set to zero. By adjusting θ and ϕ, many different IC locations and their elliptical IBs can be obtained. A pair of IC and its corresponding IB forms each candidate for the IB matching.

Fig. 3(d) shows an example of our matching process when a side-looking eye image is given. Assume that the IIC and the two VEs are obtained. Between two edges, the one with the shorter distance to C_E in x-axis is selected as the reliable edge (E_R). Next, a search region is defined by $\angle\theta$ and $\angle\phi$ near the IIC. Within the search region, the proposed method finds the candidate of which points are maximally overlapped with those of E_R. Finally, the corresponding center of the resultant candidate is determined as the final IC.

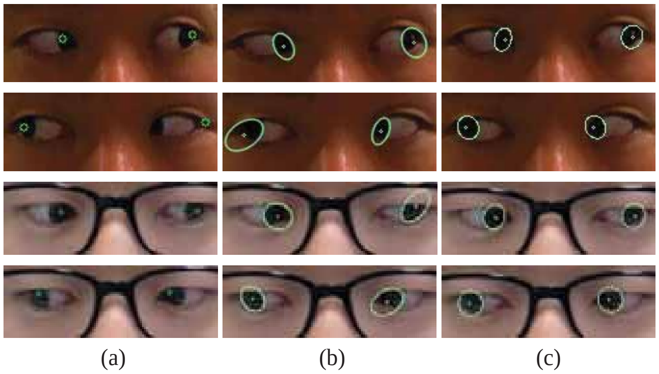


Fig. 4. Comparison of the conventional and proposed methods. (a) Valenti's, (b) Zhang's, and (c) the proposed methods.

III. EXPERIMENTAL RESULTS

For the experiment, we used a commercial webcam to capture test video. The spatial resolution of the input images is 800×600 and the distance between the screen and the user is fixed to 40cm.

In the current implementation, the proposed method initially performs conventional ellipse fitting for the iris [4]. When the resultant shape is a circle, it is assumed that the IC is located at C_E. Based on this assumption, R_I is set to the circle's radius, and R_E is determined to be the double size of R_I. Using the R_I and R_E, many different IC locations and their corresponding IBs are registered as a database. Then, the system performs the IB matching to locate the IC. Both $\Delta\theta$ and $\Delta\phi$, indicating how densely the iris matching process is conducted as shown in Fig. 3(d), are equally set to 5°. In addition, $\angle\theta=\angle\phi=\pm20^\circ$.

Two subjects, one with the glasses and the other with no glasses, were requested to stare at several points while seated in front of the screen. Fig. 4 shows the IC localization results when the users stare at left and right sides. Figs. 4(a) and (b) show the limitation of the isophote curvature method [2] and Zhang's ellipse fitting method [3] respectively. Although the iris is located at the corners of the eye, our proposed method presents accurate and reliable localization results as shown in Fig. 4(c).

REFERENCES

[1] D. Hansen and Q. Ji, "In the eye of the beholder: a survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, pp. 478-500, Mar. 2010.

[2] R. Valenti and T. Gevers, "Accurate eye center location and tracking using isophote curvature," in *Proc. CVPR*, 2008, pp. 1-8.

[3] W. Zhang, T.-N. Zhang, and S.-J. Chang, "Eye gaze estimation from the elliptical features of one iris," *Optical Engineering*, vol. 50, pp. 047003-1-047003-9, Apr. 2011.

[4] J.-G. Wang, E. Sung, and R. Venkateswarlu, "Estimating the eye gaze from one eye," *Comput. Vis. Image Und.*, vol. 98, pp. 83-103, Apr. 2005.

[5] A. Fitzgibbon, M. Pilu, and R.B. Fisher, "Direct least square fitting ellipse," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, pp. 476-480, May 1999.

[6] J. B. Roerdink, "Mathematical morphology on the sphere," in *Proc. SPIE*, vol. 1360, pp. 263-271, Sep. 1990.

Marker-based Tangible Inter[1]faces for 3D Reconstruction

Kyungboo Jung, Jong-Il Park, and Byung-Uk Choi
{Kbjung, jipark, buchoi}@hanyang.ac.kr
Dept. of Electrical and Computer Engineering, Hanyang University

***Abstract*-- In this paper, we proposed a novel marker-based tangible interface for various users to manipulate the object with intuitive and simple approaches during an authoring application of augmented reality. The proposed method makes use of marker as intuitive interface to obtain 3D geometric information of 3D reconstruction. 3D geometric information of an object surface is acquired by touching the object directly with the proposed tangible interfaces. The tangible interfaces not only support 3D reconstruction for graphical modeling but also offer features information which is used for augmented reality. Finally, we verify efficiency of the proposed method with demonstration of an augmented reality application using the proposed method.**

I. INTRODUCTION

Previous techniques of augmented reality focus on object tracking and recognition for properly augmenting virtual 3D information to fit the context of the circumference[1-3]. The early researches use markers for exactly estimating camera pose and recognizing objects. In recently, since object recognition methods are progressively developed, the development direction of augmented reality is to marker-less type.

In order to implement marker-less based augmented realty, the image processes such as stable object recognition, feature extraction, and acquiring 3D information are needed. The stable feature extraction and tracking among the methods are important for exactly extracting 3D geometry information. In general, a part of plane is used as natural marker to acquire 3D information of objects. However, the method have limit of camera viewpoint to recognize the plane. That is, a method for acquiring 3D geometrical information is needed to do not limit of object. Recently, although object recognition and feature tracking methods have been developed for marker-less based augmented reality, the methods still have some problems in stable object tracking, acquiring 3D information and management to offer to user.

In this paper, we propose marker-based authoring tools that can easily acquire 3D geometry information for object tracking as direct touch object by user. Feature extraction methods can be divided two categories for using in object tracking and application of augmented reality. The first method extracts features on specific regions after segmenting the image into object and background in offline. The second method extracts features in region that is selected by user in online.

In order to implement augmented reality, using markers to extract geometry information is primitive but exact method. If geometrical form is simple such as regular hexahedron, user can simply reconstruct the object. However, if geometrical form is complex, users must use application or device of 3D modeling and users need many costs and times to describe 3D geometry information. Therefore, we propose marker-based tangible interface and user interaction using the interface to easily acquire 3D geometry information using only user interaction that is movement after touch a pen to an object.

II. PROPOSED METHOD

A. *Tangible interface for 3D reconstruction*

In order to intuitively model objects as directly touching object, pen interface is efficiency because the interfaces are mostly used at drawing and sculpturing and used in daily life. Figure 2 shows pen interfaces proposed as authoring tool of natural multi-marker in this paper.

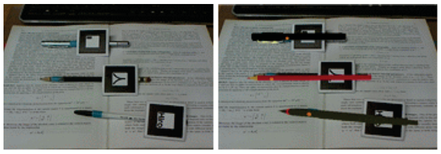


Figure 1. Interfaces proposed as authoring tools. (a) Authoring tools using marker-based pen. (b) Augmented virtual pen.

Figure 1(a) shows original appearance of tangible interface used in this paper, figure 1(b) shows augmented virtual pen on real world after recognizing marker. In order to verify a point in time of grip and release on pen, color band is stuck, as shown in figure 1(a). Also, interaction of brush is available as marker is stuck on center of pen. In this case, start and end point of tracking pen-tip is determined after checking whether the marker is hidden by finger or not. Marker is stuck on the tip of pen to calculate 3D information of the position of color band and pen-tip using transformation matrix between marker and camera. Equation (1) shows transformation relationship between center of marker and camera.

$$M = [R|t]C \tag{1}$$

where, M is 3D position vector of marker, and R is rotation matrix between camera and marker. t is 3D distance between camera and marker, C is position vector of camera. 3D information of pen-tip is calculated by matrix marker and pen-tip 3D position of marker calculated using equation (1). Equation (2) shows relationship between camera and pen-tip.

$$P = [I|D]MC = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -d \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} [R|t]C \tag{2}$$

where, I is an identity matrix, D is a translation vector that express distance, d, between marker and pen-tip. A pen-tip and a marker has not rotation relationship, only have y -axis translation. Therefore, rotation matrix is expressed identity matrix, distance transformation matrix is set 0, except y -axis translation. The 3D position of color band is also calculated using equation (2) modified d to distance value between marker and center of color band.

In order to feel pen as real tool used in modeling when an object is modeled, we propose a method so that user can move pen after gripping it. The tool has to possible to separate meaning and meaningless movement. For this, Color in area of color band is checked whether it is color band or finger color.

B. *Definition of user interaction*

If degree of freedom of hands motion increases, working ability of modeling also increases. Therefore, hand motions as various as possible must be used to reconstruct objects as fast as possible. And the hand motions must be daily motion used to model objects in real world. Since the proposed interfaces can acquire not only information of 3D translation, but also rotation, even if pose of pen is changed all motion information of pen can be acquired. Therefore, if the proposed interface is used as interface for modeling, degree of freedom of hand in state of gripping pen is allowed.

In order to use the proposed pens at modeling, user touch surface of object using pen-tip and then occlude color band. After the actions, user freely move surface of object for modeling. The position of pen-tip is continuously tracked using marker after covering color band by a finger. And then, acquiring 2D and 3D information of pen-tip position on surface of object are started. We use user interactions that are similar to human actions for understanding 3D geometrical information, watching and touch object in various viewpoints.

Control pens are additionally used to fast reconstruct various appearances of objects such as a curved surface for diversity of extracting 3D information. As movement patterns of 3D extraction pen becomes variety by control pens, complex object are reconstructed easily and fast. In this paper, we propose two control pens called rectangular control and brush control pen, respectively. The rectangular control pen determines area of rectangle consisted of the pen and a tracker pen. The rectangle area is determined by respectively moving two pens on opposite side edge of an object. The brush control pen determines line consisted of the pen and a tracker pen. The control pen and a tracker pen in any points are synchronously translated, and 3D line consisted of the two pen-tips do brush function. Finally, the rectangular control pen is able to fast reconstruct plane and the brush control pen is able to fast reconstruct curved

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

surfaces.

Figure 2 shows modeling process of a telephone with brush control pen. As shown first row and second row of figure 2, we divided to a body and a receiver of telephone for modeling. In receiver, since the appearance is a curved surface, a brush control pen is more efficient than rectangular control pen.

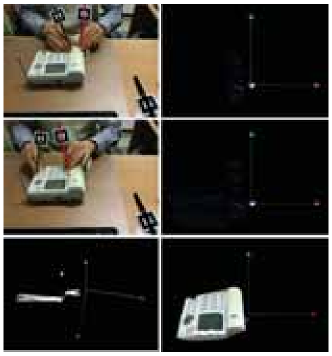


Figure 2. 3D modeling process with the brush control pen. The results of modeling the receiver of a telephone (first row), modeling the body of the telephone (second row), and the results of modeling with the brush control pen (third row).

III. Experiments and Application

A. *Accuracy of 3D modeling*

To verify accuracy of modeling, several distances between any points of reconstructed model is compared with ground truth.

Figure 3(a) shows real object, and circles being measured is superimposed. Figure 3(b) shows appearance of reconstructed model at point of view of real camera and figure 3(c) shows side appearance of result model and could verify what the result is similar to the real object.

Table 1 is the comparison result of line length of reconstructed model and real object. The lines that are consisted of circles of same color and the lines of real object are measured by ruler. In table 1, Errors of lines consisted of blue and red color are error are 0mm and 0.07mm, respectively. Since these values are in range of measurement error, it can induce that the model is exactly reconstructed. However, a line consisted of orange color have 21.83mm error. It is occurred by error of marker detection. As a distance between a marker and camera become larger, error of marker detection is increased.

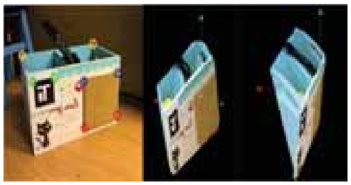


Figure 3. Verification of modeling accuracy. Target points for distance measurement(left). A virtual object at the current camera viewpoint and virtual viewpoint(middle and right).

Table 1. Accuracy of 3D modeling.

	Real distance (mm)	Measured distance (mm)
Blue circle	73	73.00
Red circle	73	73.07
Green Circle	34	38.14
Orange circle	170	148.17

B. *Application of augmented reality*

Figure 4 show a virtual object augmented on a model 3D reconstructed. It shows possibility to apply the model 3D reconstructed as multi marker to augmented reality using SURF descriptors[4] extracted on textures of an object and geometrical information. Although the result shown on figure 4 is similar to the previous marker-less augmented reality using only SURF, registration process of natural marker in previous methods is manually accomplished as separating interest object from background and storing SURFs to database. Since the result shown on figure 4 is yielded using SURF extracted by the proposed method, it is produced more quickly and intuitively than the previous method.

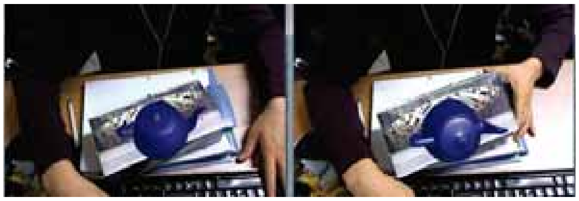


Figure 4. Example of augmented reality application.

IV. Conclusion

In this paper, we proposed two methods that allow users to 3D reconstruct by marker or user interaction as only watching object with a camera. Since the methods uses similar behaviors that are acted for understanding geometrical structure of objects by human, its shorten learning time for modeling and increase operation efficiency. The proposed 3D reconstruction method is acquiring not only geometrical information, but also recognizable feature points of object by user interaction. Since user directly extract 3D information of objects, various kinds of object structure more than previous methods that used sensor or analysis of multiple views can be reconstructed with efficiency from time and system complexity point of view.

V. Acknowledgment

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2011.

Reference

[1] Luca Vacchetti, Vincent Lepetit, and Pascal Fua, “Combining edge and texture information for real-time accurate 3D camera tracking,” Proceedings of the Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 48-57, Nov., 2004.

[2] Harald Wuest, Florent Vial, and Didier Stricker, “Adaptive Line Tracking with Multiple Hypotheses for Augmented Reality,“ Proceedings of the Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 62-69, Oct., 2005.

[3] Mark Pupilli and Andrew Calway, “Real-time Camera Tracking Using a Particle Filter,” Proceedings of the British Machine Vision Conference, pp. 519-528, Sept., 2005.

[4] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "SURF: Speeded Up Robust Features," Proceedings of the ninth European Conference on Computer Vision, pp. 404-417, May, 2006.

Implementation of Stable Video Conference System

Moongoo Lee, *Member*, IEEE

Dept. of Internet Information, Kimpo College, SEOUL, KOREA

Abstract — *In this paper, we propose an algorithm of remote video conference using silence detection algorithm to resolve the questions such as buffering method of video data in server and heavy traffic detection algorithm to the increase in participants. We apply a voice transmission algorithm and a channel management algorithm to the remote video conference system. A result of implementation we get that in consideration of average 20 frames and 30ms regardless of a number of participants, we can safely conclude that the transmission of video and voice data is stable.*

Key words — Silence detection, heavy traffic detection

I. Introduction

In previous video conference system, when the number of participants in video conference increases by n, the bandwidth and memory of n^2 is required. And also, it brings about increase in traffic and problem of a say during a conference in aspect of transmission of voice data.

In this paper, we present stable video conference system which is implemented by using a video data buffering and a technique of silence detection.

The video data buffering algorithm is not a method of broadcasting to other client in the server, but this algorithm uses two other methods; the buffering method of receiving compressed video data from clients and the indexing method for acquiring the video data of other participants in clients according to clients' bandwidth and network transmission speed.

And we apply a voice transmission algorithm and a channel management algorithm to the remote video conference system. The method used in the voice transmission algorithm is a silence detection algorithm which does not send silent participants' voice data to the server. The channel management algorithm is a method allocating a say to the participants who have priority. Finally, we can implement stable video conference system. So, in consideration of average 20 frames and 30ms regardless of a number of participants, we can safely execute that the transmission of video and voice data is stable.

II. VIDEO CONFERENCE SYSTEM

This section covers the details regarding preparation of your manuscript for submission, the submission procedure, review process and copyright information.

A. Video Data Transmission Algorithm

This paper adopted algorithm that server is transmitted video data from a client, and other participants from clients get video data according to its own bandwidth and network transmission speed. So, suggested algorithm is in contrast with the system of previous video conference that pushes video data to server and broadcast server to other clients [1]. The algorithm discussed in this paper has own video data queue according to the participants [2]-[3]. Therefore, because memory increases in order the problem of memory is solved as n^2 are modified to the form of n. Instead of having separate buffer according to the participants, we will have index from the video data that is taken lastly from our queue. If buffer queue is 10, and lastly taken index of video data is five, and video data is required, video data of index 6 is returned. This is the concept means that if next index of video data does not exist, server gets video data of relevant participant.

TABLE I
Video Data Transmission Algorithm

1. Schedule for receipt video data from client
2. Check of all attendant at conference
3. If video request is setting, get update data from user
4. If I-Frame is video data, remove video data at existing buffer
5. Add video queue

Table II
Video Data Buffering Algorithm

1.Implementation of server about video data reception of client
2. Get current session conference information
3. Get current session user information
4. Get video data that is updated from the users that I-frame chose
5. If video data has updated data, add to renewed video data list.
6. If not, request a video.
7. If the size of video queue is not reached to I-Frame, request P-Frame.
8. If the size of video queue is reached to I-Frame, request I-Frame.

B. Voice Data Transmission Algorithm

The Voice transmission algorithm sets up silence detection method and channel management method which makes unlimited assignment of says possible, says are composed of basic default of 4 units, and controllable [4].

Silence detection method is that voice data which is judged that participants do not say is not transmitted to the server although there are lots of participants. Because of this huge amount of traffic is reduced.

Table III
Voice Data Transmission Algorithm

1. Implementation of server related to the client's audio data transfer
2. Get current information of session
3. Get current information of users
4. Try to occupy audio channel
5. If you succeed to occupy audio channel, broadcast audio data to session participants

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

C. Channel Management Algorithm

The Channel management algorithm is the method that lots of people say at the same time, but a person who has priority is assigned a channel, and voice of people who are not assigned is ignored [5].

TABLE IV
CHANNEL MANAGEMENT ALGORITHM

1. Channel management function (used in audio transfer algorithm module)
2. Find the channel that you can occupy
3. You already have the channel, Update the last owned time
4. Find a difference between current time and last owned time
5.Find a participant who has the biggest difference
6. If current number of channel is smaller than maximum number, add
7. Add channel, and return added channel index

III. IMPLEMENTED OF VIDEO CONFERENCE SYSTEM

A. Configuration of Video Conference System

The web based video conference system is for client module and VCS web is a module to provide web based service, and its file name is *.ocx, and VCS conf is client implement file, and its extension is *.exe.

The VCS agent is also implement file *.exe, and it is interlocked with authentication process and plug and server, and it is designed that extension is possible. VCS server main is plug and client which is designed with automatic detachable method, and load balancing function is built in it instead of clustering technique.

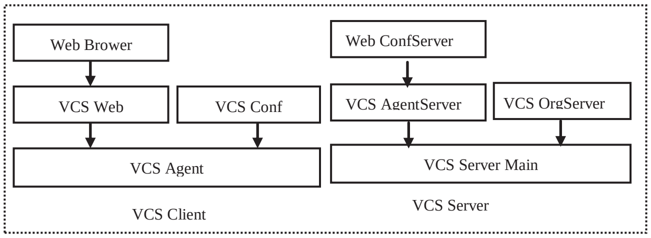


Fig. 1. Configuration of Video Conference System

B. Establishment matters

This paragraph describes circumstance that is set up for a video conference system implementation.

Default Audio Format
Type: DWORD, min: 0, max: 1, default: 1
When we progress video conference, we can set up the voice format. If we set up voice format to 0, the amount of data is reduced to 1/2, but the quality of voice data is dropped. Therefore, we set up default 1.
0 = 16bit 8 kHz (6KB/Sec),
1 = 16bit 16 kHz (12KB/Sec)

Default Audio Buffer Size
Type: DWORD, min: 1, max: 16, default: 4, unit: 20ms

A side to receive sets up the size of buffer to replay audio. Whenever it increases to 1, voice delay occurs additionally to 120ms. As the number gets small, audio delay gets reduced. Therefore, default voice delay time, which is the time that a sender's voice reaches a side to receive after passes the server, is calculated to 120ms * 4 = 0.48 second.

Default Video Quality
Type: DWORD, min: 1, max: 31, default: 4

The rate to compress video data MPEG is designated. As the number gets higher, it means high rate of compression. As the rate of compression becomes high, the size of compressed data is reduced and the amount of transfer is reduced. The video quality, however, is dropped more than a picture compressed.

Default Video Frame Rate
Type: DWORD, min: 33, max: 10000, default: 66, unit: ms

The rate of video frame per second is designated. Default 66 is a frame of (1000ms / 66ms = 15 ms). At least 33 is frame of (1000ms / 33ms = 30). Maximum 10000 is the frame of (1,000ms / 10,000ms = 0.1), so this indicates that 10 frame per second. As frame increases, PC load for video processing, and network bandwidth are required, so appropriate maximum number of frame is designated according to the network and requests of customers.

Default Video I-Frame Period
Type: DWORD, min: 1, max: 30, default: 10

Designate the video data's interval between I frame and P frame. Because the size of I frame is smaller than P Frame, we can reduce the amount of data transfer as reducing the number of occurrence for I frame.

Stream Content Length
Type: DWORD, min: 128, max: 16384, default: 8192

When we transfer the data, IDLE data (e.g., video or writing data except voice) designate current transfer size. As this amount gets bigger, it gets faster because we can transfer a lot of data at once. However, this makes voice data transfer interval larger, and it brings out problem of voice data quality. According to the status of voice transfer whether it has delay and pause or not, this amount is changed automatically within the range of maximum and minimum. This gets smaller as voice quality becomes poor, and if the quality of voice is good for a certain period, it gets bigger.

Stream Min Content Length
Type: DWORD, min: 128, max: 16384, default: 128

When we transfer the data, dedicate IDLE data minimum transfer size. Stream Content Length cannot be smaller than this size.

Stream Max Content Length
Type: DWORD, min: 128, max: 16384, default: 16384

When we transfer the data, dedicate IDLE data maximum transfer size. Stream Content Length cannot be bigger than this size.

Write Interval Time
Type: DWORD, min: 0, max: 100, default: 10, unit: ms

When we transfer the data, dedicate IDLE data transfer interval. If this size is small, we cannot transfer IDLE data a lot, so the number of opportunities to transfer the voice data is reduced, and this affects voice data quality. Maximum amount of IDLE data transfer per a second is Stream Content Length multiple 1000ms per interval time.

Stream Content Accelerate Rate:
Type: STRING, min: 1.0, max: 5.0, default: 1.5, unit: ms

When we transfer the data, dedicate a rate of acceleration of IDLE data. This is transferring the amount of IDLE data for a certain period of time that is continued. This method is that does not affect the quality of voice and expand the amount of IDLE data transfer.

Stream Accelerate Check Time
Type: DWORD, min: 0, max: unlimited, default: 360, unit: ms

When we transfer the data, dedicate a period continuing IDLE status to apply Stream Content Accelerate Rate. A designated time indicates real time data that is not transferred and IDLE status is continued. Therefore, default 360 increases the amount of IDLE transfer data to 1.50 times when we continue IDLE for 360ms.

C. The Result of Implementation

The result of implementing remote communication by using video transmission algorithm and voice transmission algorithm shows in Fig. 2 and Fig. 3. If the participants of the conference are four and six, all of the number of frame is normally 10 to 13, and average transmission speed is 142ms.

From the test result, as participants increase, problem of the amount of memory usage, problem of bandwidth, and network transmission speed problem are solved. However, traffic delay and average number of frame is below 15, and average transmission speed is 140ms.

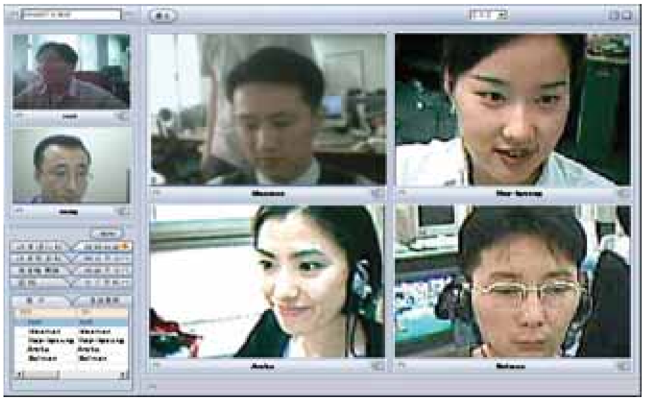


Fig. 2. In case the number of participants of video conference is four, and suggested sound or video transfer algorithm is applied, the number of frame is 10~14 and transfer speed is 140ms. But, only when the number of frame is below 15, transfer speed is kept to 140ms

Proposed video conference algorithm and voice transmission algorithm, memory, bandwidth, network transmission speed between clients and servers are solved, but for problem of resolution relevant to the number of frames and the improvement of transmission speed, there were modifications and supplements shows in Fig. 4.

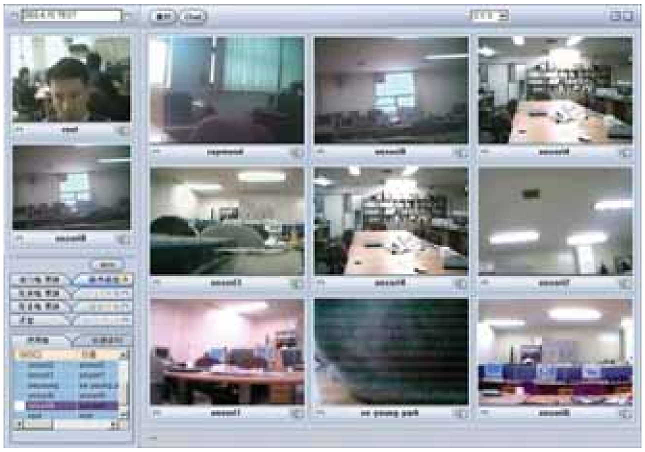


Fig. 3. In case the number of participants is increased to 6, constantly suggested voice or video transfer algorithm is applied, participants of video conference is 6, and the number of frame is 10~14, then the transfer speed is kept to 140ms. But the number of frame should be kept below 15.

The time of compression for the video data for video transmission is decided when the requirements are immediately received and background compression is prepared in advance instead of the time when clients transmit the server.

This made stable frame transmission possible. To improve the frame transmission speed with remote participants, we improved the rate of compression of I-frame, and the result was average number of frame was above 20 and 30ms continuously, transmission speed improved 4 to 5 times, and it showed clear solution.

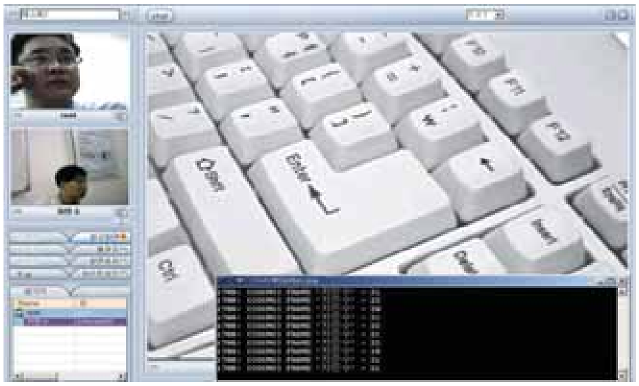


Fig.4. Not only video transfer algorithm but also channel management algorithm is applied together, and improve the rate of compression, it is transferred safely as maintaining average over than 20 frames and 30ms. As a result, the transfer speed is improved 4 to 5 times, and it showed high resolution.

Fig. 4 is the result explaining the relationship between transmission speed and the number of frames. From the graph, if the number of frame is below 5 and in case we transmit general video conference method, average transmission speed is 400- 500ms. In this case, the number of frame is below 5 in local, and in case of remote and if the number of participants increase, we will face many problems if we do not improve such as do not extend bandwidth, do not upgrade hardware or do not indemnify the performance of video conference camera as the best.

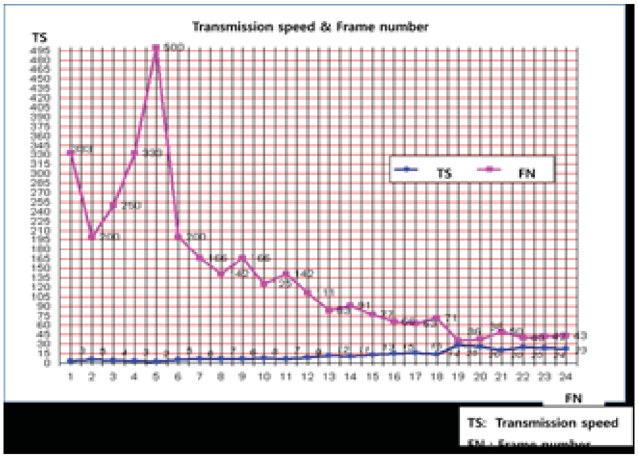


Fig. 5. Transmission speed and frame number

In case the number of frame is below 15 shown in Fig. 6, by adopting suggested algorithm index buffering method, solution problem of previous step of final modification, problem of memory according to the increase of participants, problem of bandwidth are solved, but the result was that transmission speed was considerably slow as 142ms.

In case of Fig. 5, implementation occurred by suggested algorithm, this case was that remote participants increased to 6, so there were no problem related to bandwidth and memory, but the average number of frame was below 15.

In case we use suggested algorithm, problem of bandwidth and memory were unrelated to the number of participants. In suggested algorithm, it does not compress when client transmit server, but it made background compression and transmit in advance which is shown in Fig. 4.

As enhancing the rate of I-frame, the speed was 30ms and stable transmission which is average 20 frames occurred.

IV. CONCLUSION

In this paper, to solve this problem stable video conference system with video transmission or voice transmission algorithm by using video buffering method and silence detection method.

The video buffering algorithm solved problem such as bandwidth extension, increase of memory usage as clients increase and continuous hardware upgrade. Voice transmission algorithm using silence detection method does not transmit voice data which is detected to not say among many participants. Channel management algorithm is that a participant who has a priority is assigned the say. Remote video conference system using suggested algorithm in limited memory, bandwidth and remote network transmission speed make stable transmission which is above 20 frames and average 30ms regardless of the increase of the number of clients. However, I will continue the research about a method to reduce delay time in silence detection algorithm detecting tactic participants.

REFERENCE

[1] Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman “The Design and Analysis of Computer Algorithms (Addison-Wesley Series in Computer Science and Information Processing)“Addison Wesley, 01 January, 1974.

[2] KeithJack, "Video Demystified" LLH Technology publishing, 2001.

[3] Jesus Pinto, Kenneth J. Christensen, University of South Florida, "An Algorithm for Playout of Packet Voice Based on Adaptive Adjustment of Talkspurt Silence Periods," 24th Conference Computer Networks, Lowell, Massachusetts, IEEE Computer Society, p.224, October 17 - 20, 1999.

[4] G.Y. Hong, A.C.M. Fong, Massey University, B. Fong, Lucent Technologies in Singapore, "QoS Control for Internet Delivery of Video Data," IEEE Computer Society p.458, April 08 - 10, 2002.

[5] Y. Shibata, N. Seta, S. Shimizu, Dept. of Comput. Sci., Toyo Univ., Kawagoe, Japan "Media synchronization protocols for packet audio-video system on multimedia information networks," IEEE Computer Society HICSS'95, Hawaii, USA, p.594, January 04 - 07, 1995.

BIOGRAPHIES

Moongoo Lee became a Member (M) of IEEE and IEEK in 2002. I received B.S. from Dept of Computer Science, Soongsil University in 1984, M.S. from Ewha Woman’s University in 1993, and Ph.D. from Dept of Computer Science, Soongsil University in 2000. I am now a full professor in Department of Internet Information of Kimpo College. SEOUL, KOREA. My research interests include Information Security, Algorithm Design, and other fields of Computer Science.

[6] H. Sun, W. Kwok, and J. Zdepski, “Architectures for MPEG compressed bitstream scaling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 191-199, Apr. 1996.

[7] R. Nicole, “Title of paper with only first word capitalized,” accepted for publication in IEEE Trans. Broadcast Technology.

[8] C. J. Kaufman, Rocky Mountain Research Laboratories, Boulder, CO, personal communication, 1992.

[9] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” *IEEE Transl. J. Magn. Jpn.*, vol. 2, pp. 740-741, August 1987 [*Dig. 9th Annual Conf. Magn. Jpn.*, p. 301, 1982].

Eating Activity Recognition for Health and Wellness: A Case Study on Asian Eating Style

Hyun-Jun Kim and Young Sang Choi
Intelligent Computing Lab, Future IT Research Center
Samsung Advanced Institute of Technology, Samsung Electronics Co., Ltd.

***Abstract*—In this paper, we propose an unobtrusive a method of daily life monitoring with a triaxial accelerometer embedded in a wrist band. The method recognizes each operation of the activities such as picking up food with chopsticks or eating steamed rice with a spoon and classifies the activities. We believe this method will contribute to the field of innovative health-promoting consumer electronics, which enables consumers to easily and accurately manage their daily life, especially eating activities, and consequently encourage to maintain healthier lifestyle.**

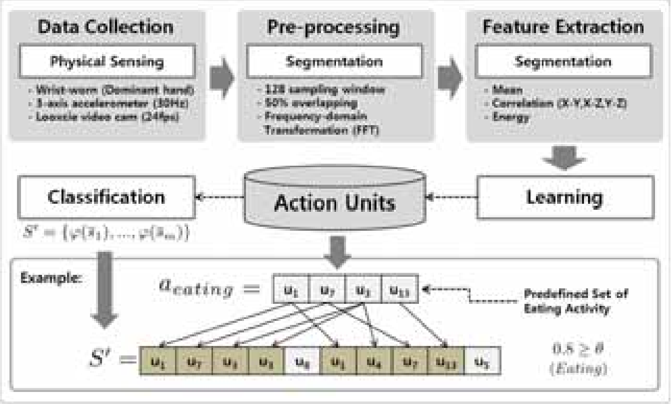

Data Collection
Physical Sensing
- Wrist-worn (Dominant hand)
- 3-axis accelerometer (30Hz)
- Looxcie video cam (24fps)
Pre-processing
Segmentation
- 128 sampling window
- 50% overlapping
- Frequency-domain Transformation (FFT)
Feature Extraction
Segmentation
- Mean
- Correlation (X-Y,X-Z,Y-Z)
- Energy
Classification
$S' = \{\varphi(\bar{s}_1), ..., \varphi(\bar{s}_m)\}$
Action Units
Learning
Example:
$a_{eating} =$
u_1 u_7 u_3 u_{13}
Predefined Set of Eating Activity
$S' =$
u_1 u_7 u_3 u_3 u_8 u_1 u_4 u_7 u_{13} u_5
$0.8 \geq \theta$
$(Eating)$

Fig. 1. Eating activity recognition process based on classified action units

I. Introduction

As people have more interest in their health, mobile healthcare is growing as an emerging technology with convergence of medicine and the information technology industry. Nowadays, chronic diseases such as type 2 diabetes or cardiovascular diseases are major threats to many people's health. Many of chronic diseases results from and aggravated by unhealthy eating lifestyle such as overeating, unbalanced nutrition, irregular meal patterns and indulgence in alcohol and caffeine. To prevent this, we need an effective method to manage our eating lifestyle by objective measurement. In spite of the importance of food intake in measuring lifesytle, there have been not much research on eating activity recognition. In a few reported studies, researchers tried to recognize a person's eating activities by using sensor network in a smart home environment[1], capturing eating activities with video camera[2] or obtrusive body worn sensors[3]. However, modifying environment or wearing sensors have limitation in practical use in daily life. In this research, we propose an unobtrusive method of eating activity recognition by using a 3-axis accelerometer embedded in the wrist band. This approach recognizes each operation of the activities such as 'picking up food with chopsticks' or 'eating steamed rice with a spoon' and classifies the activities with the start and end timestamps as shown in figure 1. With this approach, we can easily and accurately monitor our eating activity and other types of daily life activities such as smoking. We believe this technology can guide people to toward healthier lifestyle.

II. Eating Activity Recognition

The eating activity recognition procedures include preprocessing, feature extraction and recognition as shown in figure 1. Accelerometer data collected from a wrist band wearable device is firstly segmented and transformed into frequency domain data for the ease of computation. And then, some necessary features including mean, correlation and energy are extracted from the data. Finally, learning and classification are processed based on the feature data set. Notice that we used classification results as an input data for the recognition process, and we will discuss this topic in more detail in section II.B.

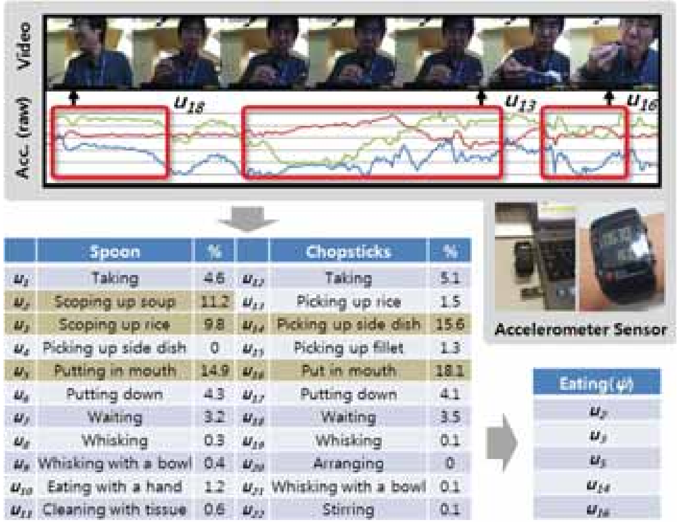
A. Action Units of Eating Activity

An eating activity can vary depending on the type of food, personal habits and cultural background. Therefore, we noticed sub activities, defined as, action units(u_i) are consisting the higher level of the eating activity. For example, we can find lots of different actions during meals such as eating steamed rice with a spoon, picking up side dishes with chopsticks, keeping the hand still during chewing and so on. These action units are commonly found on different eating activities and we classify them first and use them for the recognition process. In this study, we have defined the distinctive action units of the eating activity. From the observed data, when we assume that ψ as a predefined set of action units $\psi = \{u_1, ..., u_n\}$ and S as a set of feature data extracted from segmented data $S = \{\bar{s}_1, ..., \bar{s}_m\}$, where $\bar{s}_j = [f_1, ..., f_l]$. Eq. 1. shows Naive Bayesian method for the classification of the action units.

$$P(u_i \mid f_k) = \frac{P(f_k \mid u_i)P(u_i)}{P(f_k)} = P(f_k \mid u_i)P(u_i)$$

$$\varphi(\bar{s}_j) = \underset{u_i \in \psi}{\arg\max}\, P(u_i) \prod_{k=1}^{l} P(f_k \mid u_i) \qquad (1)$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



	Spoon	%		Chopsticks	%
u_1	Taking	4.6	u_{12}	Taking	5.1
u_2	Scoping up soup	11.2	u_{13}	Picking up rice	1.5
u_3	Scoping up rice	9.8	u_{14}	Picking up side dish	15.6
u_4	Picking up side dish	0	u_{15}	Picking up fillet	1.3
u_5	Putting in mouth	14.9	u_{16}	Put in mouth	18.1
u_6	Putting down	4.3	u_{17}	Putting down	4.1
u_7	Waiting	3.2	u_{18}	Waiting	3.5
u_8	Whisking	0.3	u_{19}	Whisking	0.1
u_9	Whisking with a bowl	0.4	u_{20}	Arranging	0
u_{10}	Eating with a hand	1.2	u_{21}	Whisking with a bowl	0.1
u_{11}	Cleaning with tissue	0.6	u_{22}	Stirring	0.1

Eating(ψ)
u_2
u_3
u_5
u_{14}
u_{16}

Fig. 2. An example of action units extracted from video sequence with corresponding accelerometer data and predefined set of eating activity.

The method φ selects the most likely class u_i with the highest probability, and finally we generate $S' = \{\varphi(\bar{s}_1), ..., \varphi(\bar{s}_m)\}$. Table 1 shows the predefined action units of the eating activity. Initially, we used 5 distinctive action units, but the number of units will be extended for the better accuracy of classification in further studies.

B. Eating Activity Recognition

From the set of action units S', the system can predict the user's activities according to the proportion of each action units, u_i. For instance, as shown in figure 2, we categorized 22 action units of eating activity($i = 22$) from the video analysis as shown in table 2. And we found that the top 5 units cover nearly 70% of overall units ($\theta \geq 0.7$). Therefore, we can recognize activities based on the probabilistic function $P(a_i|S')$ with minimum threshold θ that representing the proportion of the belonging action units, where $Activity = \{a_{eating}, a_{PCworks}, a_{walking}\}$. However, we didn't regard the sequential patterns of the action units because we couldn't discover enough evidences of patterns from the collected dataset in the initial study. When there are sequential relationships among the action units, Hidden Markov model or finite state machine will help improve the recognition accuracy.

III. Experiments

Accelerometer data was collected by wrist watch type sensor[1] at a sampling frequency of 30Hz, each window which is overlapped by 50%[4], represents 4 seconds. We collected data from 8 subjects during 2 weeks under naturalistic conditions. We extracted mean, correlation and energy features from the frequency domain transformed data set. Firstly, we tested the classification accuracy of eating and non-eating activities by using Weka[2] toolkit, and the results of f-measure are shown in table 1. For the eating activity, we collected 4 different meal types and for the non-eating activity, we collected PC work dataset during typing keyboard and using mouse and walking dataset. Table 2 shows the classification result of the type of meals including steamed rice, bread, noodle and cereal. Finally, we tested the classification accuracy when we apply different number of predefined action units as shown in table 3. Although we extracted 5 distinct action units in section II.A, we couldn't find significant classification results. When we abbreviated to 2 action units of 'doing something with a spoon' and 'doing something with chopsticks', the accuracy showed 0.776.

TABLE I
Classification of Eating and Non-Eating Activities

Type	NaiveBayes	BayesNet	Boosting	Bagging	C4.5
Eating	0.91	0.93	0.95	0.94	0.94
PC Works	0.76	0.78	0.85	0.80	0.82
Walking	0.93	0.95	0.98	0.97	0.98
W. Avg.	0.87	0.90	**0.93**	0.91	0.92

TABLE II
Classification of Eating Type

Type	NaiveBayes	BayesNet	Boosting	Bagging	C4.5
Rice	0.86	0.85	0.89	0.87	0.88
Bread	0.55	0.55	0.71	0.64	0.66
Noodle	0.41	0.41	0.61	0.55	0.59
Serial	0.37	0.36	0.37	0.08	0.38
W. Avg.	0.70	0.69	**0.78**	0.74	0.77

TABLE III
Classification of Action Units on Eating Activities

Number of Action Units($\lvert\psi\rvert$)	5	4	3	2
Accuracy	0.505	0.615	0.635	**0.776**

IV. Conclusion and Future Works

We proposed a method of recognizing eating activities by using wrist band sensor worn only on the dominant hand. In the initial study, we could accurately recognize an eating activity from non-eating activities and we also showed the feasibility of classification of meal type. For the meal type classification, we experimented with limited distinctive action units (5 and 2). The biggest challenge of the research was collecting datasets containing all types of action units. In the future work, we will collect fluent labeled datasets for training by video-based annotation. By collecting more data, we expect to improve action unit classification and consequently achieve better performance in eating activity recognition.

References

[1] Kaiser, M., Arsic, D., Hornler, B., Hofmann, M. and Rigoll, G.,"The Noldus database: Automated Recognition of Restaurant related Activities for the Restaurant of the Future", Proc. of the WIAMIS, April 2011.

[2] Tolstikov, A., Biswas, J., Thm, C. K. and Yap, P.,"Eating Activity Primitives Detection-a Step Towards ADL Recognition",Proc. of the HealthCom, July 2008, pp.35-41.

[3] Amft, O. and Troster, G.,"On-Body Sensing Solutions for Automatic dietary Monitoring",in IEEE Transaction on Pervasive Computing, April 2009, pp.62-70.

[4] Bao, L. and Intille, S. S.,"Activity Recognition from User-Annotated Acceleration Data",Proc. of the PERVASIVE, April 2004, pp.1-17.

[1] http://processors.wiki.ti.com/index.php/EZ430-Chronos

[2] http://www.cs.waikato.ac.nz/ml/weka/

Remote Actuator Control Method by Visual Feedback Using ROI with Communication Property of Multiple Channels

Yu Kudo, Akihiro Tsutsui, Ikuo Yoda
NTT Network Innovation Laboratories, NTT Corporation
M1-2F 3-9-11 Midori-cho, Musashino-Shi, Tokyo, 180-8585 Japan

***Abstract*— Many types of sensors and actuators are being connected to networks, and highly functioning or highly efficient services that use them are being investigated. One useful application of such systems might be remotely controlling actuators by using visual feedback. This method would be used to capture images of actuators and transmit those images to high-performance computers via networks for fast image analysis for obtaining information related to the actuators' state, e.g., velocity and position. This method allows the separating of sensors, computing resources, and actuators. Thus, actuators can be easily fabricated and at a low cost. However, delay and loss of data in networks are possible, which negatively affect performance. Therefore, we propose a communication method that transmits only control data and small images sufficient for controlling actuators via stable networks for less cost and more flexible system design.**

I. INTRODUCTION

Sensors, micro-controllers, and network I/F devices, such as Ethernet modules, are improving in performance, lowering in cost, and becoming smaller. Now, many types of sensor and actuator devices can be connected to networks. This allows us to collect sensor data that is useful for estimating environmental information in the home, office, or factory. Moreover, we can control actuators and mobile robots by using such sensor data via the Internet from a remote site.

Based on this background, sensor networks and network robot systems have been investigated [1], [2]. A sensor network was proposed in which many types of sensors are connected to networks. These sensors' data are collected using networked storage, and information we need is estimated or obtained by analyzing them. On the other hand, a network robot system is a coordinated system of actuators (robots), sensors, and networks. An application of a network robot system is a field trial of mobility for the elderly and disabled supported by robots in a shopping mall (public place) [3]. This system uses multiple robots, sensors installed in the robots and placed throughout the environment, and a wireless network.

Therefore, a method for coordinating actuators, sensors, and networks will play an important role in developing new services and enabling existing applications to be higher functioning or more efficient.

II. TARGET MODEL

A. Outline of Target Model

An actuator, sensors, and network coordinated system has many types of models depending on the aims or application. Accordingly, we focused on the remote control of actuators by using visual feedback and a web camera. The outline of this model is shown in Fig. 1.

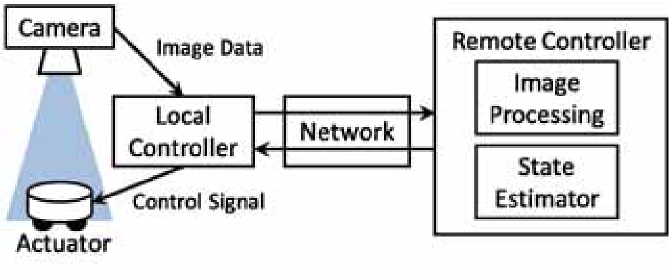


Fig. 1. Outline of target system model

B. Visual Feedback Using Web Camera

Many types of sensors, such as temperature and humidity sensors, acceleration sensing MEMS chips, and laser range finders, can be connected to networks. We determined that a web camera is the most important device for remotely controlling an actuator since it is a standard, low-cost, and easy to obtain device. It can capture images of actuators and everyday environments such as a room, home, or office.

Such an image capturing device allows us to use a control method called visual feedback control [4]. This method involves using a capturing device, such as a web camera, as a kind of sensor to detect the state, e.g., velocity or position, of actuators, and then controlling actuators by using the obtained state information as feedback signals. An example of a visual feedback control method involves capturing an image of a small mobile robot in a room by using a web camera, analyzing the captured data for the system to detect the robot in that image by using an object detection technique, estimating the position of the robot and the difference in location data between it and the goal point, and sending moving instructions to the robot. This method has the advantage of detecting objects (actuators, obstacles…etc.) in order to control or avoid them in an everyday environment.

C. Remote Visual Feedback Control System

When controlling and operating actuators from a remote site, observing the environment the actuators are in is necessary due to administration, security, and maintenance requirements. Therefore, image capturing devices are required for remote control systems. In addition, a web camera is used on the assumption that it is connected to a network for real-time communication. Therefore, we ought to be able to use image data not only for security or maintenance but also for controlling actuators by using visual feedback control for accurate (semi-) automated controlling from a remote site by using high-performance networked computer.

We define this system (e.g. Fig. 1) as remote visual feedback

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

control. In this system, control target objects (e.g., actuators or home appliances) and the environment around them is captured with a camera, and the captured image data are sent to a high-performance computer at a remote site via networks. The computer analyzes the data, estimates the physical state of the control target objects, and generates instruction signals on what the objects should do. The objects are actuated when the generated signals are sent to them via networks.

This system has the following advantages and disadvantages. Physically setting up capturing devices for visual feedback control is easy. It does not require the decomposition of control target objects for attaching sensors to detect the objects' physical values such as position, velocity, or acceleration. Thus, we can easily construct objects. Although visual feedback control requires an exceedingly high-speed computing resource for (semi-) real-time image processing and analyzing, the system can detach the resources from inside the objects and the environment. This leads to objects that are low in cost and consume less energy.

On the other hand, the system has disadvantages in terms of networks. Because of real-time transmission of images, network loads are generally large, and long network latency and data loss are likely to occur if the capacity of a network is insufficient or unstable. Latency and loss cause noise in an image, processing latency in the remote controller, and incorrect estimation or failure of control signal generation. This may give rise to insecure, unexpected actuator motion. This is detailed in Fig. 2.

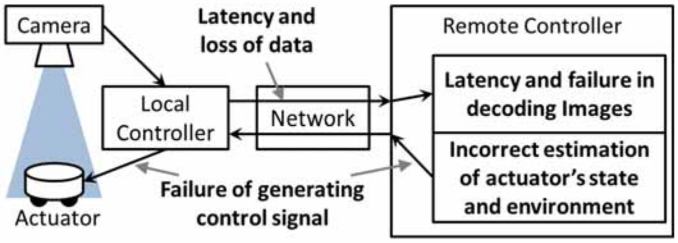

Camera
Latency and loss of data
Remote Controller
Local Controller
Network
Latency and failure in decoding Images
Incorrect estimation of actuator's state and environment
Failure of generating control signal
Actuator

Fig. 2. Disadvantages with networks

As described above, to correctly actuate a remote visual feedback system requires robustness against latency and data loss or a method for preventing these problems.

III. Proposed Method

A. Existing methods

The network latency described in Fig. 2 brings about sensing latency; therefore, a remote control system using a type of time-lag system according to control theory is currently being investigated. Similarly, data loss can be regarded as sensor noise, so the system is also equipped with a noise-full sensor. Therefore, control-theory-based methods for solving these problems have been investigated. Fujita et al. proposed a method with which input gain to a system is varied depending on network latency and showed that the method has a stabilization effect for a varying time-lag tele-operation systems [5]. Takahashi et al. used network latency and a system's physical model to forecast actuation of control targets then stabilized the system by using that forecasting state [6]. The sequence control approach has also been applied. To prevent the effect of data loss in a network with this approach, many motion scenarios are defined with intention and control messages in advance, and control targets are manipulated by starting and stopping the scenarios, e.g., Yagi et al. integrated a robot system operated using mobile phones [7].

As described above, there are many control-theory-based approaches; however, few network-based methods have been investigated for remote control using visual feedback. Almost all these control-theory-based approaches are effective on the assumption that the network parameters remain within the range defined with these approaches. However, this range is not guaranteed in an everyday environment. Thus, we propose a network-based method for the visual feedback control system.

B. ROI with low rate stable network

Networks are roughly divided into two types: low-cost, high-capacity transmission and best-effort, such as the Internet, and high-cost and stable such as a private network for business. The meaning of "stable" is that the maximum transmission latency or the maximum data-loss rate is guaranteed by the network supplier or specifications of communication devices. We propose using dual network channels, one is a best-effort network, and the other is a low bit-rate stable network. The best-effort network is used to transmit original captured images. The stable network transmits the control signal and region of interest (ROI) images sufficient for visual feedback. This is detailed in Fig. 3.

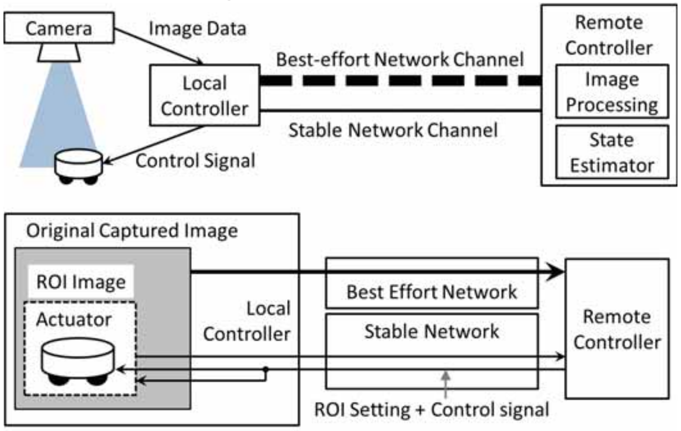


Fig. 3. Outline of proposed method

Normally, low-latency and lossless image transmission requires high–capacity networks, and the cost of high-capacity stable networks is higher than that of best-effort networks. In addition, the cost of stable networks is relative to transmission capacity. In the visual feedback system, however, all captured images do not need to be transmitted at a high frame rate. To generate a control signal, state estimation requires a high ROI frame rate, which is a small, specific region of the image that includes control targets. Therefore, the ROI is smaller than the original captured image and the required transmission capacity of the network decreases; therefore, the remote visual feedback system can use an inexpensive stable network, which exhibits a low bit-rate, low-latency, and low data-loss rate. In other words, the system can inexpensively prevent negative effects, such as transmission time lag and data loss, in networks.

Several stable (or priority) networks for consumers have been specified, e.g., ITU-T Recommendation Y.2001 defines a next generation network (NGN), and network services based on it have already started in Japan [8].

C. Process sequence of proposed method

Figure 4 describes the process sequence of the proposed method.

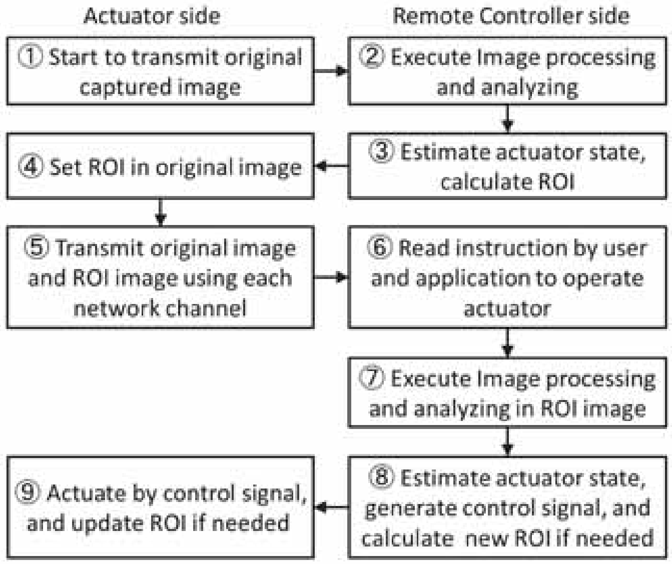

Actuator side
Remote Controller side
① Start to transmit original captured image
② Execute Image processing and analyzing
③ Estimate actuator state, calculate ROI
④ Set ROI in original image
⑤ Transmit original image and ROI image using each network channel
⑥ Read instruction by user and application to operate actuator
⑦ Execute Image processing and analyzing in ROI image
⑧ Estimate actuator state, generate control signal, and calculate new ROI if needed
⑨ Actuate by control signal, and update ROI if needed

Fig. 4. Process sequence of proposed method

First, original captured images of an actuator are sent to a remote controller with application programs or operators via a best-effort network. The images are used to search where the actuator is. Then, the remote controller estimates the actuator size and position in the images and sends sufficient ROI configuration information to keep the actuator inside its range. Next, the actuator-side local controller sets the ROI and starts transmitting ROI images via a stable network while continuing original image transmission. The remote controller then generates and sends control signals using ROI images and users (or applications) input instructions. Simultaneously, if the actuator's position is likely to go outside the ROI, a new ROI configuration is calculated. Finally, the local controller actuates the actuator and updates the ROI if needed. Processes 5 to 9 in Fig. 4 are executed repeatedly.

IV. SIMULATION MODEL

A. Implementation Example

We investigated a simulation model of our proposed method to clarify the characteristics of a remote visual feedback system, an example of which is shown in Fig. 5.

In this example, a marker and image detection program are used. The marker is placed on top of an actuator (cleaning robot). The program uses pre-built machine learning data to detect the marker. That is, personal computer 2 (PC2) recognizes the actuator's position by detecting the center of the marker in the captured image.

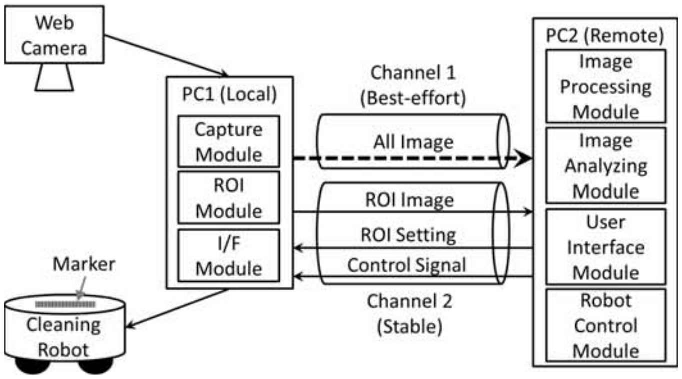


Fig. 5. Implementation example

To control the cleaning robot as intended, the position data are used for visual feedback in PC2. Additionally, the user stays in front of PC2. Thus, he/she manages and operates the cleaning robot using all images, and PC2 supports the operation using visual feedback based on the ROI images.

B. Simulation model

Our proposed method has a trade-off concerning ROI size and transmission capacity. The ROI size is smaller, and the required transmission capacity is lower. Otherwise, the ROI size is smaller, and the marker, which indicates the actuator position, easily goes outside the ROI. As a result, PC2 cannot detect it or control the actuator. Therefore, from the viewpoint of application, it is important to determine the relation between the maximum actuator speed V_{MAX} [pixel/s] and ROI size ℓ^2 [pixel^2] (regular rectangle). In addition, V_{MAX} means that PC2 fails in generating visual feedback signals when the actuator runs faster than this value, and "[pixel/s]" is the unit of the actuator's speed in the image. Of course, the actuator has physical speed; however, we can calculate it using V_{MAX}, angle of view, and position relative to the environment. Thus, we use the unit because of generality.

If ROI frame interval $1/f$ [s] (f: ROI frame rate [fps]) is longer than ROI image transmission time $T\ell^2$ [s] and processing time $P\ell^2$ [s], then V_{MAX} is assumed to be determined as the following expression.

$$V_{MAX} = \frac{\ell - M}{2} \cdot f \tag{1}$$

M: length of one side of the marker [pixel]

Equation (1) gives the length that the actuator can move per ROI frame interval $1/f$. Hence, V_{MAX} should increase in relation to ℓ.

Otherwise, if $T\ell^2$ and $P\ell^2$ are longer than $1/f$, then V_{MAX} is assumed to be determined as the following expression.

$$V_{MAX} = \frac{\ell - M}{2} \cdot \frac{1}{T\ell^2 + P\ell^2} \tag{2}$$

Equation (2) gives the length that the actuator can move per ROI update interval. Because we presuppose using stable and low transmission capacity networks with our proposed method, transmission latency is negligibly minimal; otherwise the transmission time needed to send ROI images cannot be ignored by comparing it with the frame interval under large ROI. Also, the ROI image processing time (e.g. time needed to detect the marker) is not negligible. In addition, transmission time and processing time are quadratic functions of ℓ, because an image consists of two-dimensional data.

Consequently, the ROI should be updated when the remote PC finishes estimating the actuator position. Furthermore, to keep the marker inside ROI, $T\ell^2$ and $P\ell^2$ are more important factors than *f*.

Based on the above equations, V_{MAX} is assumed to have a characteristic expressed as follows.

$$V_{MAX} = \begin{cases} 0 & (\ell < M) \\ \dfrac{\ell - M}{2} f & \left(\dfrac{1}{f} > T\ell^2 + P\ell^2\right) \\ \dfrac{\ell - M}{2} \cdot \dfrac{1}{T\ell^2 + P\ell^2} & \left(\dfrac{1}{f} \leq T\ell^2 + P\ell^2\right) \end{cases} \quad (3)$$

V. Simulation and Consideration

A. Simulation

Based on Eq. (3), we show two simulation results for V_{MAX} in Figs. 6 and 7.

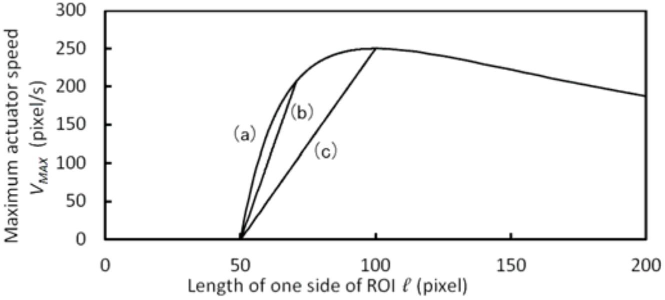


Fig. 6. Simulation results under each parameter *f*
(a) $M = 50,\ f = 40,\ T + P = 1.0 \times 10^{-5}$
(b) $M = 50,\ f = 20,\ T + P = 1.0 \times 10^{-5}$
(c) $M = 50,\ f = 10,\ T + P = 1.0 \times 10^{-5}$

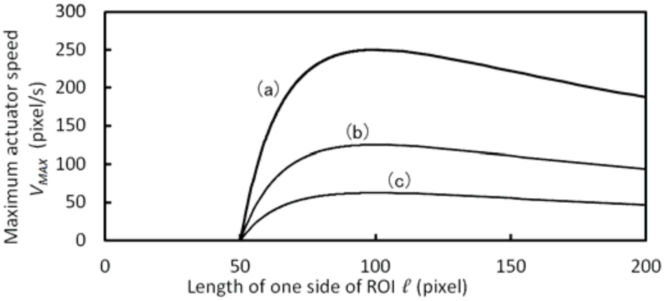


Fig.7. Simulation results under each parameter *T*+*P*
(a) $M = 50,\ f = 40,\ T + P = 1.0 \times 10^{-5}$
(b) $M = 50,\ f = 40,\ T + P = 2.0 \times 10^{-5}$
(c) $M = 50,\ f = 40,\ T + P = 4.0 \times 10^{-5}$

From these simulation results, V_{MAX} has the maximum value. In Fig. 6, discontinuous points exist at ℓ = 70 pixels and ℓ = 100 pixels under *f* = 20 fps and *f* = 10 fps, respectively. On the contrary, in Fig. 7, no discontinuous point exists. That is, when the ROI frame rate is sufficiently fast, the primary factors that restrict V_{MAX} are ROI image transmission time and processing time.

B. Consideration

From Figs. 6 and 7, V_{MAX} is considered to have the maximum value. In other words, making the ROI size excessively large would cause a reduction in V_{MAX}. Therefore, our proposed method has a range of applications subject to an available network's transmission capacity and calculating performance of computers and algorithms.

For example, in Fig. 5, the marker size is 50×50 pixel^2, the camera specifications are VGA and 20 fps, *T*+*P* is 1.0×10^{-5} s/pixel^2 (many factors affect this, however, it can be obtained by measuring the time difference between the starting time of sending ROI images and the ending time of estimating the actuator state), and V_{MAX} characteristics are assumed to be similar to those in Fig. 6 (b). In this case, our proposed method is effective in operating the actuator at speeds ranging from 0 to 240 pixel/s.

From Eq. (3) and these simulation results, to achieve a V_{MAX} value, the required ROI size can also be determined subject to *f*, *T*, and *P*. When the frame rate of the camera is low, its effect can be compensated by enlarging the ROI size.

VI. Conclusion

We discussed the importance of a sensor, actuator, and network coordinated system and defined the remote visual feedback system as one such system. To prevent negative effects of networks from controlling actuators, we used ROI and two communication channels; best-effort and stable. We conducted two simulations on maximum actuator velocity to investigate system characteristics. It became apparent that our proposed method has an effectual range with respect to ROI size.

In the future, we will build an experimentation environment and verify the validity of the simulation model.

References

[1] I. F. Akyildiz, "A survey on sensor networks," *Communications Magazine, IEEE*, vol. 40, Issue: 8, pp. 102-114, Aug. 2002.

[2] T. Akimoto, and N. Hagita, "Introduction to a Network Robot System," *Intelligent Signal Processing and Communications, 2006. ISPACS '06. International Symposium on*, pp. 91-94, 12-15. Dec. 2006.

[3] K. Kamei, S. Nishio, N. Hagita, and M. Sato, "Cloud networked robotics," Network, IEEE, vol. 26, Issue: 3, pp. 28-34, May-June. 2012.

[4] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura," Manipulator control with image-based visual servo," Robotics and Automation, 1991. Proceedings, 1991 IEEE International Conference on, vol. 3, pp. 2267-2271, 9-11 Apr. 1991.

[5] H. Fujita, and T. Namerikawa," Delay-independent stabilization for teleoperation with time varying delay," American Control Conference, 2009. ACC '09, pp. 5459-5464, 10-12 June. 2009.

[6] S. Takahashi, H. Nishimura, and Z. Weng," Visual Feedback Control with Compensation of Image Information Delay: Verification by Using Three-Link Space Robot," The Japan Society of Mechanical Engineers, vol. C-71, pp. 2225-2232, 25. Jury. 2005.

[7] A. Yagi, M. Sakai, K. Kashiwabara, K. Matsumiya, K. Masamune, and T. Dohi," Tele-care system by hoist-trolley robot arm from multi controllers and cellphone controller," IFMBE Proceedings, 2007, vol. 14, part. 17, pp. 2872-2875.

[8] ITU-T Rec. Y.2001, "General overview of NGN," Approved in 2004-12.

Two-Stage Charge Sensing Circuit for a Mutual-Capacitive Touch Screen Panel

Hoshin Cho, Sang-Jin Lee, Seok-Man Kim, Cha-Keon Cheong*, Kyoungrok Cho, *Member, IEEE*

College of Electrical and Computer Engineering, Chungbuk National University, Korea

**Dept. of System Control Eng., College of Eng., Hoseo University, Korea*

***Abstract*--In this paper, we introduce a new charge sensing circuit for a mutual-capacitive touch screen panel (TSP). The circuit senses charges in two stages: a sensing stage and an evaluation stage when the TSP is touched or untouched. This method establishes easier recognition of a touch event on the TSP and reduces the size of the on-chip sensing capacitors. The dynamic sensing range of this TSP is improved by 38% over that of a conventional TSP.**

I. INTRODUCTION

Capacitive-type touch screen panel (TSP) technologies have mostly been adopted in most mobile applications, because they provide a comfortable and intuitive user interface (UI). The projected-capacitive TSP technologies are classified into two types, self- and mutual-capacitive on the basis of their charging schemes and physical structures. Recently, a mutual-capacitive TSP, which allows multi-touch features to overcome the ghost point problem, has been in the focus [1], [2].

A mutual-capacitive TSP has a cross-bar architecture: the top part contains a current source, and the bottom part contains charge-sensing electronic nodes. The pattern on a cross-bar point forms a coupling capacitor (*Cc*), which is modulated by the touches of human fingers [1].

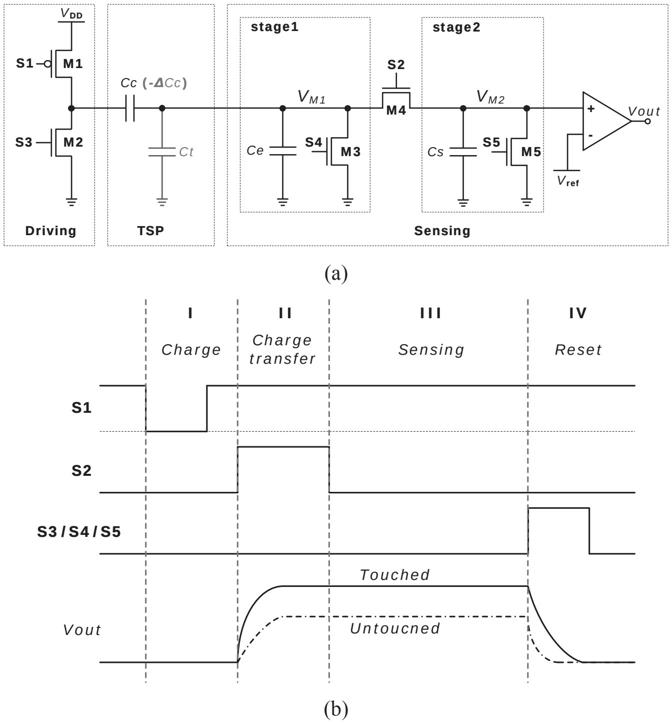
The mutual-capacitive TSP is well-adapted to a charge-to-voltage conversion scheme to capture charge modulation. It should be noted that an amount of charge on the cross-point can be converted to voltage. When a finger touches the TSP, the coupling capacitance below the finger decreases (*Cc*-*ΔCc*). The fringing electric field between the driving and the sensing line electrodes is absorbed (-*ΔCc*) by a finger that is grounded to the earth, which is an additional series touch capacitance (*ΔCc*≈*Ct*) [3], [4]. However, the mutual-capacitive TSP is more noise-sensitive than self-capacitive TSPs and requires greater chip resources to compensate for sensing errors caused by process variation [5].

In this paper, we introduce a new charge sensing circuit for a mutual-capacitive touch screen panel. The circuit enhances the ability to recognize a multi-touch event by increasing the sensing voltage variation when the TSP is touched and untouched.

II. TOUCH SENSING ARCHITECTURE

A. Proposed Sensing Circuit Architecture

The touch-sensing controller consists of a driving circuit, a TSP, two-stage sensing circuits, and a comparator, as illustrated in Fig. 1. The circuit is operated according to the timing diagram shown in Fig. 1(b).



(a)

(b)

Fig. 1. Proposed touch screen controller architecture on mutual-capacitive TSPs. (a) Circuit diagram. (b) Timing diagram, including four steps to recognize a touch position.

The driving circuit has a pMOS and an nMOS transistor for charging and resetting (discharge) of the coupling capacitance, *Cc*, on a TSP. Each TSP cell originally possesses a coupling capacitance *Cc* and a touch capacitance *Ct*, which are below the touched cell. The column-parallel sensing scheme has two stages of sensing circuits and a comparator based on a sample and hold circuit. Stage 1 comprises an external capacitor *Ce* and a reset transistor. Stage 2 comprises a sensing capacitor *Cs* and a reset transistor. In Fig. 1(a), transistor M4 separates the circuit of stage 2 from the noise component of the TSP when the capacitance is evaluated.

B. Touch Screen Control Algorithm

The algorithm of the touch screen controller shown in Fig. 1(b) completes charge sensing in four steps as follows: (i) charging, (ii) charge transferring, (iii) sensing, and (iv) reset. (i) Charging step: capacitors *Cc*, *Ct*, and *Ce* on node V_{M1} are charged when signal S1 is active. (ii) Charge transferring step: the charge on node V_{M1} is transferred to *Cs* on node V_{M2} when signal S2 is active. In this step, the left-hand side of *Cc* is

This work was sponsored by ETRI SW-SoC R&BD Center, Human Resource Development Project.

floated in order to block its effects on the sensing step. Thus, only *Ct* and *Ce* affect the sensing step. This provides greater robustness to noise from the TSP. (iii) Sensing step: the charges in *Cs* are evaluated by the comparator with a reference voltage V_{ref} that receives a touch position. (iv) Reset: capacitors *Cs* and *Ce* are discharged when S3, S4, and S5 are enabled. In this step, the next row of the TSP simultaneously proceeds to the charge step.

In the proposed scheme, node voltages V_{M1} and V_{M2} can be estimated by their relationships with the supply voltage V_{DD} and the capacitances in the network. It follows that

$$V_{M1} = V_{DD}(\frac{Cc}{Ce+Cc}),\; V_{M2} = V_{M1}(\frac{Ce}{Ce+Cs}) \quad (1)$$

$$V_{M1}^{'} = V_{DD}(\frac{Cc-\Delta Cc}{Ce+Cc}),\; V_{M2}^{'} = V_{M1}^{'}(\frac{Ce+Ct}{Ce+Ct+Cs}) \quad (2)$$

Note that V_{M*} is the voltage when the TSP is untouched, and V'_{M*} is the voltage when it is touched.

According to the research by Krah [4], *Ct* is significantly smaller than C_{C}. In this study, the effect of *C*t on V_{M1} and $V_{M2}^{'}$ can be ignored for the first stage, which is separated from the sensing stage by S2. However, charge transfer occurs when the TSP is touched. *Ct* establishes a larger sensing voltage variation when the TSP is touched and untouched, which are denoted by $V_{M2}^{'}$ and V_{M2}, respectively.

C. Comparison with Conventional Architectures

As mentioned in the previous section, *Cc* does not affect the charge transferred to *Cs* in the sensing step. The touch capacitance *Ct* is significantly smaller than *Cc*, (*C*c>>*Ct*). This allows capacitors *Ce* and *C*s to be designed smaller than those designed in conventional methods, which allows for a smaller chip area.

In a conventional mutual-capacitive touch screen controller, charging into *Ct* can be ignored because it does not affect the sensing operation [4]. In this study, however, capacitance *Ct* between a finger and the sensing line affects the amount of charge that is transferred to *Cs*. In the proposed scheme, the sensing voltage, when the TSP is touched and untouched, is improved by 38% as compared with [3]. Table I summarizes the comparison results obtained with conventional methods with those obtained with the method presented in this paper.

D. On a Chip for a TV Remote Controller

We introduce a touch-type TV remote control system using the proposed circuitry as an application. The block diagram is shown in Fig. 2, which consists of an LCD panel, sensing circuits. The remote control provides a visual interface and communication with a TV as a smart remote control. All of these IPs can be fabricated on a chip.

III. CONCLUSION

In this paper, we introduce a new charge sensing circuit for a mutual-capacitive touch screen panel (TSP). We modeled the TSP panel as a capacitor network, designed its hardware architecture, verified its operation, and compared the results with conventional touch screens. The proposed scheme improves the sensing voltage to allow for easier recognition of a touch event and reduces the sizes of the on-chip sensing capacitors. The sensing voltage is improved by 38% over that of a conventional circuit.

TABLE I
COMPARISON WITH CONVENTIONAL METHODS

Properties	[6]	[3]	This Study
Touch type	Self-capacitive	Mutual-capacitive	Mutual-capacitive
Sensing method	Charge transfer	Charge-to-voltage	Charge transfer
Multi-touch feature	No	Yes	Yes
Sensing parameter	*Ct*	Δ*Cc*	*Ct*, Δ*Cc*
Size of Cs	*Cs*>*Cc*	-	*Cs*<*Cc*
Immune to noise	Strong	Weak	Medium
Sensing voltage (mV)	-	121.6	167.9

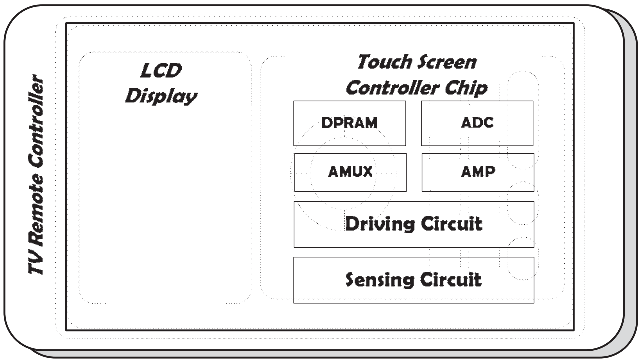


Fig. 2. Block diagram of a TV remote control system based on a microprocessor and the proposed touch screen controller circuit for a visual interface and communication with a TV.

REFERENCES

[1] G. Barrett and R. Omote, "Projected-capacitive touch technology," *Information Display*, vol. 26, no. 3, pp.16-21, Mar. 2010.

[2] H.R. Kim, Y.K. Choi, S.H. Byun, S.W. Kim, K.H. Choi, H.Y. Ahn, J.K. Park, D.Y. Lee, Z.Y. Wu, H.D. Kwon, Y.Y. Choi, C.J. Lee; H.H. Cho, J.S. Yu, M. Lee, "A mobile-display-driver IC embedding a capacitive-touch-screen controller system," *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp.114-115, 7-11 Feb. 2010.

[3] T.H. Hwang, W.H. Cui, I.S. Yang, and O.K. Kwon, “A highly area efficient controller for capacitive touch screen panel systems,” *IEEE Trans. Consumer Electronics*, vol.56, no.2, pp.1115–1122, 2010.

[4] C.H. Krah and L. Altos, Apple Inc., Cupertino, CA (US), “Analog boundary scanning based on stray capacitance” U.S. Patent 11/650,511. Jan. 3, 2007.

[5] “ATMEL capacitive touch technology,” Argussoft. [Online]. Available: http://www.argussoft.ru/webroot/delivery/files/library/argussoft/seminars/atmel_2010_04_Crocus/04-Intro to Cap Touch.pdf

[6] H. Philipp, “Charge transfer capacitance measurement circuit” U.S. Patent 6,466,036 B1. Oct. 15, 2002.

Lifestyle Improvement Support System Considering Context of a User

Tomohiro Suzuki, Masahiro Inoue, Senior Member, IEEE

***Abstract*-- Recently, a healthcare system utilizing electronic technology is attracting attention in the world. People are becoming increasingly conscious about health. Current healthcare services increase health awareness by providing information, but don't utilize the information. Moreover, medical service has problem of lacking the concept of prevention. For those problems, this research proposes a lifestyle improvement support system which shows user how to use advice. At the time of shopping and cooking, the system provides advice that match user's context. By using the context, the system provides proper advice. The previous research found that user's lifestyle is deeply related to the content of advice. In order to solve those problems, this research includes budget, work load, time in advice as context. This research performs a comparative evaluation of the proposed algorithm and the algorithm of the previous paper.**

I. Introduction

Recently, a healthcare system utilizing electronic technology is attracting attention in the world because there are many factors that increase health awareness of young generation, the evolution of machinery and the device, and improvement of communications infrastructures. As a result of poor diet and lack of exercise, lifestyle diseases are increasing. An eating habits and the lifestyle disease are related closely. "Designer foods programs" started in the United States to clarify effective food in prevention of cancer and the element from 1990. Prevention by material of food has been focused. For example, cancer can be prevented by taking vegetable. Therefore a healthcare service is expected to improve user's eating habits.

A. Current healthcare service

Current healthcare services are the following three points.

(1) A system which offers advice using a website.
The user obtains a lifestyle-related disease remedy to be related to by inputting own subjective symptoms. A user gets the lifestyle-related disease remedy which is related in inputting one's subjective symptoms. As a feature, its health condition can be grasped easily. And improvement in a user's healthy consciousness can be expected. And a user understands the action for health enhancement.

(2) A system which provides advice suitable for an individual.

Health check, advice systems have been proposed. On the basis of subjective symptoms biological information, such as pulse and blood pressure, the systems generate the appropriate advice to user.

(3) A system that provide advice about user's behavior.

The system recognizes an action such as walking and the running. And the system produces advice from over's and shorts for the targeted value.

This research proposes system to solve the problems by foods. Specifically, the research suggests lifestyle improvement support system which urges user to use advice. Using the context, it is possible to provide advice by the situation.

B. Problems of Healthcare Service

Problems of healthcare services utilizing electronic technology and medical services are the following three points.

(1) The contents of advice are common.

Since a system carries out abstract advice suitable for many people, the user has to devise the lifestyle remedy based on health condition and a life.

(2) The opportunity of the improvement is not given to a user.

Even if a user receives advice, user has to consider action. Therefore, there is a possibility of becoming useless.

(3) The advice which a system issues is not taken into consideration from both health condition and action.

The system which presents the advice about a lifestyle from action cannot be provided with advice in consideration of the user's health condition.

This research proposes system to solve the problems by foods. Specifically, the research suggests lifestyle improvement support system which urges user to use advice. Using the context, it is possible to provide advice by the situation.

II. Proposed System

A. Definition of Context and Context-Aware

Context is information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.

A system is context-aware, if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task [1].

B. System Proposal

Using the context, the research suggests a system supporting improvement of living at the time of purchasing food. The system urges user to use advice. At the time of shopping and cooking, the system provides advice that match user's context. By using the context information, it is possible to provide advice. The system provides advice at the time to use the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

advice, reduces lifestyle diseases and urges user to increase health awareness. Here is characteristic of proposed system.

(1) The system provides the advice that match the condition of user using regular diet and user's health, previous purchase.

C. Problems from previous research

Previous research[2] built and evaluated the simulation system. Figure1 shows proposed operation of the system. Previous research evaluated whether the content of advice was utilized by the user, and whether the system provides advice timely. As a result of the evaluation, the system provided advice at timing to use. Table1 is Part of the evaluation results. However new problem was found. User's lifestyle is deeply related to the content of advice. Element of money and time and effort are important for contents of advice. The user don't have enough money, might not be able to buy the recommended foods. And when a user makes recommended menu, user may need other foods. It might be difficult. Moreover, time and an effort which the user cooks a recommended menu might not be suitable for user's lifestyle.

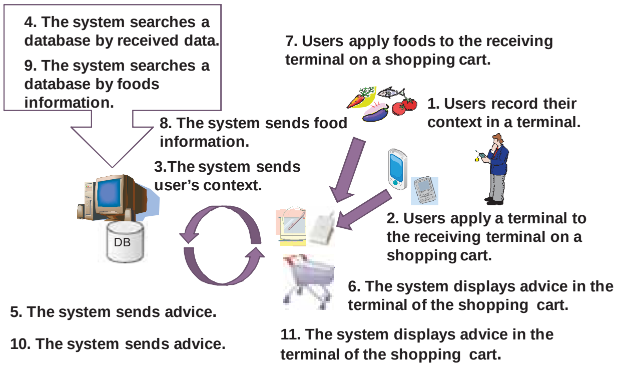

4. The system searches a database by received data.
9. The system searches a database by foods information.
7. Users apply foods to the receiving terminal on a shopping cart.
1. Users record their context in a terminal.
8. The system sends food information.
3.The system sends user's context.
DB
2. Users apply a terminal to the receiving terminal on a shopping cart.
6. The system displays advice in the terminal of the shopping cart.
5. The system sends advice.
10. The system sends advice.
11. The system displays advice in the terminal of the shopping cart.

Figure 1. Operation of Proposed System

Table1. Evaluation result

Evaluation items	Users		
	A	B	C
Did the user use advice displayed while shopping?			
Using advice for reference, the user bought foods.	+	±0	+
Using advice for reference, the user did not buy foods.	±0	-	±0
Did the user use advice displayed after shopping?			
Using recipes of advice for reference, the user cooked.	+	+	+
Using menus of advice for reference, the user cooked.	+	±0	+
The user used advice at next shopping and cooking.	-	±0	-

D. The proposal of an improvement algorithm

To urge user to use advice, the system includes budget, work load, time in advice as context. Figure2 shows context related to the content of advice. Figure3 shows the system narrows advice.

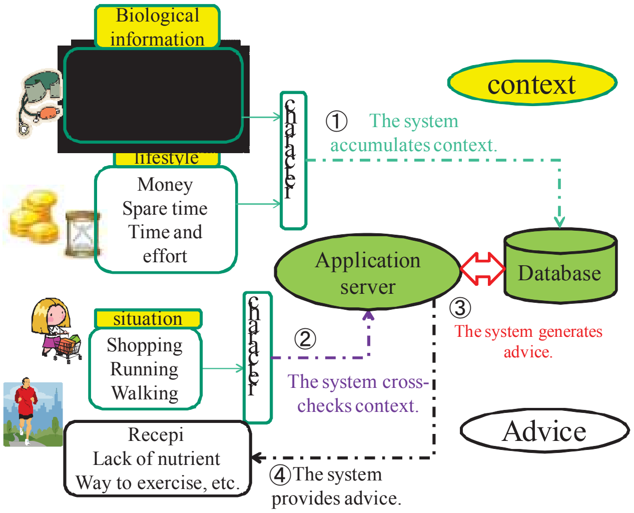


Figure 2. Relations of context and the advice

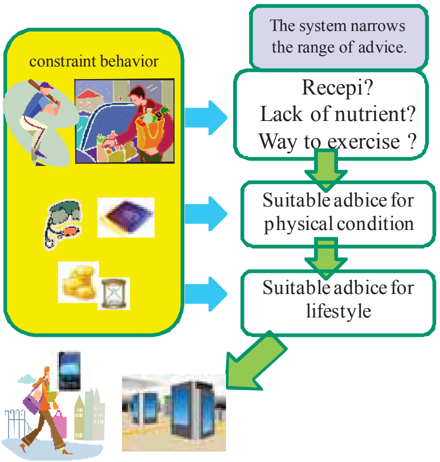

constraint behavior
The system narrows the range of advice.
Recepi?
Lack of nutrient?
Way to exercise ?
Suitable adbice for physical condition
Suitable adbice for lifestyle

Figure 3. The system narrows range of advice.

This research performs a comparative evaluation of the proposed algorithm and the algorithm of the previous paper.

CONCLUSION

Recently, a healthcare system utilizing electronic technology is attracting attention in the world. The previous research found that user's lifestyle is deeply related to the content of advice. To urge user to use advice, this research includes budget, work load, time in advice as context.

• REFERENCES

[1] Dey, A. K. and Abowd,G. D .,"Toward a Better Understanding of Context and Context-Awareness", In Proceedings of the CHI2000 Workshop on The What, Who, Where, and How of Context-Awareness, 2000.

[2] Tomohiro Suzuki, Masahiro Inoue, Lifestyle Improvement Support System Using Context IEEE Symposium on Consumer Electronics, ISCE2011, Digest of Technical Papers, Singapore, June 2011.

A Video Game Controller with Skin Stretch Haptic Feedback

Ashley L. Guinan[1], Nathaniel A. Caswell[1], Frank A. Drews[2], William R. Provancher[1], *Member, IEEE*
[1]Haptics and Embedded Mechatronics Laboratory, University of Utah
[2]Department of Psychology, University of Utah

***Abstract* — This paper presents the design of a game controller with integrated skin stretch feedback, a new form of touch feedback. This form of feedback can be used to provide directional information, as well as tactile gaming effects to a user through the input thumb joysticks. Prior testing has shown that a user can perceive and respond to single direction cues given through skin stretch feedback at the thumb joysticks. This paper presents further experiments to characterize user interaction and responses to feedback as a function of the relative timing between skin stretch cues given on both of the controller's thumb joysticks.**

I. INTRODUCTION

The video gaming industry includes a large and growing market targeted at improving the user experience and interaction with games. Many new ways for consumers to provide input to video games have recently been introduced through devices such as the Microsoft® Xbox Kinect$^{\text{TM}}$, Sony PlayStation® Move$^{\text{TM}}$, and Nintendo® Wii$^{\text{TM}}$ Remote. However, the haptic feedback provided to users has remained limited to the vibrotactile (rumble) feedback found in handheld dual thumb joystick (thumbstick) game controllers.

Haptic feedback offers a way for consumers to receive feedback through the sense of touch, rather than simply through on-screen graphics and audio feedback. Skin stretch feedback provides a way of communicating directional feedback, and was developed to fit within a compact handheld device [1]. While other haptic devices can communicate direction using pin arrays [2, 3] or vibrotactors [4, 5], those devices require more space than a small handheld device such as a game controller would allow. We have integrated two skin stretch feedback mechanisms into the thumbsticks of our game controller shown in Fig. 1. This allows us to provide two independent direction cues to a user's thumbtips.

Previous studies with skin stretch feedback [6, 7, 8, 9] have focused on feedback to one fingertip or thumbtip. The game controller uses two skin stretch displays, so it is important to determine an effective way to communicate with users through each tactor. The possibility of stimulus masking exists, where one cue may be masked by a previous or simultaneous stimulus. An experiment has been designed to investigate this possible effect and the device design and experimental results are briefly discussed in this paper.

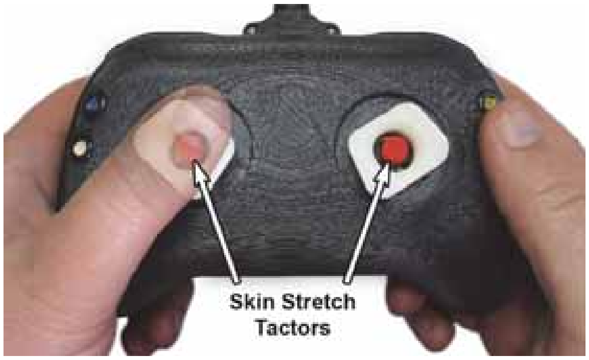


Fig. 1. Skin stretch feedback is integrated into the tops of the thumb joysticks.

II. DEVICE DESCRIPTION

Our lab's game controller is similar in function to modern game controllers, but with the addition of skin stretch feedback. Our controller includes buttons and thumbsticks to provide input similar to Sony PlayStation®3 and Microsoft® Xbox 360 controllers. Our controller also provides vibration feedback as found in current game controllers. In addition, direction cues and tactile effects are provided to the user via skin stretch tactors integrated into the top of each thumbstick of our controller. The design of the skin stretch feedback mechanisms used within the thumbsticks was previously presented in [1].

Our game controller utilizes a microcontroller to read user input from the buttons and thumbsticks, and communicate these states to a host computer. The microcontroller also controls the position of the two skin stretch feedback devices and sets the magnitude of vibration for each vibrotactor of the game controller. The microcontroller exchanges information with a PC through RS-232 serial communication, with the PC receiving button and thumbstick state information and sending tactor and vibrotactor values to the microcontroller at a rate of 60 Hz, to be compatible with the Microsoft XNA development environment. An overall view of the system architecture is shown in Fig. 2. Further details of the game controller will be presented in a future publication.

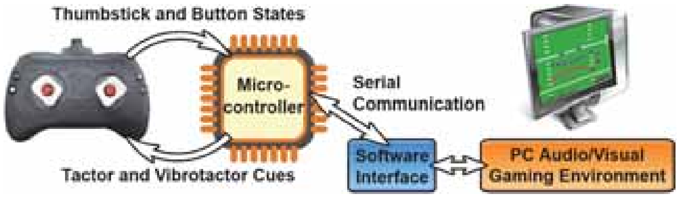


Fig. 2. System schematic. A microcontroller inside the game controller communicates thumbstick and button states to a PC and controls the vibrotactors and skin stretch feedback devices.

This work was supported, in part, by the National Science Foundation under awards IIS-0746914, IIS-0904456, and DGE-0654414. Additional support was provided by a University of Utah Research Foundation *Technology Commercialization Project* and the University Of Utah College Of Engineering Wayne Brown Fellowship.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. EXPERIMENT

A pilot experiment with nine participants was performed to investigate responses to two simultaneous or sequential skin stretch cues; one through the right thumbstick and one through the left thumbstick. A cue was delivered to one thumb, with the tactor motion limited to one of four orthogonal directions (forward, right, backwards, left). After a short delay, a second cue was delivered to the other thumb. The two cues were generally in different directions, as a combination of 16 pairs of stimuli were possible, with only 4 pairs including both tactors moving in the same direction. The order of the cues were balanced, with the first cue sometimes delivered to the right thumb and other times delivered to the left thumb. The lag times between cues tested were: 0 s, 0.05 s, 0.1 s, 0.25 s, 0.5 s, 0.75 s, 1.0 s, and 1.5 s. In cases where the lag time was zero, both tactors began their motions at the same time. A total of 128 cue pairs were included in the experiment, one each of the 16 directional combinations for each of the 8 lag times.

Participants in the experiment were told to respond to the cues as quickly and accurately as possible by moving the thumbsticks in the perceived direction of the skin stretch cue. Directional skin stretch cues were composed of a motion of the skin stretch tactor from the center position, radially outward approximately 1 mm in the cued direction, a 0.3 s pause to prevent out-back masking, followed by a return to center. Skin stretch tactor motions were approximately 10 mm/s.

Fig. 3 shows accuracy results for each tested lag time, where a response was correct when **both** thumbsticks were moved in the correct direction to match their respective cues. As previous experiments [6, 8, 9] were designed for responses with one finger, the accuracies of those tests are higher. In our prior experiments (e.g., [8, 9]), if a participant performed by responding correctly for 90% of the cues on one thumb, their performance would at best be expected to be 81% accurate when responding on both thumbs.

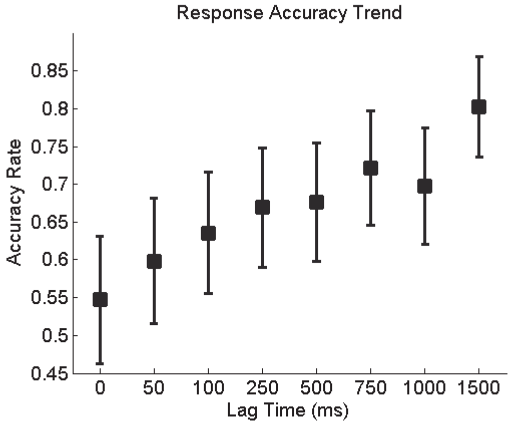


Fig. 3. Percent accuracy and 95% confidence intervals of each tested lag time. An accurate/correct answer, as reported on this plot, is composed of correct responses on both the left and right thumb.

The results from this pilot experiment show a trend of user performance improving as the time between cues increases. Accuracy seems to peak around 0.75 s, but further improvements in accuracy can be seen at 1.5 s between direction stimuli. For lag times of 1.5 s, response accuracy nears levels expected based on results from [8, 9], where participants responded to one cue at a time. Future tests with more participants will be performed to further investigate user performance for specific lag times, based on the response accuracies reported here and the average response times observed in the pilot experiment.

IV. CONCLUSION AND FUTURE WORK

A game controller has been designed to include skin stretch haptic feedback, which can aid users by providing directional information as well as tactile effects. A pilot experiment was then conducted to characterize response accuracy rates and response times for several different lag times between stimuli. As expected, results show that response accuracies increase as the lag time between two tactile cues increases.

Future tests will further investigate responses for select lag times to determine the optimum and/or minimum suggested lag time to use when providing differing skin stretch cues through the two thumbsticks of this device. In addition, experiments will be conducted to examine potential masking effects for skin stretch feedback in the presence of vibrotactile feedback, the current form of haptic feedback found in most game controllers. In addition, experiments are planned to quantify the performance benefits of skin stretch feedback in video games.

REFERENCES

[1] Gleeson, B.T.; Horschel, S.K.; Provancher, W.R.: "Design of a fingertip-mounted tactile display with tangential skin displacement feedback," IEEE Transactions on Haptics, 2010, 3(4), pp. 297-301.

[2] Wang, Q., and Hayward, V.: "Biomechanically optimized distributed tactile transducer based on lateral skin deformation," International Journal of Robotics Research, 2009, 29(4), pp. 323-335.

[3] Ki-Uk Kyung; Jun-Young Lee: "Ubi-Pen: a haptic interface with texture and vibrotactile display," IEEE Computer Graphics and Applications, 2009, 29(1), pp. 56-64.

[4] Yao, H.-Y. and Hayward, V.: "Design and analysis of a recoil-type vibrotactile transducer," Journal of the Acoustical Society of America, 2010, 128(2), pp. 619-627.

[5] Elliott, L.R., van Erp, J.B.F., Redden, E.S., Duistennaat, M.: "Field-based validation of a tactile navigation device," IEEE Transactions on Haptics, 2010, 3(2), pp. 78-87.

[6] Gleeson, B.T., Horschel, S.K., and Provancher, W.R.: "Perception of direction for applied tangential skin displacement: effects of speed, displacement and repetition," IEEE Transactions on Haptics-World Haptics Spotlight, 2010, 3(3), pp. 177-188.

[7] Medeiros-Ward, N.; Cooper, J.M.; Doxon, A.J.; Strayer, D.L.; and Provancher, W.R.: "Bypassing the bottleneck: the advantage of fingertip shear feedback for navigational cues," Proc. of the Annual Meeting of the Human Factors and Ergonomics Society, 2010.

[8] Koslover, R.L.; Gleeson, B.T.; de Bever, J.T.; Provancher, W.R.: "Mobile navigation using haptic, audio, and visual direction cues with a handheld test platform," IEEE Transactions on Haptics, 5(1), Jan.-March 2012, pp. 33-38.

[9] Guinan, A.L.; Koslover, R.L.; Caswell, N.A.; Provancher W.R.: "Bi-manual skin stretch feedback embedded within a game controller," 2012 Haptics Symposium, Vancouver, B.C., Canada, March 4-7, 2012, pp. 255-260.

Visual Saliency Based on Selective Integration of Feature Maps in Frequency Domain

Ki Tae PARK[*], *Member, IEEE*, Jeong Ho LEE[**], Student *Member, IEEE*
and Young Shik MOON[**,†], *Member, IEEE*
[*]*Center for Integrated General Education, Hanyang University, South Korea*
[**]*Dept. of Computer Science and Engineering, Hanyang University, South Korea*

Abstract—**In this paper, an automatic method for extracting visual saliency based on selective integration of feature maps in frequency domain is proposed. Feature maps are calculated by measuring the Bayes spectral entropy. In order to extract visual saliency effectively, feature maps are first generated from three images separated into Y, Cb, Cr channels, respectively. Then, by selectively integrating feature maps, visual saliency is finally extracted. Experimental results have shown that the proposed method obtains good performance of visual saliency under various environments containing multiple objects and cluttered backgrounds in natural images.**

I. INTRODUCTION

As an emerging field of computer vision, the extraction of visual saliency has been researched in various areas such as content based image retrieval, image compression (MPEG 2000), watermarking and object recognition. Due to these reasons, many approaches of the visual saliency extraction have been proposed [1-2]. The human visual system has a remarkable ability to automatically attend to salient regions, known as focus of attention. When taking a picture, people adjust the camera's focus into specific objects. The specific objects in the picture are sharper than backgrounds. Therefore, by measuring the degree of sharpness of each sub-block in an image, salient regions including the specific objects can be detected [3].

In this paper, we propose a novel approach to extracting visual saliency from the frequency domain, without converting to the spatial domain. Therefore, the proposed method is possible to directly extract visual saliency from the compressed domain such as JPEG image. To this end, feature maps are generated by calculating Bayes spectral entropy on DCT frequency domain as a measure of camera focus [4]. Finally, by selectively integrating the feature maps, visual saliency is extracted.

II. PROPOSED METHOD

The proposed method of visual saliency detection consists of two steps. In the first step, three focus maps are generated based on Bayes spectral entropy in DCT frequency domain. To this end, an input image is transformed from RGB color space to YCbCr color space. Both Cb and Cr images are resized by half their size. Then, after performing a discrete cosine transform on each Y, Cb, and Cr image, we generate three feature maps by calculating Bayes spectral entropy on DCT coefficients. In the final step, visual saliency is extracted by selectively combining the feature maps.

A. *Feature maps generation by calculating the Bayes spectral entropy*

In this paper, feature maps generated by Bayes spectral entropy are utilized for extracting visual saliency. The Bayes spectral entropy technique proposed by Kristan et al. is an approach for measuring an autofocus of a digital camera [4]. Generally, since a focus of a digital camera is adjusted into not backgrounds but main objects for obtaining an image with better quality, it is possible to extract the visual saliency based on the focus measure. The Bayes spectral entropy is to measure a sharpness value by using an image spectrum based on the discrete cosine transform. The sharpness value is calculated for each 8x8 sub-image and their mean value is considered as a focus measure of an image. In this paper, the sharpness value based on Bayes spectral entropy for each 8x8 sub-image is considered for generating feature maps. The measure of an image sharpness based on Bayes spectral entropy is defined as (1).

$$M_{Be}(t) = 1 - \frac{\sum_{\omega+\upsilon\le t} \left|F_C(\omega,\upsilon)\right|^2}{\left(\sum_{\omega+\upsilon\le t} \left|F_C(\omega,\upsilon)\right|\right)^2} \qquad (1)$$
$$\omega,\upsilon \in \{0,\ldots,7\}$$

where $F_C(\omega,\upsilon)$ implies a coefficient of DCT at (ω,υ) position. M and N denote height and width of sub-image, respectively. M and N are set to 8 in this paper. Accordingly, a sub-image transformed by DCT consists of 64 coefficients including DC and ACs. Figure 1 shows the frequency distribution of the DCT coefficients [5].

According to [5], most of the energy is occupied by some components in the lower sequence order. As shown in figure 1, it is found that six coefficients (DC and 1^{st}~5^{th} AC) contain 99% of the total energy. Therefore, we also utilize the six coefficients for generating the focus maps. More specifically DC holds 98.5% of the total energy, and 1^{st}~5^{th} AC hold 74.2% of the remaining energy. Figure 2 shows the focus maps generated by the Bayes spectral entropy technique.

This work was supported by a National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MEST) (No.2012002464).

† Corresponding author

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

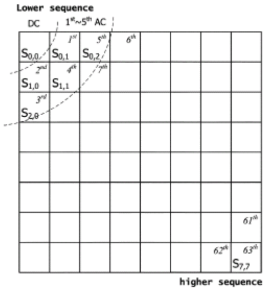


Fig. 1. Frequency distribution of the DCT coefficients

B. Visual saliency extraction by selective integration of feature maps

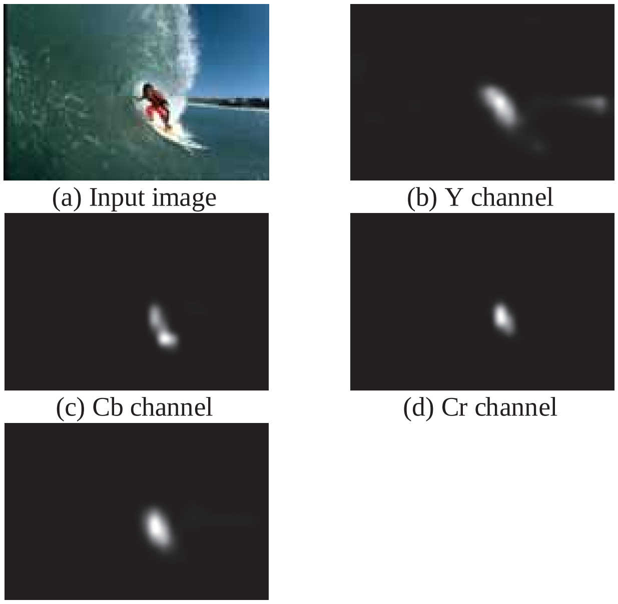
In order to extract visual saliency, we create a saliency map by combining the three feature maps. In this paper, a saliency map is generated by selectively combining the feature maps, while most existing methods combine all feature maps. To this end, the three feature maps are first combined for a reference map which is utilized for generating a saliency map. Then, three correlation values between the reference map and the three maps are calculated respectively. If the all correlation values are over 0.9, the three focus maps are combined for a saliency map. Otherwise, the corresponding focus map with the highest correlation value is selected as a core map. In order to select other maps, two correlation values between the core map and the other focus maps are calculated respectively. If the higher correlation value is over 0.9, the corresponding focus map is selected in addition to generating the saliency map. Accordingly, the two selected focus maps are combined for a saliency map. If the correlation value is under 0.9, the only core map is considered as a saliency map. Figure 2(e) shows a saliency map.

III. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed method, we have utilized the image data used in [2] and the Corel image data with various environments such as multiple object and cluttered backgrounds. Figure 3 shows extraction results of visual saliency maps extracted by the proposed method. As shown in Figure 3, visual saliency including salient objects has been extracted effectively, independent to various environments.

IV. CONCLUSIONS

In this paper, we have proposed an approach to automatic extraction of visual saliency. In order to effectively extract the visual saliency, feature maps are generated from Y, Cb, and Cr channel images by using the Bayes spectral entropy technique. Then, a reference map is created by integrating the feature maps. For constructing a saliency map, we have proposed a new method to selectively combining the feature maps by considering mutual correlation between the feature maps, while most existing methods combine all feature maps. Experimental results have shown that the proposed method extracts visual saliency effectively.

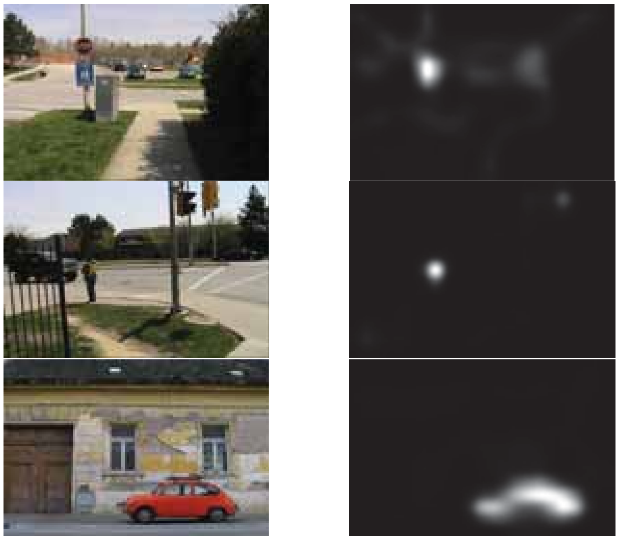


(a) Input image (b) Y channel

(c) Cb channel (d) Cr channel

(e) Saliency map

Fig. 2. Focus maps generated from each Y, Cb, and Cr channel.



(a) Input images (b) Visual saliency maps

Fig. 3. Experimental results of visual saliency extraction.

REFERENCES

[1] Yu, Y, Wang, B., Zhang, L., “Pulse Discrete Cosine Transform for Salieny-based Visual Attention,” *Proc IEEE 8th Int. Conf. on Development and Learning*, (2009), pp. 1-6.

[2] Sanchez, V., Basu, A., and Mandai, M. K., “Prioritized region of interest coding in JPEG2000,” *IEEE Trans-Circuits and Systmes for Video Technology*, Vol. 14(2004), pp. 1149-1155.

[3] Ki Tae Park, Min Su Park, Jeong Ho, Lee, and Young Shik Moon, “Detection of Visual Saliency in Discrete Cosine Transform Domain,” *Proc IEEE 30th Int. Conf. on Consumer Electronics*, (2012), pp. 130-131.

[4] Kristan, M., Pers, J., Perse, M., and Kovacic, S., “A Bayes Spectral Entropy Based Measrue of Camera Focus Using A Descrete Cosine Transform,” *Pattern Reconition Letters,* Vol. 27, No. 13(2006), pp. 1431-1439.

[5] Lee, S., Yoo, J., Kumar, Y., and Kim, S., “Reduced Energy Ratio Measure for Robust Autofocusing in Digital Camera,” *IEEE Signal Processing Letters,* Vol. 16, No. 2(2009), pp. 133-136.

Drag-and-Type: A New Method for Typing with Virtual Keyboards on Small Touchscreens

Taekyoung Kwon, *Member*, IEEE, Sarang Na, *Non-member*, IEEE, and Sang-ho Park, *Non-member*, IEEE

Abstract — *Smartphone users are experiencing a difficulty in typing alphanumeric keys with their thumbs frequently through a tiny virtual keyboard on small touchscreens. We explore a new style of typing method, Drag-and-Type, on the touchscreen*[1].

Index Terms — Smartphone, touchscreen, virtual keyboard.

I. Introduction

A smartphone is now becoming a part of our lives and turns out to be one of the most popularly used consumer electronic devices today. Its small flat touchscreen enables us to navigate various services and applications very easily, promptly and intuitively with our own fingers. The small touchscreen is also changing the way of typing alphanumeric characters. Without a physical keyboard, today's smartphones popularly present virtual keyboards in software, under the high-resolution of small touchscreen, e.g., 4.8" 1280x720 pixels (306 ppi) and 3.5" 640x960 pixels (326 ppi) in commodities. It is now commonplace to see users who are tapping their fingers on the small virtual keyboard, aka soft keyboard but there are three concerns that motivate our study. Firstly, users are frequently experiencing a difficulty in typing alphanumeric characters exactly on minute keys by their thick thumbs [1]. Although the higher resolution of touchscreen facilitates smaller keys for constructing a full size keyboard layout, users may prefer a larger key for typing characters by their thumbs more easily. Such a larger key may only allow a partial keyboard layout on the small touchscreen, e.g., separate layouts for alphabets and numeric (and/or special) characters, and pop-up keys for more characters at best. This tendency may less benefit from the recent and future advance in the high-resolution touchscreen. Note that the partial keyboard layout eventually requires a number of switches among several layouts, whereas the full keyboard layout does not. The well known typing method for blinded users, e.g., VoiceOver, also works on the partial keyboard. Secondly, the visual echo of entered key, that is, the most widely used response method in virtual keyboard, can be hidden possibly under a thick thumb and its blunt touch, as illustrated in Fig. 1-(a). This phenomenon may hinder users from being aware of the real entry, even worse in the case of password entry where no textual echo is preferred for security reasons, as illustrated in Fig. 1-(b). Finally, there is a concern that the touch event and its geometric data can be captured and exploited by spyware if a user types secret characters, such as passwords. Thus, we are motivated to study a new method for typing alphanumeric characters on a full layout of virtual keyboard presented on a small but high-resolution touchscreen, in the way of resisting spyware as well as making large-thumb users enter keys more accurately. In this paper, we devise a new *drag-and-type* method quite different from a conventional direct typing. In particular we consider two kinds of drag-and-type methods; *drag-and-tap* and *drag-and-drop* on a full layout of virtual keyboard. The former method considers a standard full size keyboard and works with separate tapping actions, while the latter method operates with a split keyboard like the Natural Keyboard and with dragging actions only.

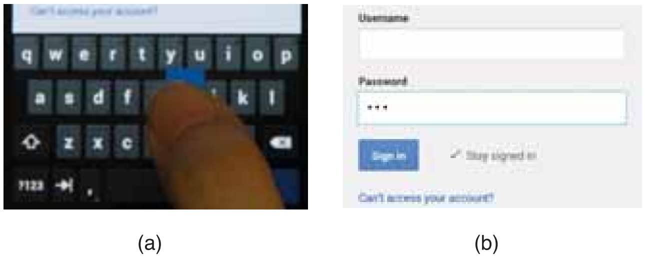


Fig. 1. Hidden echo problem. (a) Echo of entered key 'b' is hidden under a thumb. (b) No textual echo in many cases of password entry.

II. Related Work

A number of virtual keyboards with various layouts, such as OPTI, Atomik, Metropolis and FITALY [3,4,5], have been proposed to rethink the standard qwerty keyboard with regard to convenience in mobile electronic devices. However, they are still painful on small touchscreens due to requiring a user's thumb on tiny keys for direct typing. An attempt to overlay larger split-keys in a pie menu on a virtual keyboard [2] improves such a tendency but, on the other hand, causes visual occlusion and two layered typing, which may be undesirable for the fast and/or consecutive typing of characters.

III. Drag-and-Type Methods

On the flat touchscreen, finger touch actions are classified into two actions, i.e., tapping and dragging. The former is

[1] This work was supported in part by the IT R&D program of MKE/KEIT [10039180, Intuitive, convenient and secure HCI-based usable security technologies for mobile authentication and security enhancement in mobile computing environments]. This research was also supported in part by MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2012-H301-12-4008) supervised by NIPA (National IT Industry Promotion Agency).

T. Kwon is with the Dept. of Computer Engineering, Sejong University, Seoul, 143-747, Korea (e-mail: tkwon@sejong.edu).

S. Na is with the Dept. of Computer Engineering, Sejong University, Seoul, 143-747, Korea (e-mail: tot@sju.ac.kr).

S. Park is with the Dept. of Computer Engineering, Sejong University, Seoul, 143-747, Korea (e-mail: superh1@sju.ac.kr).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

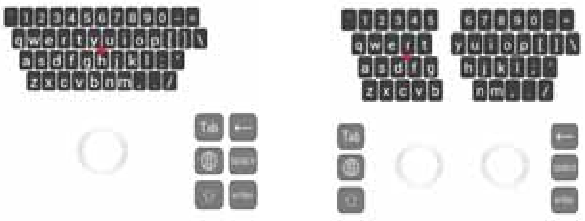
activated usually for a click event, while the latter is done for scrolls and/or more functions, such as pointing and navigating. Multiple touch actions may involve simultaneous and/or consecutive actions of tapping and dragging. We consider those actions for devising a new typing method. First of all, we point out that the dragging action enables more accurate targeting to a tiny item, that is, a tiny key on virtual keyboards. Another point is that the user's thumb typing on a small touchscreen is done eventually as like the hunt and peck typing, aka two-fingered typing, on a real keyboard. Thus, it is expected that if a small touchscreen represents a full size keyboard on which tiny keys are located close to each other, then the dragging actions of pointing would be quite familiar as well as more accurate than the direct tapping actions. Although the accuracy is obtained at the cost of dragging time, it would be reasonable to think that less erroneous typing is also attractive in a large number of applications. So we devise two sorts of drag-and-type methods in that sense.

A. *Drag-and-Tap*

The first method presumes a full layout of standard qwerty keyboard in small size and makes a user navigate the virtual keyboard by dragging one finger, e.g., the left thumb, and type a highlighted (selected) character by tapping on any blank area with another finger, e.g., the right thumb. Note that the split tapping is also used in other methods, such as VoiceOver. Fig. 2-(a) illustrates a prototype layout of our drag-and-tap keyboard. The red dot is used to select a target key, while a larger grey circle indicates a dragging area. Deep grey keys are used for rendering more functions onto the keyboard, such as tab, language, shift, backspace, space, and enter. Fig. 3-(a) is a snapshot of the drag-and-tap method in use.

B. *Drag-and-Drop*

The second method presumes a full layout of split qwerty keyboard that is similar to the Natural Keyboard, and makes a user navigate the virtual keyboard by dragging two fingers simultaneously, e.g., the left and right thumbs, and type a highlighted (selected) character by releasing (dropping) the corresponding finger. Fig. 2-(b) illustrates a prototype layout of our drag-and-drop keyboard. The red and blue dots are used to select target keys, respectively, while larger grey circles indicate dragging areas. Deep grey keys are also split for rendering more functions onto the keyboard. Fig. 3-(b) is a snapshot of the drag-and-drop method in use.

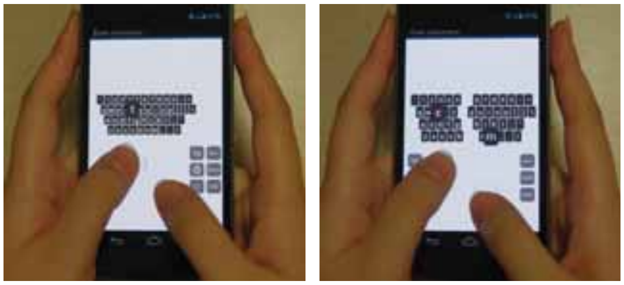


(a) (b)

Fig. 2. Prototype designs of drag-and-type methods. (a) Drag-and-tap. (b) Drag-and-drop. (Large circles indicate places to touch, whereas red and blue dots signify the pointers.)

IV. IMPLEMENTATION AND EXPERIMENT

We implemented a prototype system on the Google Android platform and conducted a user study. As illustrated in Fig. 3, users can drag-and-type alphanumeric characters by fingers in both methods. We gained impressive results in the user study. We compared two drag-and-type methods and the regular virtual keyboard with regard to entry time and error rates. We recruited 8 participants (6 males, 2 females) with academic education and average age of 27.4 years. The participants were assigned three methods in random sequences in the within group study, and asked to type alphabets in sequence, i.e., from a to z, for 5 times, and decimals in sequence, i.e., from 1 to 0, for another 5 times, after training themselves up to twenty minutes. The average entry time was 19.074 sec, 23.097 sec, and 23.986 sec for alphabets, respectively, in the regular, drag-and-tap, and drag-and-drop methods. The average number of backspaces was 4.375, 0.625, and 0.5, respectively. The number of mistyped trials was 3, 1, and 0, respectively. As for decimals, the average entry time was 6.442 sec, 8.204 sec, and 10.297 sec, respectively. The average number of backspaces was 2.500, 0.125, and 0.125, respectively. The number of mistyped trials was 1, 0, and 0, respectively. When considering that participants are already used to the regular virtual keyboard, the experimental results are fairly remarkable. We will manipulate more details of implementation and through user experiments in the full paper.



(a) (b)

Fig. 3. Screenshots of drag-and-type methods in use. (a) Drag-and-tap. (The right thumb is used for tapping.) (b) Drag-and-drop.

V. CONCLUSION

The drag-and-type method is a novel typing method for both normal and blinded users on a small flat touchscreen. We will explore more applications and functions for this method.

REFERENCES

[1] M. Agarwal, M. Mehra, R. Pawar, and D. Shah, "Secure authentication using dynamic virtual keyboard layout," In Proc. of *ICWET'11, ACM press*, pp. 288-291, 2011.

[2] K. Go and L. Tsurumi, "Arranging Touch Screen Software Keyboard Split-Keys based on Contact Surface," In Proc. of *CHI'10, ACM press*, 2010.

[3] I. S. Mackenzie and S. X. Zhang, "The design and evaluation of a high-performance soft keyboard," In Proc. of *CHI'99, ACM press*, pp. 25-31, May 1999.

[4] S. Zhai, M. Hunter, and B. A. Smith, "Performance optimization of virtual keyboards," *Human-Computer Interaction*, vol. 17, 2002.

[5] J. D. Ichbian, "Method for designing an ergonomic one-finger keyboard and apparatur therefor," In *US patent 5487616*, issued 1996-01-10.

A Context Aware Engine for Multimedia Applications on Smartphone

Sangdo Park, Junghyun Park, Paul Barom Jeon
Intelligent Computing Lab.
Samsung Advanced Institute of Technology
{sdpark, jhpsy.park, paul.barom.jeon}@samsung.com

Su Myeon Kim
Data Intelligence Lab.
Software R&D Center, Samsung Electronics Co., Ltd.
sumyeon.kim@samsung.com

***Abstract*—High complexity of understanding user's current situation is a crucial barrier to popularize context aware applications on smartphone. In order to overcome this barrier, we propose a context aware engine which eases the development of intelligent applications. Key feature of the engine is dynamic reconfiguration of the knowledge base depending on the current environments using linked data. We validated the usefulness of this feature via two in-house applications.**

I. Introduction

The rapid growth of smartphone's performance has been reshaping our daily life. People can shorten waiting time at a bus stop equipped with a live positioning service of public transportations and can connect to digital world at everywhere. However, context aware applications are not widely used even though understanding user's current situation and performing useful actions without user's intervention have been the subject of many researches. We believe that this is largely because existing works in context awareness have assumed an environment in which sensors and devices are tightly linked to specific application. Application developers and/or users are thus required to deal with each context aware application individually, which increases implementation cost considerably, limits flexibility, and may make unrealistic demands on the user.

In this paper, we present a mobile context aware engine for multimedia applications. Our engine can dynamically update multimedia content information from linked data as well as domain knowledge in order to adapt to different situations and scenarios, and it provides context aware services designed to support 3rd party applications. The goal of the context engine is to ease the development of a wide range of intelligent applications that can exploit the services provided in order to furnish context aware behavior. We believe the proposed context aware engine will help intelligent applications be popularized like the success of mobile applications.

II. Context Aware Engine

To support context aware multimedia application, our engine manages information about the mobile device, its environment and the user. When the user starts to control multimedia contents, our engine reconfigures knowledge model and reasons current context using semantic technology Web Ontology Language(OWL) and DL reasoning.

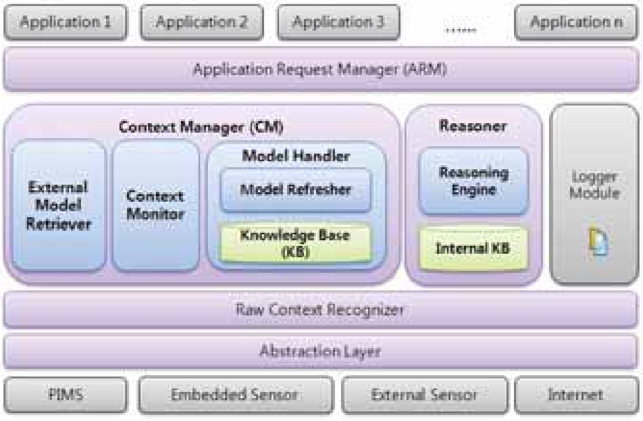


Fig. 1. Context aware engine architecture.

A. Architecture

Figure 1 shows the architecture of proposed context aware engine embedded in Mobile devices. Context Monitor in Context Manager continuously monitors the environmental changes or device status such as new multimedia content played. External model retriever communicates with 3rd party server on the internet and gets proper knowledge information. Model handler has the responsibility to update knowledge base and to refresh knowledge models on-the-fly. Reasoner continuously updates the conclusions drawn from its knowledge base. Concluded knowledge are saved into internal knowledge base.

Application Request Manager(ARM) interfaces with applications which want to get contexts concluded by the Reasoner. Abstraction Layer(AL) and Raw Context Recognizer(RCR) formalize sensor data regardless of OS and hardware.

B. Reconfiguration Method

We designed three-layered context model hierarchy as shown in Figure 2. Initially, smartphone equips with a default model, which is essential to understand basic context, such as person's relation, location, time and sensor data. Temporally, default model can be extended with domain knowledge which describes specific situations. For example, "playing movie" domain knowledge is inserted to the Reasoner when user starts playing a movie. To conclude detailed context, domain model also can be extended with exhaustive information about

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

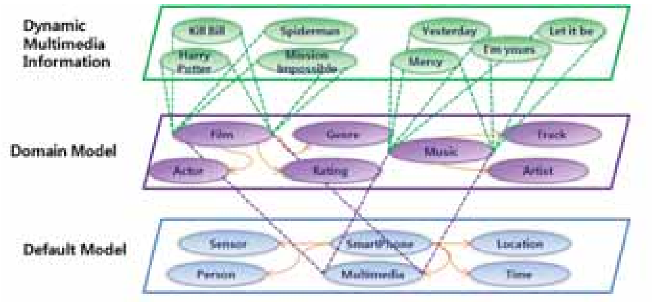


Fig. 2. Context model hierarchy.

```
PREFIX movie: <http://data.linkedmdb.org/resource/movie/>
DESCRIBE  ?subject ?genre ?actor ?director ?performance ?rating
WHERE
{
  ?subject  <http://purl.org/dc/terms/title> "Kill Bill".
  ?subject    movie:genre ?genre.
  ?subject    movie:actor  ?actor.
  ?subject    movie:director                ?director.
  ?subject    movie:performance             ?performance.
  ?subject    movie:rating ?rating.
}
```

Fig. 3. SPARQL query for movie ”Kill Bill”.

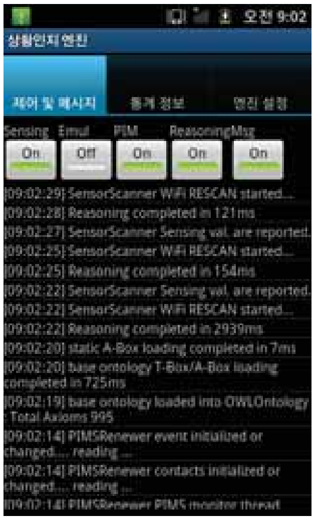
multimedia contents. All of the extended context models are integrated semantically with the default model.

III. EXPERIMENTS

We implemented the context aware engine using Java and context models as a form of ontology. To overcome computational resource limitation, we have developed a mobile semantic reasoner called delta-reasoner[1]. It gave acceptable completeness and decidability on Android smartphone.

When a user controlled a multimedia player application on smartphone, Context Monitor detected status changes through broadcast messages sent by the media player. Firstly, Model Handler added multimedia domain model into Reasoner. Secondly, External Model Retriever generated SPARQL query according to the received broadcast messages and retrieved content information from Linked Open Data(LOD)[2,3]. Figure 3 shows the automatically generated query requesting information about the movie Kill Bill. After multimedia content information was merged to default model, new conclusion about current context was made on-the-fly. Figure 4 shows control UI of the context aware engine service run as a Android background service. This service was so light as to conclud once every three seconds on the smartphone.

We made two sample applications to validate our context aware engine with only a few lines of code. One was generating warning sign to prohibit child to watch adult movies and the other was recommending music. Context aware engine concluded if the user was allowed to watch the playing movie by comparing user information in the PIMS[1] with movie rating . It was also concluded that if there was a remake version of music for the recommendation. Figure 5 shows sample



```
[09:02:29] SensorScanner WiFi RESCAN started...
[09:02:28] Reasoning completed in 121ms
[09:02:27] SensorScanner Sensing val. are reported.
[09:02:25] SensorScanner WiFi RESCAN started...
[09:02:25] Reasoning completed in 154ms
[09:02:22] SensorScanner Sensing val. are reported.
[09:02:22] SensorScanner WiFi RESCAN started...
[09:02:22] Reasoning completed in 2939ms
[09:02:20] static A-Box loading completed in 7ms
[09:02:20] base ontology T-Box/A-Box loading
completed in 725ms
[09:02:19] base ontology loaded into OWLOntology
: Total Axioms 995
[09:02:14] PIMSRenewer event initialized or
changed.... reading ...
[09:02:14] PIMSRenewer contacts initialized or
changed.... reading ...
[09:02:14] PIMSRenewer PIMS monitor thread
```

Fig. 4. Screen shot of context aware engine control UI.

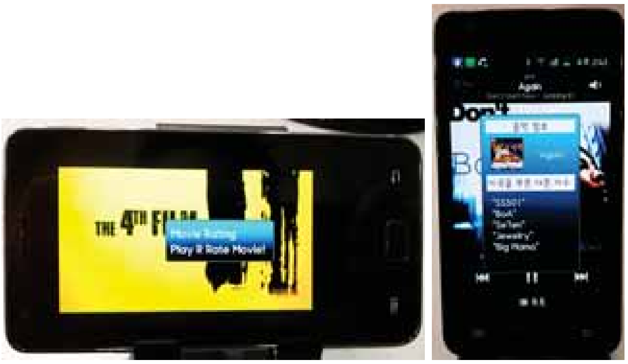


Fig. 5. Sample application UIs.(Left) warning message of prohibited user and (Right) recommended music list

applications UIs - the warning message overlapped on the screen and the list of recommended music respectively.

IV. CONCLUSION AND DISCUSSION

In this paper, we designed context aware engine for mobile devices. Our context aware engine, contrary to other existing ones, reconfigures its knowledge base on-the-fly by adopting semantic technologies to reflect the context of user, environments, multimedia and multimedia contents. We have validated this functionality using LOD and two in-house scenarios. The result shows that various intelligent apps can be easily implemented just by utilizing the proposed context aware engine.

REFERENCES

[1] Boris Motik, Ian Horrocks, and Su Myeon Kim, “Delta-Reasoner: a Semantic Web Reasoner for an Intelligent Mobile Platform,” in *Proc. WWW*,pp. 63–72, 2012.

[2] W3C, *SPARQL Query Language for RDF,* http://www.w3.org/RT/rdf-sparql-query/, 2008.

[3] Bizer Christian, Heath Tom and Berners-Lee Tim, “Linked Data - The Story So Far ,” *Int. J. Semantic Web Inf Syst.*, vol. 14, 5, (3), pp. 1–22, 2009.

[1] Personal Information Management System

A Seamless Remote User Interface System Supporting Multi-Screen Services in Smart Devices

Yuseok Bae and Jongyoul Park
Next Generation Computing Department, Big Data S/W Research Laboratory, ETRI, Korea

Abstract-- **Many smart devices getting introduced to the market provide multimedia services and various interactive applications. Efficient collaboration among these devices helps improve user convenience, user mobility, and multi-screen services. In this paper, we propose an architecture for a seamless remote user interface system supporting multi-screen services in smart devices that guarantees smooth user interface transition through efficient collaboration.**

I. INTRODUCTION

Efficient collaboration among smart devices is becoming an important factor for the continuous success of fast emerging smart devices such as phones, tablets, and TVs. In addition, multi-screen services manipulating smart devices are in the spotlight. The Remote User Interface (RUI) is a typical device collaboration service among smart devices. Currently, the Virtual Network Computing (VNC) [1] is one of the most popular Remote User Interface (RUI) solutions. It relies on the Remote Frame Buffer (RFB) Protocol [2] in order to transmit the frame buffer content of a server to clients. However, it is not suitable for providing a smooth RUI solution including A/V streaming data due to insufficient network bandwidth and delay when the UI needs to be updated. Meanwhile, the RVU [3] included in the Digital Living Network Alliance (DLNA) [4] interoperability guidelines tries to resolve the problem by separating the paths for graphical UI and A/V streaming data. It uses bitmaps to transfer graphical UI, A/V stream over HTTP/DTCP-IP protocol to transmit A/V streaming data, and XML formatted graphics commands to deliver keycodes. However, the RVU is also insufficient for providing a seamless RUI with smart devices in that it does not consider media transformation. Since smart devices support different codecs and different streaming protocols, it is inadequate to provide media transformation for the seamless RUI supporting multi-screen services among these devices.

In this paper, we propose an architecture for a seamless RUI system supporting multi-screen services in smart devices. The proposed architecture provides a series of capabilities such as automatic device discovery, remote UI sharing, remote event processing, media transcoding, and streaming to account for the characteristics of smart devices. Moreover, it not only guarantees smooth UI transition by utilizing the separated paths for delivering graphical UI and A/V streaming data, but also supports efficient collaboration among smart devices including real-time media transcoding and streaming so as to accommodate various smart devices.

II. THE SEAMLEASS RUI SYSTEM

Fig. 1 shows the system architecture of the seamless RUI system supporting multi-screen services consisting of servers and smart devices in home networks.

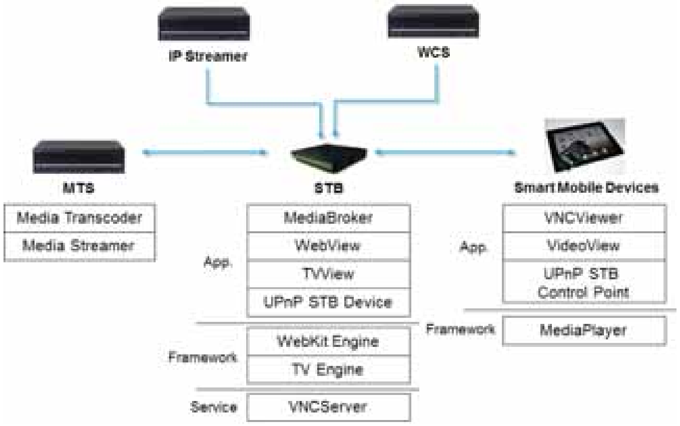


Fig. 1. System architecture

The IP Streamer transmits MPEG-2 transport streams with H.264/AAC formats via IP multicast protocol and the Web Content Server (WCS) delivers web-based UI contents coupled with A/V streaming data. The Media Transcoder and Streamer (MTS) performs media transcoding according to a transcoding request message and transmits converted media streaming data to smart devices.

The Set-top Box (STB) takes a role of a bridge device for brokering control messages between the MTS and Smart Mobile Devices (SMDs). The WebView renders web-based UI contents based on the WebKit Engine. The TVView receives MPEG-2 transport streams via IP multicast protocol, demuxes them, and displays A/V streaming data with the help of the TV Engine. The UPnP STB device is a logical device for device discovery and communication with SMDs and the VNCServer works as a service to provide the RUI. When an SMD connects to the STB, the MediaBroker requests a media transcoding about the current channel to the MTS and transfers access information about converted media stream to the SMD.

The SMDs have the VNCViewer which handles the remote UI rendering and user input events while VideoView deals with A/V streaming data. In addition, UPnP STB Control Point discovers and controls the UPnP STB device. The MediaPlayer controls playback of A/V data as part of the Android framework.

This work was supported by the IT R&D program of MKE/KEIT, [10039202, Development of SmartTV Device Collaborated Open Middleware and Remote User Interface Technology for N-Screen Service].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. IMPLEMENTATION AND EVALUATION

In order to demonstrate the feasibility of the seamless RUI system for smart devices in home networks, we apply the VNC's RFB protocol to smart devices to handle remote UI updates and remote events. In addition, UPnP protocol is installed for device discovery, control actions, and events. Moreover, we utilize the FFmpeg [5] to transform media formats in real-time and the Real-Time Streaming Protocol (RTSP) to transmit converted media, respectively.

We compose real-time IPTV services using HTML5-based UI contents coupled with A/V streaming data. The IP Streamer transmits 1080i MPEG-2 transport streams with H.264/AAC formats to 6.5 Mbps / 59.94 fps via IP multicast protocol. The MTS transcodes them into 480p streams with MPEG-4 A/V formats and transfers converted streams to 3Mbps / 29.97 fps using RTSP protocol.

Fig. 2 shows a sequence diagram for the seamless RUI supporting multi-screen services between servers and smart devices about real-time IPTV services.

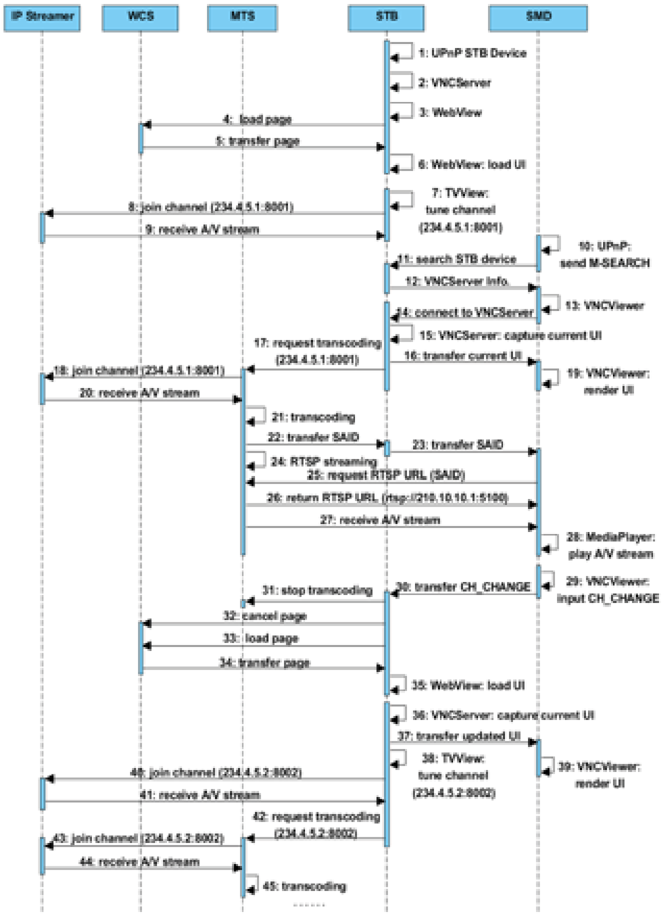


Fig. 2. Sequence diagram

An SMD searches the UPnP STB device by sending M-SEARCH message to 239.255.255.250:1900 and then connects to the STB. The VNCServer transfers the current UI and the VNCViewer renders received UI. Meanwhile, the STB requests a media transcoding to the MTS about the current channel. The MTS receives multicast A/V streaming data and transcodes them. After that, it transfers a service access identifier (SAID) related to RTSP streaming. The SMD makes an inquiry of the RTSP streaming URL using the SAID with HTTP protocol and obtains RTSP URL from HTTP redirection message. Finally, it plays A/V streaming data.

Fig. 3 illustrates the test-bed and a screenshot for the seamless RUI supporting multi-screen services using smart devices in home networks.

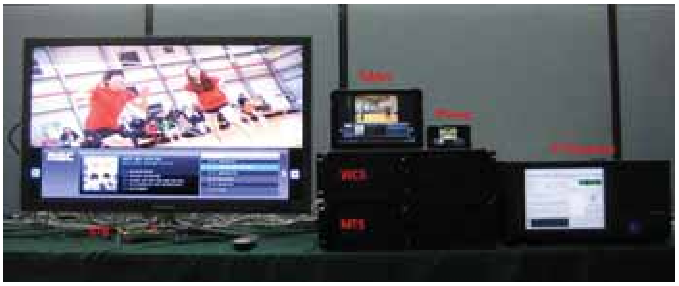


Fig. 3. Test-bed and screenshot

We built a test-bed for the seamless RUI about real-time IPTV services including SMDs and three kinds of servers such as IP Streamer, WCS, and MTS. The STB has a 1GHz dual-core processor based on Android 4.0. Likewise, an Android 4.0 smartphone with a 1GHz dual-core processor and an Android 3.1 tablet with 1GHz dual-core processor are used as SMDs. In the experiment, SMDs have about 6 seconds of latency to play transformed A/V streaming data due to real-time media transcoding and RTSP streaming. However, our approach shows it is effective to provide multi-screen services as well as device collaboration among smart devices in home networks.

IV. CONCLUSIONS

We have presented the architecture for the seamless RUI system supporting multi-screen services in smart devices based on the Android platform in home networks. The system guarantees smooth UI transition through not only the separated paths but also collaboration among smart devices. Moreover, it supports real-time media transcoding and streaming for the seamless RUI. Future works will include performance optimization to reduce the latency and development of an improved architecture including abstraction layer to accommodate heterogeneous platforms.

REFERENCES

[1] T. Richardson, Q. Stafford-Fraser, K. Wood, and A. Hopper, "Virtual Network Computing," IEEE Internet Computing, Vol.2, No.1, pp.33-38, Jan.-Feb. 1998.

[2] T. Richardson and J. Levine, "The Remote Framebuffer Protocol," IETF RFC 6143, Mar. 2011.

[3] RVU Alliance, http://www.rvualliance.org.

[4] Digital Living Network Alliance, http://www.dlna.org.

[5] FFmpeg, http://ffmpeg.org.

Background Display for Visually Impaired People in Mobile Touch Devices

Heesook Shin, Jeong-Mook Lim, Jong-uk Lee and Ki-Uk Kyung
Electronics and Telecommunications Research Institute, Daejeon, Korea

Abstract—**Existing visual representation and control restricts the access of visually impaired people to information in touch screen devices, but a flexible touch interface and multimodal representation can provide easy accessibility to blind users. To improve the accessibility of touch screen devices for visually impaired people, we propose a new approach to information representation and interaction using the concepts of 'background display' and 'foreground display'. Background display is defined as indirect representation elements, such as color, size, and location. Foreground display is defined as direct representation elements, such as text. We use auditory and tactile modality to present background display, and auditory display is used for foreground display. We found that blindfolded users understood the information on buttons more effectively and performed touch input tasks more quickly when background display was offered.**

I. INTRODUCTION & RELATED WORKS

Touch screens are a common interface element of mobile devices such as smart phones. Varied and flexible interaction on touch devices without physical buttons has many advantages for sighted users, but visually impaired people are restricted in accessing touch devices because touch interaction requires a lot of visual input.

In response to this limitation, there have been several studies. Some have used touch gestures to improve the input method. Slide Rule [1] uses a multi-touch gesture interface to access complex information intuitively. Some eye-free gesture interfaces for sighted people can be adopted for blind users, such as EarPod [2]. These technologies mostly use speech-based descriptions for information delivery and audio feedback to react to gesture inputs.

Some researchers have also tried to improve the output method. To provide complex layouts and special positioning information to visually impaired users, Timbermap [3] uses a sonification interface on mobile touch devices and is an attempt to overcome the fact that speech-based descriptions are weak to provide geometric descriptions by using sound cues.

Geometric information, such as location and size is also important in presenting various data, such as web pages and maps. Therefore, in this paper, we describe a new approach to provide information about location, size, color, and shape with sound and tactile modalities to blind users of touch screen mobile devices. We use a touch gesture interface for user input and a text-to-speech (TTS) interface for output information.

II. BACKGROUND DISPLAY

Information can be represented in various forms. Like 'Text', the meaning of some information can be expressed directly. 'Location', 'Size' and Color' forms can deliver information implicitly and intuitionally. These kinds of forms can make the arrangement and accessibility of information easier and are useful in mobile devices with small screens.

We define these two ways of representing information as 'foreground display' and 'background display' in the following Fig. 1.

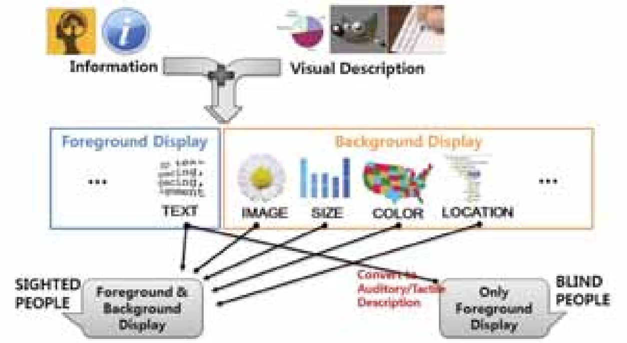


Fig 1. Representation of information by foreground display and background display

Although most information is represented by a combination of foreground display and background display as shown in the web page of Fig. 2, visually impaired people cannot access background display. Foreground display becomes the dominant means of representing information through auditory and tactile modalities such as TTS and Braille display.

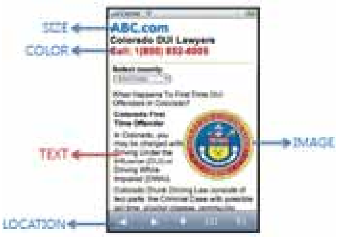


Fig 2. Web page expressed in a combination of background and foreground display

Then how can we provide 'background display' to visually impaired people? How can 'background display' be effectively provided to them in practical mobile touch devices?

To answer these questions, we propose auditory and tactile modalities to provide background display as they can be used easily and commonly in mobile touch devices. In this research, we tested the effectiveness of background display to input characters by touching buttons.

This work has been supported by the S/W Computing R&D program of MKE/KEIT(10035360-2010-01, TAXEL: Visio-haptic Display and Rendering Engine).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. EXPERIMENTS

A. Environments

We compared auditory background display, tactile background display, and non-background display. In this experiment, background display indicates the information of button location, shape and size. We used TTS technology as the foreground display. To distinguish the auditory foreground from the background display, we used different tones and different spatial (left and right) conditions of sound source. For the case of tactile modality, we generated and applied a vibration pattern using a (ACE) TACTAID VBW32 transducer [4]. Because this vibro-tactile actuator works through the sound source, we could make create the same condition of working the auditory and tactile actuator.

Fig. 3 shows the ways of expressing foreground and background display. The auditory and tactile stimulus of background display is provided when the user's finger is on the area of a button. They are generated on the left side channel while foreground auditory display is generated on the right side channel to ensure simultaneous delivery of background and foreground display.

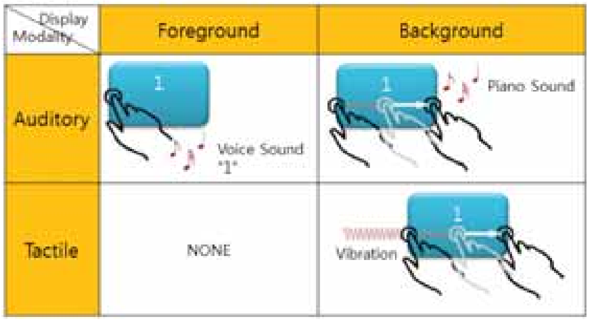


Fig 3. Auditory and Tactile stimulus of foreground and background display

A calculator program was implemented for the user experiment. It is a representative application requiring button pressing activity and has shape, size, and location information of buttons as the background display.

(a) (b)

Fig 4. (a) Implementation of experiment (b) Example of calculator program

We carried out blindfolded experiments with 6 participants (4 males, 2 females, 36.2 average age) to examine the effects of background display. We asked the participants to enter simple formulae, for example '1+2='. The tasks were carried out under three conditions, namely, without background display (WBD), auditory background display (ABD), and tactile background display (TBD). Each block has 10 formulae, and 3 blocks were operated for each condition. A total of 720 characters were input for each condition. We measured the mean time to touch one button and the average errors per trial for each condition.

B. Results

The result of the user experiment showed that the use of a background display led to rapid input speed. The mean time for tasks using the auditory background display was 4.08 seconds (SD = 0.85), and that for the tactile background display was 3.89 seconds (SD = 0.5), while the mean time for non-background display was 4.82 seconds as shown in the Fig. 5. Both kinds of background displays resulted in significantly faster button input speed than could be achieved without background display (p-value < 0.006 for ABD and p-value < 0.003 for TBD).

There was no significant difference between the auditory background display and tactile background display. The average errors per trial were 0.028 for WBD, 0.025 for ABD, and 0.028 time for TBD.

The results of user interviews and a questionnaire survey clearly showed that personal preferences for sound or vibration varied, but all participants preferred to have a background display rather than having none.

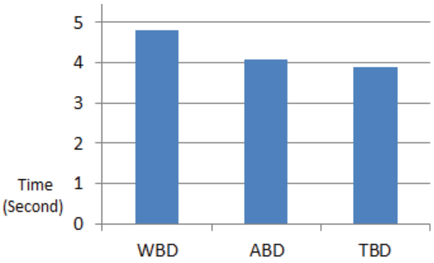


Fig 5. Mean time for each condition (without background display(WBD), auditory background display(ABD) and tactile background display(TBD))

IV. CONCLUSION

We defined concepts of 'foreground display' and 'background display' as different types of representation of information. Our experiment showed that when provided with a background display through auditory and tactile modalities in mobile touch devices, blindfolded users could understand geometric information and carry out button touch tasks more effectively than without a background display.

We plan to apply tactile background display to web page browsing application for visually impaired people and will work to develop various feelings of texture to improve tactile background display.

REFERENCES

[1] S.K. Kane, J.P. Bigham and J.O. Wobbrock, "Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques," *ASSETS'08*, pp. 73-80.

[2] S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan and P. Baudisch, "Earpod: eyes-free menu selection using touch input and reactive audio feedback," *CHI'07*, pp. 1395-1404.

[3] S.Jing, "Timbremap: Enabling the Visually-Impaired to Use Maps on Touch-Enabled Devices", *MobileHCI2010,* pp. 17-26.

[4] S. Brewster, "Tactons: structured tactile messages for non-visual information display", *AUI2004*, pp. 15-23.

Depth Boundary Reconstruction Method Using the Relations with Neighboring Pixels

Donghyun Kim, Seungchul Ryu, Sunghwan Choi and Kwanghoon Sohn
Department of Electrical and Electronics Engineering, Yonsei University,
50 Yonsei-ro, Seodaemun-gu, Seoul, 120-749, Korea
Email: khsohn@yonsei.ac.kr

***Abstract*—In this paper, we propose a depth boundary reconstruction method for effective view synthesis. To prevent the introduction of undesired depth value, the depth boundaries are reconstructed by selected pixel value through the relations with neighboring pixels. Experimental results show how the proposed method works well for view synthesis.**

I. Introduction

Recently, three-dimensional video (3DV) has received much attention as a next generation multimedia system since it provides a depth impression and realistic feeling to the viewer. With increasing the interest of consumers, viewers demand to directly control and obtain the depth impression without wearing glasses via advanced 3D displays, e.g., auto-stereoscopic displays or free viewpoint TV (FTV). These displays use multi-view plus depth video (MVD) [1] system since it can generate virtual views by view synthesis and provide 3D effect.

Depth videos are implicit in the information of distance between camera and objects. Thus, they can be used as geometry data to map the corresponding pixels between different views. Therefore, the quality of virtual views depends on the quality of supplementary depth videos as well as multiple color videos.

However, during the depth video coding process, we used the conventional video coding method, e.g., H.264/AVC [2], which uses a block-based transform and quantization of high frequency components. Consequently, blocking and ringing artifacts are introduced in sharp edges along depth boundaries in the depth video. These artifacts not only affect the quality of the depth video but also the synthesis results.

To overcome this problem, a number of researchers have been devoted to reduce the depth coding artifacts. Liu *et. al* have proposed a joint trilateral filter [3] which is based on similar characteristics. Similarly, Oh *et. al* have proposed a filter using fequency, similarity and distance [4]. However, when using this method to reconstruct the coded depth boundary, undesired depth pixel values are introduced, and it leads to geometric distortions in view synthesis process.

In this paper, we propose a depth boundary reconstruction method based on the the relation between bilateral kernel and neighboring pixels for preventing the occurrence of undesired depth pixel values. Consequently, we can reduce the coding artifacts around the depth boundaries and improve the quality of synthesized views.

The remainder of this paper is organized as follows. In section II, we explain the bilateral filter [5], which is the basis of the proposed method. Section III describes the proposed depth boundary reconstruction method. Experimental results are given in section IV. Finally, conclusions and future research directions are given in section V.

II. Bilateral Filter

Generally, the bilateral filter is used to preserve the depth boundaries since it is a non-iterative and simple method. This filter obtains the filtered result by using non-linear Gaussian filter having weights based on geometric closeness and photometric similarity. For a given position p, the filtered result is as follows:

$$D'(\mathrm{p}) = \frac{1}{W_{\mathrm{p}}} \sum_{\mathrm{q}\in\Omega} C(\mathrm{p},\mathrm{q}) \cdot S(D(\mathrm{p}), D(\mathrm{q})) \cdot D(\mathrm{q}) \qquad (1)$$

where C is the geometric closeness kernel taking the pixel locations p and q. S is the photometric similarity kernel with the corresponding depth pixel values $D(\mathrm{p})$ and $D(\mathrm{q})$. Ω denotes the set of pixels used in calculating the filtered results $D'(\mathrm{p})$ for central pixel position p. W_{p} is a normalization factor which equal to the sum of the $C \cdot S$ filter weights.

The geometric closeness and photometric similarity functions are based on Gaussian filter and computed as follows:

$$C(\mathrm{p},\mathrm{q}) = \exp(-\|\mathrm{p}-\mathrm{q}\|_2/2\sigma_c^2) \qquad (2)$$

$$S(D(\mathrm{p}), D(\mathrm{q})) = \exp(-\|D(\mathrm{p}) - D(\mathrm{q})\|_2/2\sigma_s^2) \qquad (3)$$

where σ_c and σ_s are the parameters controlling the fall-off of the weights for closeness and similarity, respectively.

III. Proposed Method

Unlike color images, most regions in depth images are smooth except at object boundaries. However, during depth map coding, the object boundaries are degraded by coding artifacts.

Although the bilateral filter is usually used to reconstruct the depth boundaries, it generates the undesired depth pixel values because this filter is based on Gaussian filter. These undesired pixel values cause geometric distortions such as gradient distortions along the object boundaries. Since the gradient distortions can be recognized as smoothing regions in 3D space, they may cause synthesis errors during view synthesis at the decoder side. Therefore, we should prevent the introduction of undesired depth pixel values when reconstructing the depth boundaries.

We first adopt the Canny method to detect color edge information which informs the locations of depth boundaries that need to be reconstructed. Then, to reconstruct a typical pixel $\mathrm{p} = (x, y)$ along the detected edge, we use the relations with neighboring pixels of p as weights, and sequentially update the depth value of p as follows:

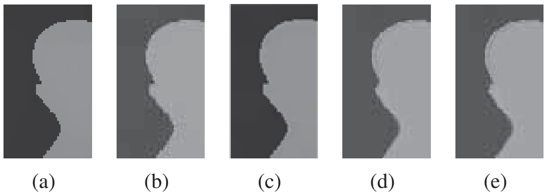
$$D'(\mathrm{p}) = \arg\min_{d\in\tilde{\mathrm{d}}_{\mathrm{p}}} \frac{\sum\limits_{\mathrm{u}\in\tilde{\mathrm{u}}_{\mathrm{p}}} \sum\limits_{\mathrm{v}\in\tilde{\mathrm{v}}_{\mathrm{p}}} G(\mathrm{u},\mathrm{v}) \cdot w(\mathrm{u},\mathrm{v},d)}{\sum\limits_{\mathrm{u}\in\tilde{\mathrm{u}}_{\mathrm{p}}} \sum\limits_{\mathrm{v}\in\tilde{\mathrm{v}}_{\mathrm{p}}} G(\mathrm{u},\mathrm{v})} \qquad (4)$$

$$G(\mathrm{u},\mathrm{v}) = C(\mathrm{u},\mathrm{v}) \cdot S(D(\mathrm{u}), D(\mathrm{v})) \qquad (5)$$

$$w(\mathrm{u},\mathrm{v},d) = |D(\mathrm{u},\mathrm{v}) - d| \qquad (6)$$

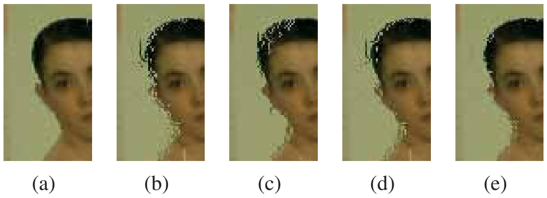
where $\tilde{\mathrm{u}}_{\mathrm{p}}$ and $\tilde{\mathrm{v}}_{\mathrm{p}}$ denote the set of horizontally and vertically neighboring pixel values of p in $(\mathrm{u},\mathrm{v}) \in W$, respectively. $\tilde{\mathrm{d}}_{\mathrm{p}}$ is a set of immediately adjacent four pixel values of p as $\tilde{\mathrm{d}}_{\mathrm{p}} = \{D(x-1,y), D(x,y-1), D(x+1,y), D(x,y+1)\}$.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a) (b) (c) (d) (e)

Fig. 1. Reconstructing results comparison, view 1, 1st frame (cropped): (a) original depth image, (b) encoded depth image by HM 5.0 with QP 34, (c) reconstructed depth image by the bilateral filter [5], (d) the joint trilateral filter [3], and (e) the proposed method.



(a) (b) (c) (d) (e)

Fig. 2. Synthesized view quality comparison, view 2, 1st frame (cropped): (a) synthesized with uncompressed depth image, (b) encoded depth image by HM 5.0 with QP 34, and reconstructed depth images by (c) the bilateral filter [5], (d) the joint trilateral filter [3], and (e) the proposed method.

TABLE I
EXPERIMENTAL RESULTS FOR *"Ballet"*

	Depth PSNR (dB)			Synthesized PSNR (dB)		
QP	Original	[3]	Proposed	Original	[3]	Proposed
34	39.57	37.57	38.85	34.90	35.13	35.96
39	36.52	35.01	35.73	33.71	33.94	35.05
42	34.82	33.62	33.94	32.79	33.28	33.94
47	33.31	32.17	32.63	31.83	32.02	32.11

TABLE II
EXPERIMENTAL RESULTS FOR *"Breakdancers"*

	Depth PSNR (dB)			Synthesized PSNR (dB)		
QP	Original	[3]	Proposed	Original	[3]	Proposed
34	39.76	37.99	38.65	38.44	38.44	38.80
39	37.33	36.13	36.84	37.24	37.29	37.58
42	35.84	34.83	35.48	36.15	36.20	36.36
47	34.46	33.87	34.12	34.80	34.81	34.98

As in (4), we calculate each bilateral kernel of the adjacent four pixels with the horizontally or vertically neighboring pixels in W. Since the relation between the adjacent four pixels and the pixels of W is used to weight bilateral-like kernel of p, we can select a depth pixel which has the highest percentage of the neighboring pixels of p. Therefore, we update the depth value of p by using the existing depth value of the adjacent four pixels. Consequently, we can prevent the occurrence of undesired depth value on p and preserve the depth boundaries.

IV. Experimental Results

In this section, we show some experimental results of the proposed method. The proposed method is implemented on a modified version of the High Efficiency Video Coding (HEVC) test model (HM 5.0) as a post-processing. The proposed method kernel is of size 16×16 and $\sigma_c = 20$ and $\sigma_s = 60$. We have tested the proposed method for *"Breakdancers"* and *"Ballet"* sequences [6]. View 1 and 3 were selected as reference views and virtual view 2 was synthesized by using Depth Image Based Rendering (DIBR) technique [7]. Depth images are encoded by HM 5.0 with QP 34, 39, 42, and 45.

In order to show how the proposed method works well for depth coding, Fig. 1 shows the cropped depth images. Fig. 1(a) shows an original depth image and Fig. 1(b) shows an encoded depth image by using conventional HM 5.0. The results, from 1(c) to 1(e), show the corresponding depth images which is reconstructed by the previous methods and the proposed method, respectively. As compared with Fig. 1(c) and 1(d), the proposed method can obtain the sharpened depth boundaries as can be seen in Fig. 1(e).

In addition, we can show that how the proposed method affects the quality of synthesized view in Fig. 2. Fig. 2(a) shows the synthesized results by using an original color and depth image. Compared to Fig. 2(a), Fig. 2(b) shows distortions along object boundaries since it was synthesized with the compressed depth image which has severe blurring distortions along those depth boundaries. However, we can obtain the improved quality of synthesized results as can be seen from Fig. 2(c) to 2(e), since the previous methods and the proposed method reconstructed the depth boundaries as shown from Fig. 1(c) to 1(e). As compared with Fig. 2(c) and 2(d), the proposed method obtains the better synthesizing results as can be seen in Fig. 2(e).

Table I and II show the objective performance of the trilateral and the proposed method for test sequences. We evaluate the PSNR of coded depth as well as the synthesized results with respect to the original view. Although the reconstruction methods underperform at the objective performance of the coded depth image, they improve the performance of the synthesized view. Especially, the proposed method has shown the best efficiency for improving the quality of synthesized view among the depth boundary reconstruction methods.

V. Conclusion

In this paper, we propose a depth boundary reconstruction method. To inform the locations of depth boundaries, we first detect the color edge using the Canny method. Then, we use the relations with neighboring pixels as weights, and calculate the bilateral-like kernel of typical pixel along the edge. The value of typical depth pixel can be updated by the existing depth value of the adjacent four pixels. Therefore, the proposed method can prevents the introduction of undesired depth pixel values which commonly occurs when using the bilateral filter. Consequently, the proposed method preserves the depth boundaries better and provides the improved synthesized views. In the future, we plan to use this method as a loop filter for improving coding efficiency.

References

[1] A. Smolic, K. Mueller, P. Merkle, N. Atzpadin, C. Fehn, M. Mueller, O. Schreer, R. Tanger, P. Kauff, and T. Wiegand, "Multi-view video plus depth (mvd) format for advanced 3d video systems," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q*, vol. 6, p. 2127, 2007.

[2] I. Rec, "H. 264, advanced video coding for generic audiovisual services," *ITU-T Rec. H. 264-ISO/IEC 14496-10 AVC*, 2005.

[3] S. Liu, P. Lai, D. Tian, C. Gomila, and C. Chen, "Joint trilateral filtering for depth map compression," in *Proceedings of SPIE*, vol. 7744, p. 77440F, 2010.

[4] K. Oh, A. Vetro, and Y. Ho, "Depth coding using a boundary reconstruction filter for 3-d video systems," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 3, pp. 350–359, 2011.

[5] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*, pp. 839–846, IEEE, 1998.

[6] MSR 3-D Video Sequences [Online], Available: http://www.research.microsoft.com/vision/ImageBasedRealities/3DVideoDownload.

[7] C. Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," in *Proceedings of SPIE*, vol. 5291, p. 93, 2004.

Confidence Stereo Matching Using Complementary Tree Structures and Global Depth-Color Fitting

Young Ju Jeong, Jiwon Kim, Ho Young Lee, Du-sik Park

***Abstract*—In this paper, a new stereo matching algorithm is proposed that separately estimates high and low-confidence region disparities. First, a mutually complementary tree structure decides on the high-confidence region, estimating its disparity map using dynamic programming optimization. Later, a disparity fitting algorithm restores low-confidence region disparity using high-confidence disparity and the information from one view color, fulfilling global optimization. The confidence stereo matching algorithm enhances both the occlusion areas and difficult-to-estimate such as thin objects, resulting in a high-quality disparity map.**

I. Introduction

Ideal 3D TV causes viewers to experience the real three-dimensional world through displays surpassing existing 2D TV, which merely represents two-dimensional images or movies. Currently, stereoscopic 3D technology is commonly used in 3D movies, DVDs, and games.

In order to convert 3D capturing systems and displays, such as stereo input contents, to multiview displays and change the 3D depth, a structure should be constructed because the multi camera system acquires 2D contents and the 3D depth structure of the camera is rough and incorrect.

Stereo matching is a method of reconstructing 3D information from stereo images. The stereo matching method is categorized into two areas: the local algorithm and the global algorithm, differing mainly by internal disparity smoothness [1]. Smoothness is the relationship between self-pixel disparity and neighborhood pixel disparity, and is defined by the Markov Random Field [2]. The local algorithm, also called the window-based algorithm, depends only on the area within the window. Disparity smoothness can be determined using the aggregation technique, but this is not an explicit estimation step method [3].

II. Confidence stereo matching

In this paper, a high confidence stereo matching algorithm that uses different algorithms to estimate high and low-confidence regions is proposed. In order to detect and estimate high confidence regions and their disparities, mutually complementary tree structures are used for dynamic programming, covering four neighborhoods of pixel smoothness. The low-confidence regions are restored by a global disparity fitting algorithm, using the information from one color input and its high-confidence disparities. This successfully reduces disparity errors in hard-to-estimate areas by separating high and low confidence areas and using stereo images only in high confidence areas. Moreover, global disparity fitting is good at occlusion areas and hard-to-estimate areas, which are difficult to estimate owing to lack of information.

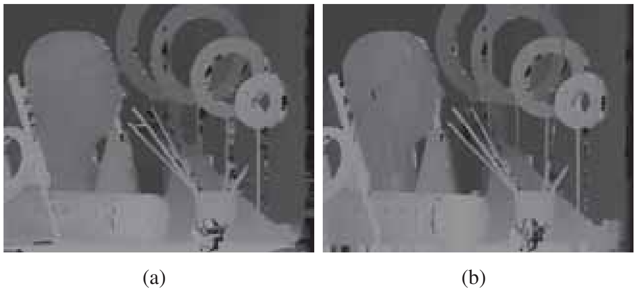

(a) (b)

Fig. 1. Mutually complementary structural disparities (a) Disparity of horizontal energy, Dh , (b) disparity of vertical energy, Dv

A. Mutual Complementary Tree Structures - High Confidence Region Detection

Mutually complementray tree structures are composed with vertical tree structure and horizontal tree structure which are introduced in Bleyer's paper [4].

Equations (1) is the energies of the horizontal tree structure. Eh is the horizontal energy. $Cv(x)$ is one of the child nodes of $p(x,y)$, which are $p(x-1,y)$ and $p(x+1,y)$. $Cv(x)$ is one of the child nodes of $p(x,y)$, which are $p(x,y-1)$ and $p(x,y+1)$. $Eh(w)(p(x,w),d)$ is the horizontal child energy of the $p(x,w)$ node when its parent node is $p(x,y)$. It does not contain $p(x,y)$ to $p(x,w)$ energy propagation.

$$\begin{aligned} Eh(p(x,y),d) =& Ed(p(x,y),d) \\ &+ \sum_{w\in cv(x)} \min_{di\in \mathrm{D}}[Ew(p(x,y),di)+Es(di,d)] \\ &+ \sum_{w\in cv(y)} \min_{di\in \mathrm{D}}[Eh(w)(p(x,w),di)+Es(di,d)]. \end{aligned} \tag{1}$$

Equations (2) is the energies of the vertical tree structure and its energy propagation is opposition with horizontal structure.

$$\begin{aligned} Ev(p(x,y),d) =& Ed(p(x,y),d) \\ &+ \sum_{w\in cv(y)} \min_{di\in \mathrm{D}}[Ew(p(x,y),di)+Es(di,d)] \\ &+ \sum_{w\in cv(x)} \min_{di\in \mathrm{D}}[Ev(w)(p(w,y),di)+Es(di,d)]. \end{aligned} \tag{2}$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

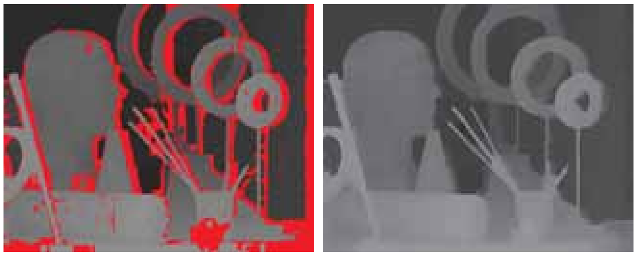

(a) (b)

Fig. 2. Disparity fitting (a) Non-red regions are high-confidence disparities and the red region is the low-confidence region. (b) Output of the disparity fitting algorithm.

Fig. 1 (a) is the disparity of horizontal energy, Dh and Fig. 1 (b) is the disparity of vertical energy, Dv . Dh has horizontal error propagation near vertical thin objects. On the other hand, Dv has vertical error propagation, even though the vertical edge is well estimated.

Only when Dh and Dv are the same, which means the mutual complementary tree structure converges to the same result, is it defined as a high-confidence region.

The left and right disparity consistency check works well at the occlusion error region. The small peak is high error disparity probability. Peak errors are removed from the high-confidence region at the last step. The small peak disparity and the region in which left disparity is not consistent with right disparity are excluded from high-confidence disparity.

B. Disparity Fitting - Low Confidence Region Manipulation

The low-confidence region is restored by the high-confidence disparity and its color information. The energy function is defined by high-confidence disparity and its color information, and its minimum solution is optimized using quadratic programming and the Lagrange multiplier.

$$E(\tilde{\mathbf{d}}) = \sum_{\tilde{d}i \in \tilde{\mathbf{d}}} \left\{ \tilde{d}i - \sum_{dj \in N(di)} \alpha ij \tilde{d}j \right\}^2 = \frac{1}{2} \tilde{\mathbf{d}}^T \mathbf{Q} \tilde{\mathbf{d}} \qquad (3)$$

Equation (3) is the new disparity energy function. $\tilde{\mathbf{d}}$ is the final disparity vector. $\tilde{d}i$ is the i-th element of the final disparity vector. N is the neighbor implemented by the 3×3 window. αij is the color similarity of the i-th color pixel, which is the same location as $\tilde{d}i$ and the j-th color pixel, which is the same location as $\tilde{d}j$.

The constraint is considered using the Lagrange multiplier such that final disparity is similar to the input disparity in the high-confidence region.

$$J(\tilde{\mathbf{d}}) = E(\tilde{\mathbf{d}}) + \lambda(\tilde{\mathbf{d}} - \mathbf{d})^T \mathbf{D} (\tilde{\mathbf{d}} - \mathbf{d}) \qquad (4)$$

Equation (4)is the final step in the solution. $\mathbf{D}$ is a diagonal matrix whose diagonal elements are 1 for constraint pixels and 0 for all other pixels. λ is the Lagrange parameter.

$$(\mathbf{Q} + \lambda \mathbf{D}) \tilde{\mathbf{d}} = \lambda \mathbf{d} \qquad (5)$$

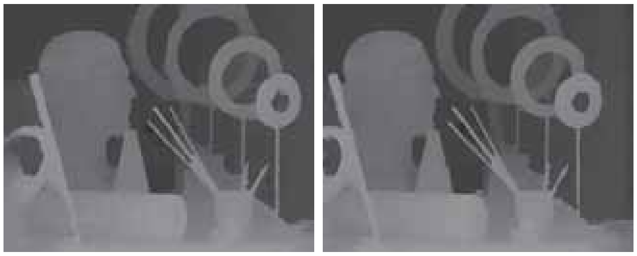

(a) (b)

Fig. 3. Final results (a) Left post-processing disparity (b) Right post-processing disparity

Since (4) is a quadratic system of $\tilde{\mathbf{d}}$, the global minimum can be found using quadratic programming. The global minimum should satisfy the zero vector by differentiating (3) by $\tilde{\mathbf{d}}$.

Equation (5) is a sparse linear system of the global minimum solution of (11) [5].

C. Post Processing

Disparity Fitting is a one view operation. The resulting disparity of disparity fitting probably has disparity error, which does not satisfy left disparity and right disparity consistency.

The boundaries of the final disparities are clear because the background and foreground are separated through left and right consistency checks.

The background region can be restored by foreground disparity in the disparity fitting step when color information is insufficient or the color is similar. This error can be corrected by left and right disparity consistency checks.

III. Conclusions

The idea of this paper has been to estimate a disparity map from a stereo image by separately solving the high-confidence and low-confidence regions. These two regions are decided using mutually complementary tree structures. A region becomes a high-confidence region only when the results of the two structures are the same. Stereo input is not used in the low-confidence region, in which only the high-confidence region disparity and its color input are used to solve for global energy minimization. The merit of this algorithm is the identification of low disparity matching error regions, because it detects high-confidence matching points.

References

[1] Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-rame stereo correspondence algorithms. (2001) 131–140 in Proc. IEEE Workshop Stereo and Multi-Baseline Vision.

[2] Li, S.: Markov random field modeling in image analysis. Springer (2001)

[3] Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. (2006) 650–656 IEEE Trans. Pattern Analysis and Machine Intelligence.

[4] Bleyer, M., Gelautz, M.: Simple but effective tree structures for dynamic programming-based stereo matching. in VISAPP **2** (2008) 415–422

[5] Levin, A., Lischinski, D., Weiss, Y.: A closed form solution to natural image matting. **30** (2008) 228–242 in Proc. IEEE Conference on Computer Vision and Pattern Recognition.

A Novel Hole Filling Method Using Image Segmentation-Based Image In-Painting

Jinkyu Hwang, Kyungjae Lee, Jaesung Kim, and Sangyoun Lee, *Member, IEEE*,
Dept. of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea

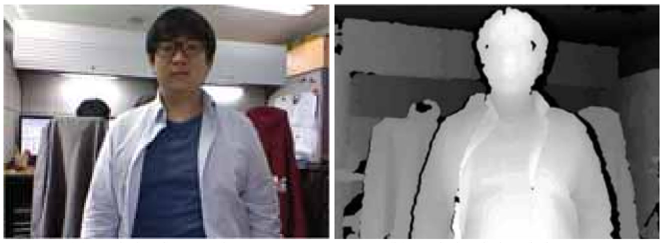
***Abstract*--In this paper, we focus on a problem that hole caused by dis-occlusion in depth image have a bad effect on stereoscopic image quality. To solve the problem, we propose a novel hole filling method using image segmentation-based image in-painting. Our hole filling method shows not only high quality result but also possibility of general-purpose method as real-time application. At the end of the paper, we generate natural stereoscopic image using DIBR.**

I. INTRODUCTION

The DIBR (Depth Image Based-Rendering) [1] is one of the methods that generate virtual stereoscopic images from monoscopic image associated per-pixel depth information. To acquire monoscopic image, the TOF (Time of Flight) sensor is generally used; however, this sensor is somewhat expensive to use commercially and has low resolution (QVGA) of a depth image. Recently, the Kinect sensor is frequently used in computer vision and graphics as an alternative to TOF sensor. The Kinect sensor can support a maximum of VGA resolution of depth and color image. In addition, it is much cheaper than TOF sensor. Figure 1 describes a color (RGB) image and a depth image acquired by the Kinect sensor.

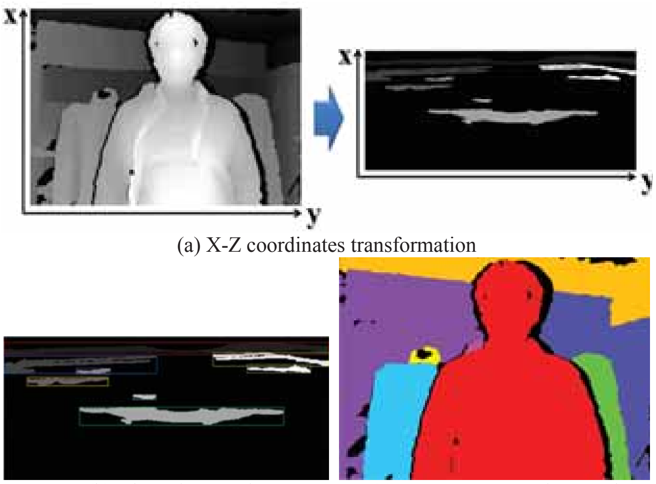
However, restoring the hole should be carried out before applying the Kinect sensor to DIBR. The hole means the regions that have no depth information. It is mostly caused by optical noise on the reflective surface and occluded areas, which are occluded in original view, can be visible in different views. This hole should be restored because it has a bad effect on stereoscopic image quality.

In this paper, we propose a novel hole filling method using the image segmentation in a depth image. The overall procedure of proposed method has two primary steps: the image segmentation in the depth image, restoring the hole in the depth image. This paper is organized as follows. In section 2, we address our image segmentation method using depth information and hole filling method using information of the image segmentation. In section 3, we evaluate proposed method. Finally, conclusions and future works are drawn in section 4.



(a) Original RGB image (b) Original depth image

Fig. 1. Original color(RGB) image and depth image acquired by Kinect



(a) X-Z coordinates transformation

(b) Clustering result (c) Image segmentation result

Fig. 2. The object Segmentation in the depth image.

II. PROPOSED METHOD FOR RESTORING HOLE REGION

A. *Image segmentation in depth image*

To segment object in the scene, we use depth information instead of color information because depth can clearly discriminate objects in spite of complex background. At first, we transform X-Y plane depth image into X-Z plane depth image by projecting all points on the depth image onto X-Z plane as shown in Figure 2(a). Next, points on transformed depth image can be clustered as SEGMENT$_{OBJ}$ by applying connected component labeling algorithm [2] after applying morphological dilation to connect neighboring valid points as described in Figure 2(b). Figure 2(c) shows the image segmentation result that expresses each segment in the X-Z plane to the X-Y plane.

B. *Hole filling in depth image*

Before restoring the depth information in the hole region, we assume that hole occurs in the background object not foreground one. To restore depth information in the hole region according to the assumption, we apply the method after modifying Telea's image in-painting approach [3-5]. As the assumption, we additionally define two regions, NOBAND and NONEIGHBOR. The NOBAND means a region that can't be a BAND region and the NONEIGHBOR means a region that can't be a neighborhood region. At first, whole hole regions are clustered as SEGMENT$_{\text{HOLE}}$ by applying connected component labeling in order to establish the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

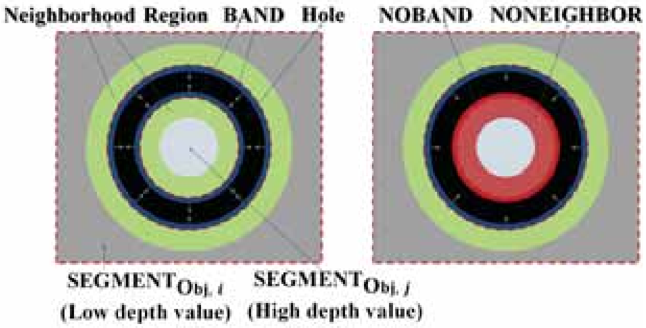


Fig. 3. Modified Telea's image in-painting algorithm for implementation

NOBAND. Because we have the image segmentation result, we can easily establish NOBAND by finding a region of segment that has maximum average of depth values in the region that overlaps with the BAND as described in Figure 3.

$$\text{NOBAND} = \text{BAND} \cap \text{SEGMENT}_{\text{OBJ},i^*}$$
$$i^* = \max_i \left(\sum_{i=1}^{n} I\left(\text{BAND} \cap \text{SEGMENT}_{\text{OBJ},i}\right)\right) \tag{1}$$

where   is the label of segment and   means the number of segment.

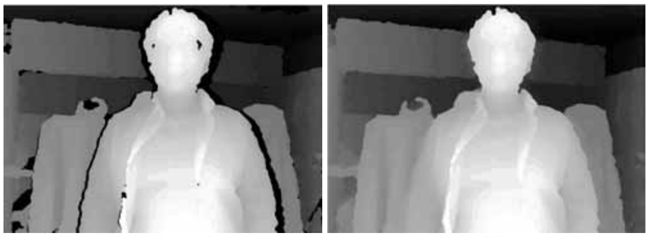
The NONEIGHBOR is established after determining . Once the label of the segment is determined, then the regions relevant to the segment are set to NONEIGHBOR. Figure 4 shows the neighborhood region, BAND, NOBAND and NONEIGHBOR regarding each $\text{SEGMENT}_{\text{HOLE}, i}$

$$\text{NONEIGHBOR} = (\text{Neighborhood Region}) \cap \text{SEGMENT}_{\text{OBJ},i^*} \tag{2}$$

In restoring procedure, each $\text{SEGMENT}_{\text{HOLE}}$ is iteratively restored. If all   are restored, whole procedure of hole-filling is finished. Figure 4 shows that the result of hole-filling in the depth image.

III. EXPERIMENTAL RESULT

To evaluate proposed method's performance, we compare stereoscopic image generation result of proposed method with linear interpolation. Figure 5 shows that proposed method's image distortion is less than linear interpolation methods Additional contribution of proposed method is that computational complexity is low (11 fps). By using this advantage, we establish stereoscopic image generation system with Kinect sensor and then we generate stereoscopic image in real time. Figure 6 shows stereoscopic image generation results, which are left, right image and synthesized image in the way of interlaced L-R display, using our system.



(a) Original depth image (b) Hole filling result

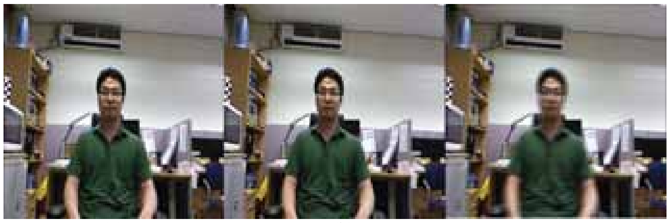
Fig. 4. Depth image hole filling result using object segmentation

(a) Original depth image (b) Hole filling result

Fig. 5. Depth image hole filling result using object segmentation

IV. CONCLUSION

In this paper, we propose a novel hole-filling method using image segmentation-based image in-painting. Our hole filling method shows not only high quality result but also possibility of general-purpose method as real-time application. Additionally, we carry out background synthesis by using proposed image segmentation method and acquire natural background synthesis result.



(a) (b) (c)

Fig. 6. Stereoscopic generation result. (a) left image (b) right image and (c) synthesized image (interlaced L-R)

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0011472).

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST)(No. 2011-0016302)

EXAMPLES OF REFERENCE STYLES

[1] C. Fhen, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV", In Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI (2004), pp. 93-104

[2] Weijie C, Maryellen L. Giger and Ulrich Bick, "A Fuzzy C-Means (FCM)-Based Approach for Computerized Segmentation of Breast Lesions in Dynamic Contrast-Enhanced MR Images". Academic radiology (Academic Radiology) 13 (1): 63–72

[3] T.Chan and J.Shen, "Mathematical Models for Local Deterministic Inpainting", In Proc. VIIP 2001, pp. 261-266, 2001

[4] Weijie C, Maryellen L. Giger and Ulrich Bick, "A Fuzzy C-Means (FCM)-Based Approach for Computerized Segmentation of Breast Lesions in Dynamic Contrast-Enhanced MR Images". Academic radiology (Academic Radiology) 13 (1): 63–72

[5] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. "Image Inpainting." In Proceedings SIGGRAPH 2000, Computer Graphics Proceedings, Annual Conference Series, edited by Kurt Akeley, pp. 417—-424, Reading, MA: Addison-Wesley, 2000

A New Hybrid Approach for View Extrapolation and Hole Filling

I. Koreshev, *Student Member, IEEE,* M. T. Pourazad, *Member, IEEE,* and P. Nasiopoulos, *Senior Member, IEEE*

***Abstract*—Generating high quality synthesized views is key factors in the successful adoption of multiview and stereoscopic 3D TV technology. One of the main issues regarding synthesized views is the occluded regions which appear as holes in the virtual views. Filling these holes is even more challenging when only one view is available and other views need to be generated via extrapolation. These holes must be filled with texture data that resembles the real texture that would exist in the real view. We present a new method for filling these holes, which allows for realistic texture data to be used, thus improving the perceived quality of the extrapolated view. Our subjective results show that our method outperforms the current state-of-the art synthesizing approach.**

I. Introduction

Three-dimensional (3D) video provides users with a more engaging and realistic impression of scenes than traditional two-dimensional (2D) video. Users can perceive depth in 3D videos the same way as they would perceive depth if they were looking at a live scene. However, a major hurdle in the proliferation of 3D display technology is the availability of 3D content. As of now the majority of available content is still 2D, and so consumers buying a stereoscopic 3D TV end up using the display less often for watching 3D content. With the development of multiview displays that provide viewers with a wider viewing angle and do not require the use of glasses for watching 3D content, the problem of lack of content becomes even more pronounced, since 3D content in the form of multiview includes several views of the scene. Multiview content production is expensive and highly demanding in terms of camera configuration and post processing.

One possible solution to the lack of multiview and stereoscopic 3D TV content is converting available 2D videos into 3D format. There has already been work done on automated depth map generation from 2D videos for 2D-to-3D video conversion purposes [1]. Yet, after estimating the depth map, the remaining challenge is synthesizing other views. Most emphasis is put on the need for synthesizing virtual views for multiview applications, which are generated by interpolating two or three real (available) views. There is, however, less discussion and research done regarding view extrapolation when only one view is available.

The main issue with the synthesizing process is related to estimating the information of occluded areas [2]. During the synthesizing process, areas of the background that were occluded by foreground objects in the 2D video become visible in the synthesized views. These areas, known as holes, must be filled with realistic data to avoid noticeable artifacts. The hole filling issue is more challenging in the case of view extrapolation compared to view interpolation, since the information of only one view is available. A well-practiced solution is to this problem is to apply interpolation to estimate the missing texture of the hole pixels. This approach has been utilized in the existing state-of-the-art view synthesis reference software (VSRS). VSRS has been selected by the ISO/IEC Moving Pictures Experts Group (MPEG) 3D Video (3DV) ad-hoc group to synthesize test sequences for future 3D video compression standardization activities [3]. VSRS uses nearest neighbor interpolation to fill the created holes during the view synthesizing process (by selecting neighboring pixels and assigning their average value to the hole pixels). The downfall of interpolation-based hole-filling methods is that the interpolated texture does not resemble the true texture of the occluded areas, but instead looks as if a clone tool was applied to those areas, in a sense that small parts of the neighboring texture are simply replicated over and over. This approach usually produces a similar looking color to the true background, but fails to reproduce texture that exists in those areas, which reduces the quality of the synthesized views and hampers the 3D effect. To avoid creation of holes in the synthesized view, a group of researchers in Disney Research Zurich have proposed to use warping to generate synthesized views from available views [4]. This method separates foreground and background objects and then by warping (stretching or compressing) the objects tries to shift them to the left or right directions (depending on which view is being synthesized) and synthesize views without creating any holes. The problem here is that due to warping, some deformities can be produced in the generated virtual views especially around foreground objects close to background objects with well-defined vertical edges.

In this study we improve the existing extrapolation-based view synthesizing techniques by using a new approach for hole filling to generate virtual views that resemble real views more closely. Our approach takes some general ideas from the existing VSRS approach and the Disney approach to fill the holes created during the extrapolation process. This hybrid approach generates higher quality views due to the hole regions being filled more accurately with matching texture.

II. Hybrid View Extrapolation

The first part of our proposed technique uses the depth and texture information similar to the existing VSRS approach to create a virtual view. This involves shifting the objects based on their depth value left or right depending on the position of the virtual camera with respect to the real camera (to be on the left side or right side). The formula that we use for this process does not require any additional parameters that are not already provided to the regular VSRS approach. The amount of shift per pixel, p_{pix} for each level of depth is calculated as follows [5]:

$$p_{pix} \approx -x_B \frac{N_{pix}}{D}\left(\frac{m}{255}\left(k_{near} + k_{far}\right) - k_{far}\right) \qquad (1)$$

where m is the depth level, D is the viewer's distance from the display, k_{near} and k_{far} are the distance of the closest and farthest object to the camera, and N_{pix} is the user defined parameter controlling the maximum parallax. The maximum parallax determines the depth of the closest object to the viewer when watching the scene on the screen. This process creates holes similar to the regular VSRS approach. We classify the generated holes into two kinds: 1) cracks (with one to two pixels width), and 2) large holes (more than two pixels width). To fill the cracks, similar to VSRS, the nearest neighbor interpolation technique is applied. For filling larger holes, first an image mask is created based on the synthesized view by setting regions with available texture data to the value of one and the areas without texture data (i.e., holes) to the value of 0. This mask is used to generate the list of warp points.

To fill up the holes, we apply warping to the background area of the image. To do this, first the warping start-points (the points where the hole-areas start with a small overlap towards the background) and the warping end-points (the points where the hole-areas end with a small overlap towards the foreground) are identified. To avoid

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

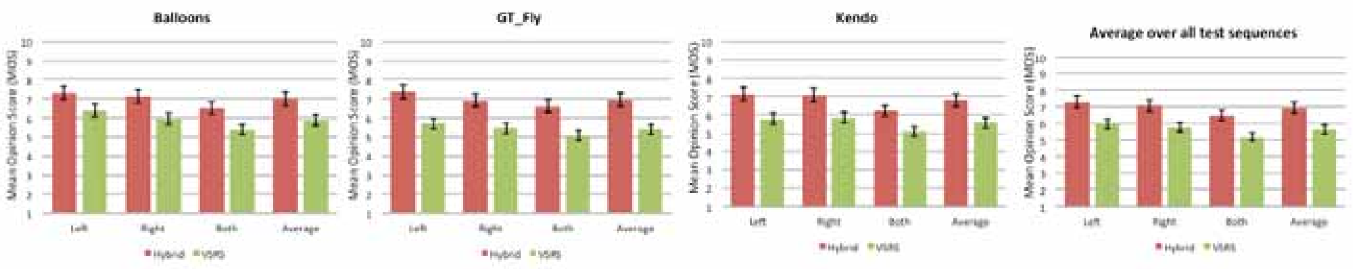


Fig. 1. Mean opinion scores for individual scenes and the average for all scenes combined. The black bar shows the 95% confidence interval.

vertical parallax, the warping process for filling the holes should be done in the horizontal direction, so the vertical coordinate of the warping start-point and end-point are equal. We also restrict the warping process to not use the information of the corner of the synthesized image. We do that because there is not enough texture data at the corners that can guarantee effective warping. Then, Piecewise cubic Hermite interpolation [6] is applied to the synthesized view so that the hole areas identified by their ending and starting points are filled by stretching the background area. Once the warped image is created, hole areas in the virtual view (which are marked in the mask image) are filled with the data from the warped image. The generated virtual view contains more realistic texture data in the hole regions than in the case of simple interpolation. Our approach is different from the Disney approach [4], since only the background area is warped. This prevents deformation of foreground objects and results in a more visually pleasant 3D effect.

III. Subjective Evaluations

The performance of our method is evaluated based on subjective tests and is compared to that of the existing VSRS package (version 3.5) [2,3]. For this evaluation we used three test sequences, namely “Balloons” (1024x768, 30fps, 300 frames), “Kendo” (1024x768, 30fps, 300 frames) and “GT_Fly” (1920x1088, 25fps, 250 frames).

The viewing conditions were set according to the ITU-R Recommendations BT.500-13 [7]. Eighteen subjects (21 to 28 years old) participated in our test. All subjects had none to marginal 3D image and video viewing experience. They all were screened for Color and visual acuity, and also for stereovision. The evaluation was performed using a 46” Full HD 3D TV with passive glasses. The TV settings were as follows: brightness: 80, contrast: 80, color: 50, R: 70, G: 45, B: 30.

At the beginning of each evaluation session, a demo sequence (“Dancer”, 1920x1088, 25fps) with different levels of synthesizing artifacts was played for the subjects to become familiar with the artifacts and the testing process. After that, the viewers were shown the synthesized stereoscopic test sequences in random order, so that they would watch two different synthesized versions of the same sequence consecutively, without knowing which video is generated by our method or VSRS. Between test videos, a ten-second gray interval was provided to allow the viewers to rate the perceptual quality of the content and relax their eyes. Here, the perceptual quality reflects whether the displayed scene looks pleasant in general. In particular, subjects were asked to rate a combination of “naturalness”, “depth impression” and “comfort” as suggested by Huynh-Thu et al. [8]. For ranking, there were 11 quality levels, 10 indicating the highest quality and 0 the lowest quality. Three test scenarios were examined: 1) right-view is synthesized, 2) left-view is synthesized, and 3) both views are synthesized. Switching the synthesized view between the right and the left eye compensated for the effect of eye dominancy (we had 8 left-eye dominant and 10 right-eye dominant subjects).

IV. Results and Discussion

The first step after collecting the experimental results was to check for outliers according to the ITU-R Recommendations BT.500-13 [7] (there were none). The mean opinion scores (MOS) were then calculated with a 95% confidence interval as shown in Fig.1. As the subjective results show, our hybrid approach scored consistently higher than VSRS. Even in the case where both views are synthesized, the MOS score for our hybrid approach is higher than that of VSRS. As Fig. 1 shows, the difference between the quality of our synthesized view and the ones generated by VSRS is higher in the case of “GT-Fly”. This is due to the high accuracy of the depth map of this computer-generated stream. Since our method relies on precise movement of objects at different depth levels, a cleaner depth map allows our method to shift only the specific objects and exclusively warp the background area and create a high quality synthesized view.

According to the subjective test results, our proposed method allows for better hole filling and view synthesis compared to VSRS. While our approach cannot replicate the true texture that is missing, it will provide more realistic looking texture by warping the background areas.

V. Conclusion

We presented a hybrid view synthesizing method, which allows for realistic texture data to be used for hole filling, thus improving the perceived quality of the extrapolated views. Our method utilizes warping to stretch the background and cover the hole areas. By filling only the hole region with the warped data, we limit the distortion associated with image warping to only those regions while maintaining the rest of the scene intact and preserving the more visually important foreground objects. Our subjective evaluations show that our method outperforms the current state-of-the-art view synthesizing method.

References

[1] M. T. Pourazad, P. Nasiopoulos and A. Bashashati, “Random Forests-Based 2D-to-3D Video Conversion”, the 17th IEEE International Conference on Electronics, Circuits, and Systems, ICECS 2010, pp. 150-153, December 2010.

[2] ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding,” N11630, Oct. 2010.

[3] ISO/IEC JTC1/SC29/WG11 “View Synthesis Software Manual,” MPEG, release 3.5, Sept. 2009.

[4] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, “Nonlinear disparity mapping for stereoscopic 3D,” ACM SIGGRAPH 2010 papers (SIGGRAPH '10), Hugues Hoppe (Ed.). ACM, New York, NY, USA, Article 75, 10 pages, 2010.

[5] ISO/IEC JTC 1/SC 29/WG 11. “Committee Draft of ISO/IEC 23002-3 Auxiliary Video Data Representations,” N8038, April 2006.

[6] F. N. Fritsch and R. E. Carlson, "Monotone Piecewise Cubic Interpolation," SIAM J. Numerical Analysis, Vol. 17, 1980, pp.238-246.

[7] ITU-R, “Methodology for the subjective assessment of the quality of television pictures,” ITU-R, Tech. Rep. BT.500-13, 2012.

[8] Q. Huynh-Thu, P. L. Callet, and M. Barkowsky, “Video quality assessment: from 2D to 3D - challenges and future trends,” in Proceedings of IEEE International Conference on Image Processing (Institute of Electrical and Electronics Engineers, New York, 2010), pp. 4025–4028.

An adaptive Error Concealment Method Selection algorithm for Multi-view Video Coding

Pei-Jun Lee, *Member, IEEE,* and Kuei-Ting Kuo, *Student Member, IEEE,*
Department of Electrical Engineering, National Chi Nan University, Puli, Nantou, Taiwan

Abstract-- **This study proposes an adaptive error concealment method selection for multi-view transmission. This study uses the neighboring block motion vector of the damaged block to define two motion correlations, motion degree and homogenous degree, which are used to select the suitable error concealment method for the damaged block construction by the proposed fuzzy reasoning. The simulation results show that the proposed algorithm can select suitable error concealment algorithm for the damaged multi-view video to reduce concealing time on homogenous damaged block and to improve the reconstructed quality on non-homogenous damaged block. In the performance evaluation, the PSNR of the proposed method have 28.11 dB to 32.4 dB.**

I. INTRODUCTION

To improve the coding efficiency of the multi-view video sequence, the Joint View Team (JVT) proposes a multi-view video coding (MVC)[1] which adopts the hierarchical B picture (HBP)[2] coding structure for the multi-view video coding. Because the spatial similarity between the interviews is higher and the movement between the successive frames is closely, HBP coding structure can obtain the good coding efficiency by the temporal and interview predictions. However, the prediction correlation of HBP coding structure easily results the serious error propagation, once a block lost in the multi-view coded bit-stream will lead not only a successive frames damaged in the current view but also the frames damaged in the neighboring views. Thus, the error concealment (EC) methods in MVC should be developed to take account of the interview and temporal correlations.

The paper [3] uses the property of the same timestamp frames on the difference views which have motion similarity. Therefore, the damaged MB can be concealed by the motion information of the corresponding MB in the reference view. Paper [4] and [5] study the performance of solutions that exploit the neighborhood spatial, temporal and inter-view information for the error concealment scope. Furthermore, different boundary distortion measurements, motion compensation refinement and temporal error concealment of Anchor frames were exploited to improve the results obtained by the basic error concealment techniques. Above studies use a unitary EC method to reconstruct the all damaged blocks in the MVC video sequence, which will increase the computation time in the MVC decoding. The EC method of [6] selects the difference EC methods, boundary matching algorithm (BMA) by candidate MVs, pixels in the blocks are interpolated using a weighted average of the neighboring pixels, and copying previous frame to conceal the slice losses, the inter-coded frame losses, and frame losses, respectively. Since to take account of the predictions correlation of HBP coding structure, such as the interview and temporal correlation will increase the reconstructed quality and reduce the reconstructed complexity for different conditions of the damaged block.

In this paper, an adaptive EC method selection is proposed to determine the suitable EC method for the damaged multi-view video to reduce concealing time on homogenous damaged block and to improve the reconstructed quality on the non-homogenous damaged block. The proposed algorithm uses the neighboring blocks motion vector of the damaged block to define motion degree and homogenous degree, then the fuzzy reasoning is adopted to select the suitable EC method with these two motion correlations. The simulation results show that the proposed algorithm can succeed in selecting suitable EC method by the fuzzy reasoning for the different damaged block construction.

II. PROPSED ALGORITHM

To diminish the error propagation in multi-view video transmission, taking account of the interview and temporal correlations is necessary for the suitable EC method selection. Based on the correlation between the motion and the disparity from neighboring cameras in multi-view video generation, the high motion correlation exists between inter-view and intra-view. The block motion correlation of encoded frames of the neighboring views (frames) from the same instant (view) can be used to select suitable EC methods to conceal the damaged block. In this paper, the motion correlations of the current block indicate the regional similarities between adjacent frames (views) by motion degree, $Mean_c$, and the homogenous degree, Var_c, respectively.

$$Mean_C = \frac{1}{8}\sum_{k=1}^{8} MV_k \text{ , } Var_C = \sqrt{\frac{1}{8}\sum_{k=1}^{8}(MV_k - Mean_C)}$$

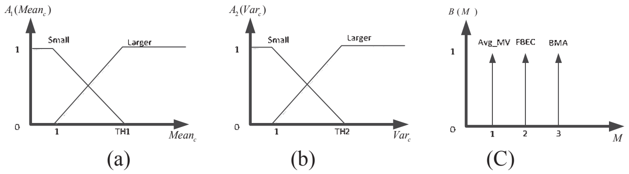
where MV_k, $k = 1, 2, 3, 4, 5, 6, 7, 8$ denotes the MV of the top-left, top-right, bottom-left, bottom-right, left-top, left-bottom, right-top, and the right-bottom block of damaged MB, respectively.

When the homogenous degree and motion degree are small, the motions of the neighboring blocks have similar motion degree. Because this region has unity motion, the EC method can use the lower complexity EC method to keep the reconstructive quality and to reduce concealment time. In this study, we adopt the compensated pixels from the reference frame by using the average MV of the neighboring block (Avg_MV) to conceal the damaged block. When the motion degree and the homogenous degree are large, which indicate the damaged block belongs to the high motion region. To obtain the higher reconstructed quality, the proposed algorithm uses motion vectors of the neighboring blocks as a possible candidate for the motion vector of the lost block. Among the possible candidates motion vector, BMA is applied

978-1-4673-1363-6/13/$31.00 ©2013 Crown

to find the block that smoothen the boundary between the lost block and the neighbors. In the other case, when motion degree is low and the homogenous degree is high, the motion correlation of the neighboring block is un-reliable. To solve this problem, the proposed algorithm uses the feature-based EC (FBEC) [7] to conceal the damaged block. This EC method estimate the relationship of the feature points in the neighboring pixels of damaged block, and the corresponding feature points in the reference frames (views) by an affine model. Therefore, the pixels of the damaged MB are concealed by this model from the reference frames (views).

In this study, we adopt the fuzzy reasoning to select the suitable EC method for the damaged block concealment. In the fuzzy reasoning, the antecedent memberships are the above two motion correlations (the motion degree and the homogeneous degree) as shown in Fig.1 (a) and (b). The consequent membership is three EC methods, Avg_MV, FBEC, and BMA, respectively, as shown in Fig.1(c). The fuzzy rule sets up by this antecedent and consequent memberships. The “Minimum Inference Engine” generates the output fuzzy set and the center average defuzzification method gets the suitable EC method. Figure 2 is the flowchart of the proposed EC method selection algorithm for MVC.



(a) (b) (C)

Fig. 1. The membership functions of (a) the antecedent of motion degree, (b) the antecedent of homogeneity degree, (c) the consequent.

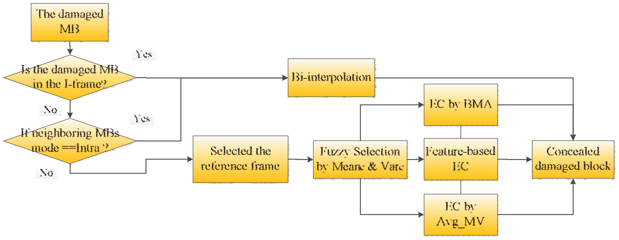

The damaged MB
Is the damaged MB in the I-frame?
Yes
No
If neighboring MBs mode ==Intra ?
Yes
No
Bi-interpolation
Selected the reference frame
Fuzzy Selection by Meanc & Varc
EC by BMA
Feature-based EC
EC by Avg_MV
Concealed damaged block

Fig. 2. The flowchart of the proposed EC method selection algorithm.

III. SIMULATION

This section shows that the reconstructed results of the proposed EC method selection are applied to the multi-view video sequences. The test sequences are Ballroom and Race 1. The quantization parameter (QP) is 37. The check error patterns with lost rate of 3%, 5%, 10%, and 20% are used in the experiment. Figure 3 shows the reconstructed result of ballroom sequence. The results are compared with error free, no EC, Avg_MV EC method, BMA EC method, FBEC method, and the proposed EC method selection algorithm. The red blocks in Fig. 3 indicate the damaged block in the faster and complex motion region. The reconstructed object in Fig3.(c) and (e) are incomplete. The result of Fig. 3(d) has apparent inaccuracy in the homologues region. The reconstructed result of the proposed EC method selection algorithm is better than other EC method.

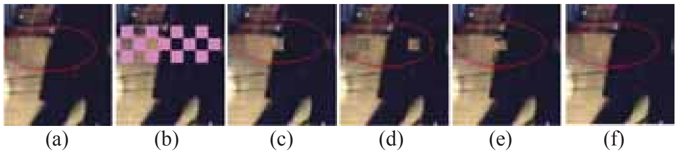

(a) (b) (c) (d) (e) (f)

Fig. 3. The 111th picture for view 1 of Ballroom sequence with 5% check lost rate by the (a) error free (b) no EC (c) Avg_MV (d) BMA (e) FBEC (f) proposed EC method selection algorithm.

Table I shows the subjective results of PSNR and computation time. It shows that the reconstructed quality of the proposed EC methods selection better than FBEC EC method 0.68 dB. And, this reconstructed time most saves 1.11 second/frame at 20% lost rate in the ballroom sequence.

TABLE I
THE RESULTS COMPARISON WITH DIFFERENCE EC METHOD IN DIFFERENCE VIDEO SEQUENCES

Ballroom : 31.74.dB								
lost rate	3%		5%		10%		20%	
	ΔPSNR(dB)	Time(ms/f)	ΔPSNR(dB)	Time(ms/f)	ΔPSNR(dB)	Time(ms/f)	ΔPSNR(dB)	Time(ms/f)
No EC	-10.71		-12.79		-15.66		-18.59	
EC_Average MV	-0.83	12.95	-1.36	13.02	-2.57	13.33	-4.76	13.79
EC_FBEC	-0.55	218.01	-1.21	370.55	-2.23	672.56	-3.37	1337.78
EC_BMA	-0.97	18.65	-1.45	20.50	-2.44	30.07	-3.97	48.66
EC_propsoed	-0.49	53.74	-0.89	63.22	-1.78	124.98	-3.63	225.18
Race1 : 33.06dB								
lost rate	3%		5%		10%		20%	
	ΔPSNR(dB)	Time(ms/f)	ΔPSNR(dB)	Time(ms/f)	ΔPSNR(dB)	Time(ms/f)	ΔPSNR(dB)	Time(ms/f)
No EC	-9.66		-11.67		-14.54		-17.53	
EC_Average MV	-0.56	12.90	-0.99	13.00	-1.99	13.45	-3.92	14.18
EC_FBEC	-0.58	167.92	-0.94	270.23	-1.81	531.79	-4.00	1034.96
EC_BMA	-1.62	18.33	-2.30	21.92	-3.47	31.68	-4.99	48.76
EC_propsoed	-0.60	82.95	-0.76	129.11	-1.72	243.09	-3.32	467.17

ACKNOWLEDGMENT

The authors thank the National Science Council of Taiwan, the Ministry of Economic Affairs of Taiwan, and AUO Corporation, under Contract No. NSC 98-2221-E-260-038-MY3, 101-EC-17-A-02-S1-201, and 1106058, respectively.

REFERENCE

[1] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” *IEEE Transactions Circuits System Video Technology,* vol. 17, no. 9, pp. 1103-1124, Sept. 2007.

[2] P. Merkle, A. Smolic, K. Muller, and T. Wiegand,” Efficient Prediction Structures for Multiview Video Coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, issue 11, pp. 1461-1473, Nov. 2007.

[3] S. J. Liu, Y. Chen, Y. K. Wang, M. Gabbouj, M. M. Hannuksela, and H. Q. Li, “Frame Loss Error Concealment For Multiview Video Coding,” *IEEE International Symposium on Circuits and Systems, 2008.*, pp.3470-3473, May 2008.

[4] L. Liang, R. Ma. P. An, and C. Liu, “An effective error concealment method used in multi-view video coding,” *International congress on Image and Signal Processing*, vol. 1, pp 76-79, Oct. 2011.

[5] B.W. Micallef, C.J. Debono, and R.A. Farrugia, “Performance of Enhanced Error Concealment Techniques in Multi-view Video Coding Systems,” *International Conference on Systems, Signals and Image processing*, pp.1-4, Jun 2011.

[6] E. Kurtepe, A. Aksay, C. Bilen, C. G. Gurler, T. Sikora, G. B. Akar, and A. M. Tekalp, ”A Standards-Based, Flexible, End-to-End Multi-View Video Streaming Architecture”, *Packer Video 2007*, pp.302-307,2007.

[7] P. J. Lee, Homer H. Chen, W. J. Wang, and L.G. Chen, ”Feature-Based Error Concealment for Object-Based Video,” *IEICE Transactions on Communications*, vol. E88-B no.6 pp.2616-2626, June 2005

Improving the Streaming Experience of Large Personal Music Libraries to Mobile Devices

Petros Belimpasakis, *Senior Member, IEEE*
Bang & Olufsen, Pullach, Germany

***Abstract*—Streaming audio content to mobile devices typically comes with buffering latencies and audio gaps, when switching tracks. We present a solution for personal music libraries, which combines song segment pre-downloading with streaming, in order to give an improved experience to users, as if the whole library is on the mobile device and immediately available. Used along with media transcoding, the solution imposes storage requirement at the client side which is only a fraction of the original music library.**

I. Introduction

The amount of digital content at home is growing at high rate, and it is estimated that advanced users host terabytes of data at their homes [1]. At the same time, cloud services allow users to store their personal content on-line, and access it from any location, whenever they need it. While advances in technology have increased the storage capacity of mobile devices, there are still not in a stage where a user could have his entire personal music library on his smart phone, or in his car's jukebox. Modern smartphones might have storage capacities of 128GBs, but the automotive industry is following in slower pace, as strict requirements (e.g. related to temperatures and shock loads), significantly increase the costs of such storage devices. Thus, modern high end cars are typically found to include hard disk drives of up to 64GBs. This storage capacity is simply not enough to accommodate the entire music library of a user, especially in high quality.

In this paper we present a solution for allowing users to "virtually" have access to their entire personal music library, hosted at home or on the cloud, via mobile devices, but without comprising the experience. We provide to the users the feeling that their entire music library is available on their mobile device, instantly playable, with a clever combination of media side-loading and on-line streaming.

II. The Problem

As there is a great mismatch between the storage capacity available at home, and the storage available at mobile devices, a user with a large music collection basically has two choices: a) either synchronize only a subset of this home music library to his mobile devices (either directly or via the cloud), or b) stream music from home [2] or cloud services [3] to his mobile device. The use case of streaming music to a mobile is gaining a lot of traction, especially as most modern mobile devices include also wireless communication capabilities.

Remotely browsing the entire personal music library, and streaming the desired songs, gives to the users the flexibility of accessing any of their songs, almost as if they were all available on their mobile devices. However, there is one usability drawback: the initial media buffering which is required before a song can start playing on the mobile device, can be "awkward" and long. This is imposing in the streaming music experience audio "gaps", especially when the user browses or skips tracks, in a relatively fast way. Imagine a user driving, and trying to find a specific song in his favourite album, by pressing multiple times the "next" button on his steering wheel controls. He would have to wait a few seconds after each button press, for the streaming song to be buffered and played, before deciding if he needs to press "next" again, or not. This is especially problematic in the driving scenario, as the user might not be able to take his attention off the road, to look at the head unit screen of the car.

So, how do we make the entire music library that a user has at a media server (e.g. at his home), seem available to his remote devices (e.g. car or smartphone), with the best possible user experience, i.e. without buffering gaps?

III. Proposed Solution

Our solution suggests a unique combination of the two previously mentioned methods (synchronizing and streaming), to provide an optimal experience for the end user. More specifically, we suggest that only part (e.g. first 10 seconds) of every media item in the user's media library is pre-downloaded/synchronized in the mobile device, along with its appropriate metadata (e.g. song title, duration, album cover, etc.). So, that means not the whole media file, but only the beginning of it, as shown in **Figure 1**.

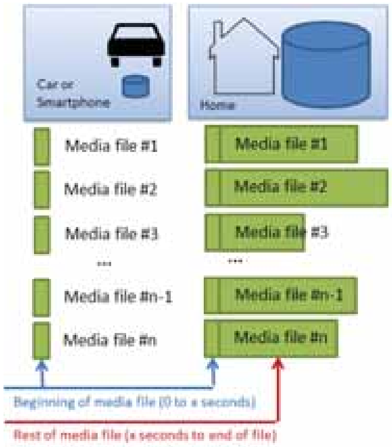


Figure 1. Illustration of media segment storage at client compared to the original media at media server (e.g. at home)

The client (car or smartphone) makes sure that it periodically communicates to the media server and gets any updates of the music library (e.g. when the car is in the garage over WiFi), for example utilizing the DLNA [4] "Content Synchronization" Guidelines. It is even possible when the car

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

is away, over a cellular connection. The client has all needed information to display and allow the user to browse the entire music library, plus listen the first 10 seconds of every song.

Combining this with the streaming option, when the user selects a song to start playing, the system starts rendering it immediately (from the local 10 second storage), but at the same time it immediately starts downloading / buffering (from the media server, via the cellular or other wireless connection), the rest of the media item from the 10-second point, to the end of the media item. The rest of the media file can be downloaded at a variable quality, based on the current network conditions (based on the 2G/3G/LTE availability, the client can request from the server the media file to be transcoded and down-sampled, on the fly). Ultimately, after the first 10 seconds of media playing have elapsed, the client seamlessly switches to rendering/playing of the newly buffered stream (from point 10-second and beyond), in such a way that the user does not notice any audio gap. This provides a unique user experience, as the entire media library appears instantly playable (without buffering gap), and at the same time minimizing the storage requirements at the client (e.g. car).

IV. THE PROTOTYPE

We implemented a full prototype of the solution, utilizing a custom made Android smart phone application as the client, and a custom made media server, on the home. The media server was able to expose Web APIs via which the content could be remotely streamed, at the clients and transcoded at the same time, on-the-fly. The logic and interaction between the components is illustrated in Figure 2.

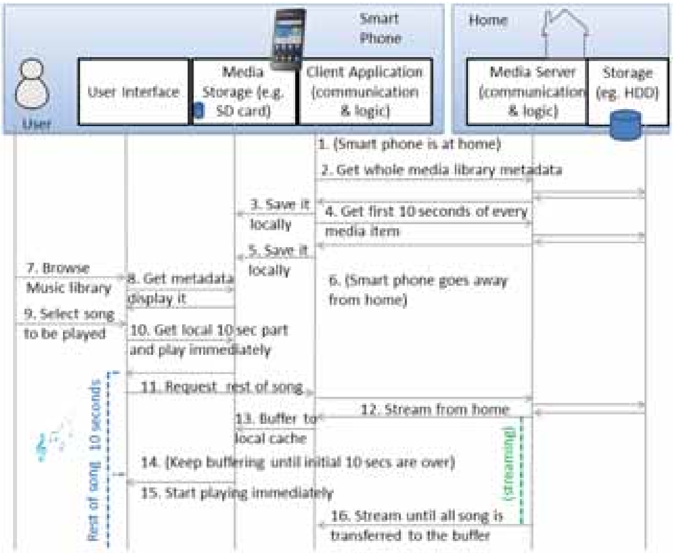

Smart Phone
Home
User
User Interface
Media Storage (e.g. SD card)
Client Application (communication & logic)
Media Server (communication & logic)
Storage (eg. HDD)
1. (Smart phone is at home)
2. Get whole media library metadata
3. Save it locally
4. Get first 10 seconds of every media item
5. Save it locally
6. (Smart phone goes away from home)
7. Browse Music library
8. Get metadata display it
9. Select song to be played
10. Get local 10 sec part and play immediately
11. Request rest of song
12. Stream from home
13. Buffer to local cache
14. (Keep buffering until initial 10 secs are over)
15. Start playing immediately
16. Stream until all song is transferred to the buffer
Rest of song 10 seconds
(streaming)

Figure 2. Interaction between components of prototype implementation

Our biggest concern of making sure that audio switch is gapless, from the pre-downloaded segment to the newly buffered song part, proved not to be a problem with our client platform. There was no audible gap for the end user, and the experience seemed totally seamless to the end user.

Calculating on the storage requirements that this solution has, we assume that average song duration is about 3.5 mins, i.e. 210 seconds. At this duration, the average file size would be 8MBs (for a 320kbps .mp3 file), or about 25 MBs (for a WMA lossless). An audio system with 1 TB of storage at home can hold a maximum of 40 000 songs, in lossless format. Obviously, storing them at a car or smartphone would require 1TB, which is unrealistic, plus cars typically do not support .WMA lossless files. Assuming that the whole library is transcoded to 320kbps MP3 files and copied to the client, would still require about 320GBs of storage, still unrealistic.

But, holding just the first 10 seconds of each one of those 40 000 songs (in 320kbps MP3 files), would require only about 15 GBs, which is a very realistic amount. Actually, even holding and buffering just 5 seconds of each song (that would "cost" 7.5 GBs of storage at the client) would be enough to provide a sufficient user experience in cases of streaming over a 3G network. Different combinations between user music library size, and the required storage at the client, for the pre-downloaded segments, are illustrated in Table 1.

Table 1. Relation between library size and client storage

User library (songs)	Storage at home (lossless)	Pre-downloaded song segment duration (sec)	Storage required at clients for segments (320 MP3)
40 000	1 TB	10	15.25 GB
30 000	0.75 TB	10	11.44 GB
20 000	0.5 TB	10	7.63 GB
10 000	0.25 TB	10	3.81 GB
40 000	1 TB	5	7.63 GB
30 000	0.75 TB	5	5.72 GB
20 000	0.5 TB	5	3.81 GB
10 000	0.25 TB	5	1.91 GB

The storage required at the client, for the transcoded 10 second song segments, is only 1.5% of the original storage required at home for the whole lossless songs. While, in the case of 5 second segments, this ratio is only 0.75%.

V. CONCLUSION

We presented a solution that allows users to stream their large personal music library, from their home or cloud service, to mobile devices (e.g. smart phone or car), while getting the feeling of zero-buffering time and thus significantly improving the experience. Combining song segment pre-downloading with transcoded streaming, the storage requirements at clients is only a fraction of the original media storage requirements.

REFERENCES

[1] T.M. Coughlin, "Personal storage for mobile applications", Journal of Magnetism and Magnetic Materials, Vol. 320, No. 22, pp. 2860-2867, 2008.

[2] P. Belimpasakis, S. Moloney, V. Stirbu, J. Costa-Requena, "Home Media Atomizer: Remote Sharing of Home Content - without Semi-trusted Proxies", IEEE Transactions on Consumer Electronics, vol.54, no.3, pp. 1114-1122, August 2008.

[3] G. Kreitz, F. Niemela, "Spotify -- Large Scale, Low Latency, P2P Music-on-Demand Streaming", In Proceedings of Tenth International Conference on Peer-to-Peer Computing (P2P), August 2010.

[4] Digital Living Network Alliance (DNLA), Available: http://www.dlna.org

A Smart Phone Peripheral with Bi-Manual Skin Stretch Haptic Feedback and User Input

Markus N. Montandon and William R. Provancher, *Member, IEEE*
Haptics and Embedded Mechatronics Laboratory, University of Utah

***Abstract*— We have developed and calibrated a bi-manual smartphone peripheral for rendering skin stretch feedback to a user's thumbs. The device's compact skin stretch displays are capable of providing repeatable haptic cues to a variety of users. Results from single- and dual-handed direction identification tests for judging 16 equally distributed direction cues show significantly better performance when the user is simultaneously provided with a direction cue on both thumbs.**

I. Introduction

The growing markets of smartphones and tablet computers have stimulated advancements in visual and auditory displays for these portable devices. The improvements of these interfaces allow for more engaging user experiences with applications and games of increasing complexity. However, minimal progress has been made in improving communication through a user's sense of touch. Haptics in handheld electronics has been primarily relegated to vibration-based feedback. While these vibrations can be used effectively, the information these vibrations communicate can be limited.

By providing a higher fidelity haptic interface it is possible to communicate a wider variety of unique and distinguishable cues. These cues can be used to supplement the sounds and visual effects of games or menu navigation and can lower the cognitive load of the user. This feedback can also allow for new experiences in which users can tactilely interact with the sensor data available on smart devices.

Application of vibrations and other haptic stimuli have been used as aids for a diverse range of human tasks. It has been observed that a vibrotactile actuator providing feedback to accompany the use of a touch screen keyboard improves accuracy and results in a strong user preference for the tactile feedback [1]. Wearable devices have been tested to create more dynamic haptic cues that simulate human physical contact such as twisting, squeezing or dragging a finger across the wrist [2]. Rotational stretching of the skin at the forearm has been used to successfully direct a user to match rotations at a control knob [3]. Lateral skin stretch applied to the fingertip has been shown to be a reliable method for communicating cardinal direction cues (i.e., North, South, East, and West) [4].

Fingertip skin stretch involves the user placing their fingerpad against a rounded 7 mm high-friction contactor (or tactor). A 16 mm round opening, also referred to as an aperture, is used to restrain the lateral movement of the fingerpad [5]. The aperture aids in holding the finger in place while the tactor moves laterally to induce stretch to the fingerpad skin.

We have developed a smartphone peripheral that is capable of delivering responsive and repeatable lateral skin stretch to both of the user's thumbs (Fig. 1). Each tactor includes a force sensor and tactile button to allow the user to provide dual analog input and button selections at the same contact point. This peripheral communicates wirelessly over Bluetooth with a smartphone running the Android operating system.

II. Device Description

A. Hardware

The two tactile displays used in our smartphone peripheral (Fig. 1) to provide planar skin stretch have improved characteristics and additional features from previous designs [6]. By orienting the display's RC servomotors flat on their sides a lower profile is achieved, to be more compatible with a flat smartphone. Each servo's motions are transmitted to a sliding plate, which is constrained from above and below, through spring steel wires. These wires maintain a rigid linkage along their respective axes while allowing lateral deflection when the servo on the orthogonal axis is actuated. The display's tactor is mounted atop a force sensor which is mounting to the sliding plate. The tactor has a planar workspace of 6.2 mm x 6.2 mm square. The skin stretch display housing is 45.9 mm x 67.7 mm x 18.5 mm.

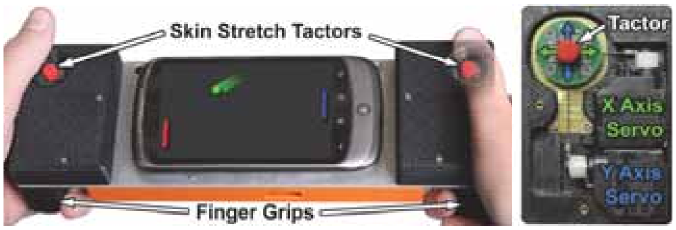


Fig. 1. Bi-manual skin stretch peripheral provides skin stretch feedback while allowing thumb-based input [left]. Internal actuation mechanism of a single skin stretch display [right].

B. Supporting Electronics

Within each tactor is a force sensor that communicates lateral forces induced by the user for device input. The user may also press down on the tactor to depress a tactile button mounted below each tactor. Inputs from both of these sensors are processed by a custom printed circuit board (PCB). The onboard microcontroller (MCU) relays these inputs to the smartphone through an RN-42 serial Bluetooth modem. The MCU also manages tactor positioning updates.

C. Control Software and Method

Software on the Android smartphone runs as a haptic service to manage skin stretch cues and user input. This service can interact with other Android software to enable haptic feedback

This work was supported, in part, by the US National Science Foundation under awards IIS-0746914 and IIS-0904456.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

in applications (e.g., games) or to enhance the user interface.

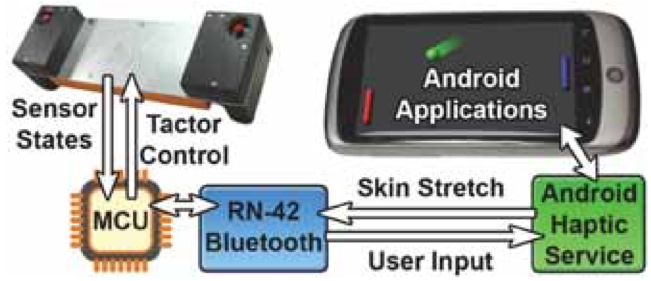


Fig. 2. Schematic of peripheral’s hardware and software components.

Each of the tactile displays on our peripheral is capable of rendering direction cues at any arbitrary angle; however, cues were calibrated for 16 unique directions that stretch the skin radially outward from the center position, spaced every 22.5°. The number of cues chosen was based to slightly exceed the angular accuracy results from previous studies [7], [8].

Initially skin stretch direction cues were specified using a single endpoint to specify the motion of the two servos. The motion of the tactor that results from these commands can be seen in Fig. 4(center), whose paths were recorded using a pair of orthogonal US Digital linear encoder probes (part # PE-500-2-I-S-L), which have a resolution of 12.5 μm.

To compensate for the nonlinear kinematics of our tactile displays, control the velocity, and improve upon the angular accuracy of our directional skin stretch cues these cues were specified as a trajectory of waypoints with 100 Hz update cycles. Twenty three waypoints were specified for each direction cue to achieve the tactor motions shown in Fig. 4(right). To characterize the accuracy of these cues, 1600 repetitions were rendered using the device to 10 subject’s thumbs while tactor position data were logged using a set of linear encoders [9]. This showed that direction cues are rendered with an average distance of 1.55 mm and absolute angular error of 1.67 degrees across all 16 cues.

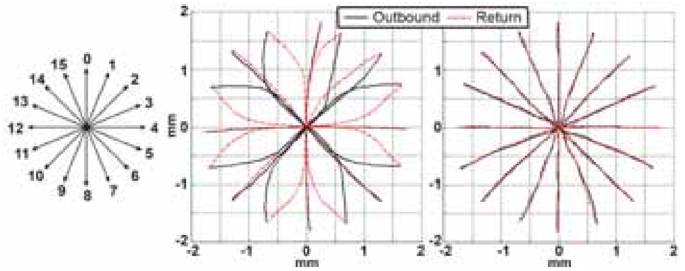


Fig. 4. Desired directional skin stretch cues [left]. Outbound and return tactor paths for initial, uncorrected direction cues that only specify the two endpoints [center] and calibrated direction cues with 23 waypoints [right].

III. Experiment

As an initial demonstration of device utility, we conducted a simple identification study with 12 participants (6 male, 6 female, average age of 32.8 years) using our smartphone peripheral. Each participant was given skin stretch cues in 16 directions (per Fig. 4(right)), with each direction repeated 10 times, and delivered in a random order. This was done in order to test the feasibility of providing more subtle direction cues for unstructured navigation, than tested in prior work [4]. Skin stretch cues were composed of a radial outbound tactor movement of approximately 1.55 mm at 30 mm/s, a 0.3 s pause, and a return to the center position at 15 mm/s.

Skin stretch direction cues were delivered to the right thumb only in one experiment and were delivered to both of the thumbs of each participant in a second experiment. The participants’ held the device with both of their thumbs pointed straight forward on the device. Upon perceiving a direction the participant responded verbally with the corresponding numerical value, as shown in Fig. 4(left). Identification accuracy for cues delivered to the right thumb resulted in an average accuracy of 29%, while the two thumb configuration resulted in an average of 49% correct answers. This difference in accuracy was found to be statistically significant based on a paired t-test, $t(1918) = 13.454$, $p < 0.001$, $\alpha = 0.05$.

It was observed that participants’ incorrect responses for the tests where cues were only delivered to their right-thumb were predominantly in the counter-clockwise direction from the rendered cue direction, which is consistent with prior testing [5]. This systematic bias is not observed in the experiments were cues were simultaneously delivered to both of the participants’ thumbs.

IV. Conclusions

A smartphone peripheral was developed that allows for responsive and repeatable haptic feedback to supplement or replace audio and visual communication methods on a smart phone. We have completed preliminary tests for user direction identification with the right- and dual-thumb configurations for cues in 16 evenly spaced directions (every 22.5°). Future tests will include identification tests in additional hand configurations and we will explore making ergonomic improvements to the device.

References

[1] S. Brewster, F. Chohan, and L. Brown, “Tactile feedback for mobile interactions,” *SIGCHI - CHI ’07*, pp. 159-162, 2007.

[2] A. A. Stanley and K. J. Kuchenbecker, “Design of body-grounded tactile actuators for playback of human physical contact,” *IEEE World Haptics Conference*, pp. 563-568, Jun. 2011.

[3] K. Bark, J. Wheeler, G. Lee, J. Savall, and M. Cutkosky, “A wearable skin stretch device for haptic feedback,” *IEEE World Haptics Conference*, pp. 464-469, 2009.

[4] B. T. Gleeson, S. K. Horschel, and W. R. Provancher, “Perception of Direction for Applied Tangential Skin Displacement: Effects of Speed, Displacement and Repetition,” *IEEE Transactions on Haptics*, vol. 3(3), pp. 177-188, 2010.

[5] Gleeson, B. T., Stewart, C. A., & Provancher, W. R. (2010). Improved Tactile Shear Feedback: Tactor Design and an Aperture-Based Restraint, *IEEE Transactions on Haptics*, 4(4), pp. 253–262, 2011.

[6] B. Gleeson, S. Horschel, and W. Provancher, “Design of a Fingertip-Mounted Tactile Display with Tangential Skin Displacement Feedback,” *IEEE Transactions on Haptics*, vol. 3(4), pp. 297-301, 2010.

[7] M. Salada, P. Vishton, D. Ph, and J. E. Colgate, “Two Experiments on the Perception of Slip at the Fingertip,” *HAPTICS '04*. pp. 146-153, 2004.

[8] K. Drewing, R. Zopf, M. O. Ernst, and M. Buss, “First Evaluation of A Novel Tactile Display Exerting Shear Force via Lateral Displacement” *ACM Transactions on Applied Perception*, vol. 2(2), pp. 118-131, 2005.

[9] M. Montandon, “Design, testing and implementation of a smartphone peripheral for bi-manual skin stretch feedback and user input,” M.S. thesis, Mech. Eng., Univ. of Utah., Salt Lake City, UT, 2012

Compacted Codeword Huffman Decoding Method for MPEG-2 AAC Decoder

Eun-Seo Lee, Jae-Sik Lee, Kyou-Jung Son and Tae-Gyu Chang, *Senior Member, IEEE*

***Abstract*-This paper proposes a new MPEG-2 AAC Huffman decoding algorithm which is designed to find multiple symbols in a single search. The analysis and experimental results show that the computational complexity of the proposed method is lower by more than 46% when compared with those of the up-to-date methods.**

I. INTRODUCTION

Huffman decoder is known to be one of the major processing blocks which occupies about 30% of overall computational complexity of an MPEG-2 AAC decoder [1]. This paper proposes a new Huffman decoding method for the purpose to improve the implementation efficiency of the MPEG-2 AAC decoder.

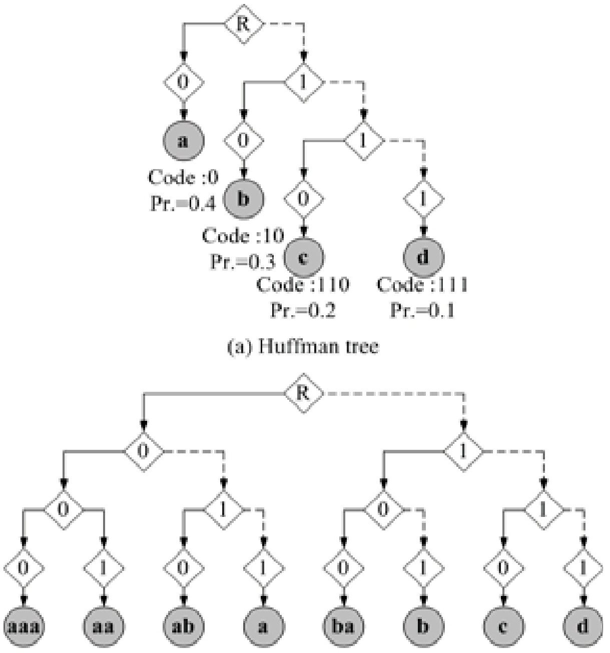
The proposed algorithm uses a reconstructed Huffman table in a form of a direct look-up table so that the symbol search can be performed by direct memory reading. Moreover, the reconstructed Huffman table allows decoding of multiple symbols in a single search, resulting in significant enhancement of searching efficiency.

The memory usage and the searching efficiency are the major design parameters which directly affect the power consumption of the Huffman decoding algorithm [2-4]. The trade-off relation between the amount of memory usage and searching efficiency of the proposed algorithm is also analytically investigated. In this way, the highly skewed statistical distribution of VLC(variable length code) can be best capitalized by utilizing the trade-off relations in determining the codeword length of the Huffman table.

To verify the performance of the proposed method, it is implemented for MPEG-2 AAC decoder and its processing speed(i.e., computational complexity) is measured and compared with those of the other three conventional methods including the sequential search method, the binary tree search method, and the hybrid method [5]. The average computational complexity of the proposed algorithm tested for 63 MPEG-2 AAC files is shown as 84%, 63% and 46% less than those of the conventional methods.

Such complexity reduction of Huffman decoding algorithm can be considered as a significant contribution, especially for the purpose of low power implementation of MPEG2 AAC audio decoder.

II. Compacted Codeword Huffman Decoding Method



(a) Huffman tree

(b) Reconstructed Huffman tree

Fig. 1. An example of reconstructed Huffman tree of compacted codeword based Huffman decoder.

The conventional skewed binary tree is fully expanded to a complete binary tree as shown with an example in Fig.1. Each leaf node in Fig.1(b) corresponds to a compacted codeword having three bit wordlength. Each compacted codeword can accommodate multiple number of short Huffman codes. Then a content addressable memory is constructed to include the information of ‘number of symbols’, ‘sequence of symbols’, and ‘consumed bit length’. The reconstructed Huffman table is shown in Table. I. The accommodation of multiple symbols yields a highly condensed memory structure.

TABLE I
The Example of The Reconstructed Huffman Table

Index (Compacted codeword)	Number of symbols	Sequence of symbols	Consumed bit length
0 (0 0 0)	3	**a**, **a**, **a**	3
1 (0 0 1)	2	**a**, **a**	2
2 (0 1 0)	2	**a**, **b**	3
3 (0 1 1)	1	**a**	1
4 (1 0 0)	2	**b**, **a**	3
5 (1 0 1)	1	**b**	2
6 (1 1 0)	1	**c**	3
7 (1 1 1)	1	**d**	3

Huffman decoding can be performed by direct memory access with the incoming bits. Multiple symbols can be decoded from a single access of the memory address. After each cycle of decoding one compacted codeword, the unconsumed number of bits in the codeword is concatenated with the next input bits to make up a new compacted codeword. When the length of Huffman codeword exceeds the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

length of compacted codeword, it is exceptionally treated to decode using the sequential search method.

III. Analysis of The Computational Complexity of The Proposed Method

Decoding complexity of the proposed method is analytically derived to obtain a closed-form equation as shown in (1). It is compared with the decoding complexity of the standard binary search method as given in (2).

$$C_{proposed} = ((3L+2C) + \sum_{i=K+1}^{N} p(s_i)\cdot(3L+2C)\cdot(l(s_i)-D))/\alpha \quad (1)$$

$$C_{binary} = \sum_{i=1}^{N} p(s_i)\cdot(3L+2C)\cdot l(s_i) \quad (2)$$

where D : compacted codeword length,
N : total number of symbols in the Huffman codebook,
K : number of symbols for which the code length is shorter than D,
α : average number of symbols decoded by one search,
L, C : numbers of instructions for a load operation and a compare-and-jump operation, respectively,
$p(s_i)$, $l(s_i)$: occurrence probability and the bit length of the i-th symbol s_i, respectively.

The decoding complexity ratio, $C_{proposed}/C_{binary}$, is computed using (1) and (2). In order to obtain the relative memory usage ratio, the total required number of memory space is computed for all of eleven MPEG-2 AAC Huffman tables, and it is divided by the sum of the number of Huffman symbols in the eleven Huffman tables. For each different wordlength of compacted codewords, the decoding complexity ratio and memory usage ratio are plotted together in Fig. 2, where the trade-off relationship is well confirmed.

The result shown in Fig. 2 can provide a useful design guide to determine a proper codeword length considering the memory and processing speed of the implementation environment. As marked in the Fig. 2, in case of designing with 5-bit compacted codewords, the processing complexity of the algorithm is significantly improved to be less than 37% in comparison to that of the conventional binary search algorithm. On the other hand, the 1.2 times increase of memory usage can be considered as reasonable considering the improvement in computational complexity.

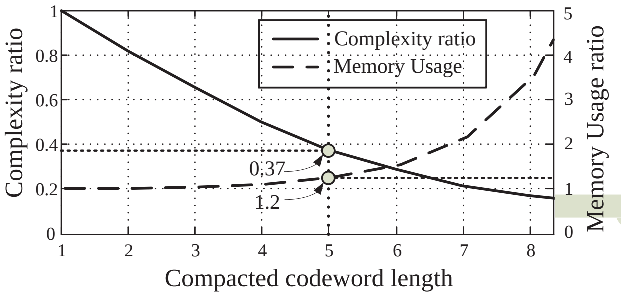


Fig. 2. Memory usage ratio and relative processing complexity ratio of the proposed method against the binary search method.

IV. Measurements and Results

The performance of the proposed Huffman decoding algorithm is evaluated by measuring the instruction cycles on the TMSC64xx DSP platform. The length of the compacted codeword is five in the design. The total of 63 MPEG-2 AAC files are tested and compared with the results of the other three conventional methods, i.e., the sequential search method, the binary tree search method, and the Hybrid search method.

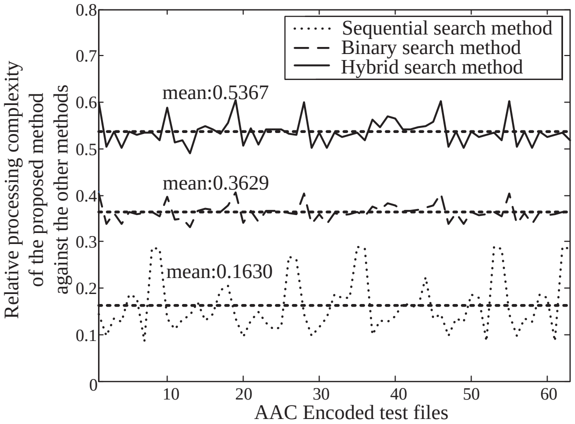


Fig. 3. Relative processing complexity of the proposed method in comparison to those of the other methods (5 bits compacted codeword length).

The results show that the average computational complexities are reduced to be 16%, 36% and 54% with respect to the three conventional methods, respectively, as shown in Fig. 3.

V. Conclusion

This paper presents the compacted codeword Huffman decoding method, which significantly reduces the computational complexity and increases the memory efficiency compared to those of the conventional methods. The trade-off relation between the searching efficiency and the memory usage of the proposed algorithm is also analytically derived. The excellence of the proposed algorithm is also verified through the experimental results performed for 63 MPEG AAC files. Where the complexity of the proposed algorithm is improved to be less than 54% even when it is compared with that of the Hybrid Huffman decoding algorithm[5], which is known as one of the most efficient ones. The proposed Huffman decoding method is expected to be utilized broadly especially for low power implementation of portable multimedia platforms.

VI. References

[1] M.A. Watson and P. Buettner, "Design and implementation of AAC decoders", IEEE Transactions on Consumer Electronics, vol. 46, no. 3, pp. 819- 824, Aug. 2000.

[2] R. Freking and K. Parhi, "Low-memory, fixed-latency Huffman encoder for unbounded-length codes," in Proc. 34th Asilomar Conf. Signals, Syst., Comput., vol. 2, pp. 1031–1034, Pacific Grove, CA, Nov. 2000.

[3] K. Chung and J. Wu, "Level-compressed Huffman decoding," IEEE Trans. Commun., vol. 47, no. 1-, pp. 1455–1457, Oct. 1999.

[4] S. Ho and P. Law, "Efficient hardware decoding method for modified Huffman code," Electron. Lett., vol. 27, no. 10, pp. 855–856, May 1991.

[5] J.S. Lee, J.H. Jeong, and T.G. Chang, "An Efficient Method of Huffman Decoding for MPEG-2 AAC and Its Performance Analysis", IEEE Transactions on Speech and Audio Processing, vol. 13, no. 6, pp 1206-1209, Nov. 2005.

An Added MDCT Harmonic Coding for the G.718-SWB Super Wideband Speech Codec

Yoon-Jin Kim, Jong-Ha Im, and In-Sung Lee, *Chungbuk Nationanl University*

***Abstract*-- In this paper, we present a MDCT domain harmonic coding method for use in the G.718-SWB standard. The proposed method resulted in a 0.9 dB SNR improvement compared to the standard sinusoidal mode.**

I. INTRODUCTION

The standardization of the super wideband speech codec has encouraged international research into speech coding algorithms. The SWB (Super-wideband) codecs have expanded from the original WB (wideband) codecs. The G.718-SWB codec is an expansion of the WB codec. The SWB is separately encoded in the MDCT transform domain [3], which is primarily used in audio coding. The audio content part of SWB is relatively new, and so a lot of redundancy still exists.

II. THE HARMONIC CODING ALFORITHM FOR G.718SWB CODEC

The G.718 wideband codec and were provided for compatibility. Embedded variable bit-rate codec, G.718 wideband codec except based on the existing wideband, 7000 ~ 14000 Hz was added to the super-wideband band [1]. Existing G.718 wideband codec consists of five tiers in the three layers by adding an additional 36 - 48 Kbit / s codec is embedded in the form of variable bit rate [2]. Generic mode and Sinusoidal mode, expansion of the band is used for the first layer. Two modes are determined for super-wideband signals that are input Tonality, improve the quality of super-wideband band extended an additional layer is coded using the Sinusoidal mode. Sinusoidal mode is also split-band super wideband signals are represented by a predetermined number of signals. Efficient harmonic component has a problem and can't encode. By segmenting the codec is focused on efficient coding. Generic mode and Sinusoidal mode other than the existing features of the audio signal is added considering Individual-Line mode and harmonic mode. The proposed harmonic coding algorithm and the original Super Wide Band coding algorithm were processed simultaneously. Added modes, including super-wideband coder is shown in Fig. 1. For convenience, represented by the dotted line was added mode. The first step in the proposed harmonic coding algorithm is to extract the fundamental frequency in order to model the harmonic structure. As can be seen in Fig. 2, the harmonic is represented as repeated pulses at every integer multiple of the fundamental frequency. Other non-pulse MDCT coefficients are mostly masked by these pulses. This means that the only the pulse MDCT coefficients are valid in the harmonic structure of the MDCT domain [3]. In the regular SWB extended speech codec, a great number of bits are needed to represent the pulse position. Therefore, imagine that we could represent the pulse positions by only using one fundamental frequency; a great many bits could then be reserved for quantization.

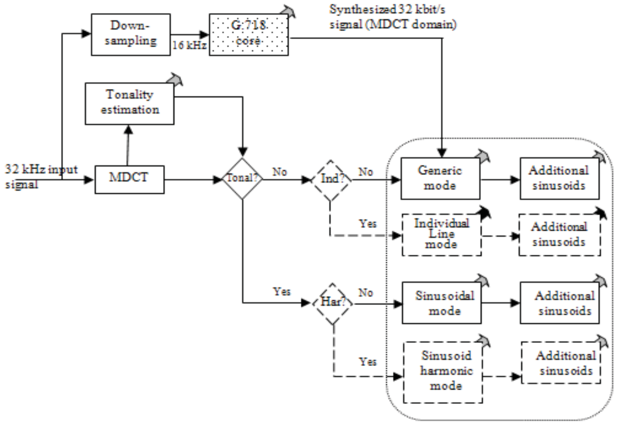


Fig. 1. The proposed harmonic coding for The G.718-SWB

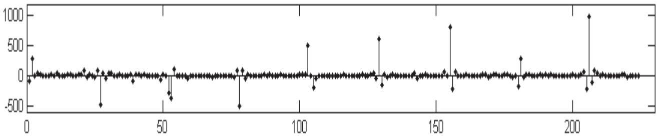


Fig. 2. The sinusoidal harmonic signal characteristics in the MDCT domain

A. Mode selection

Harmonic mode is distinguished, depending on the degree of Tonality Sinusoidal mode. Part of the signal of MDCT coefficients are divided according to the extracted energy. We obtain two set of 10 synthesis sinusoidal, one is from harmonic coding, and the other one is from original SWB codec. Error signal is calculated respectively, to choose better coding mode for current frame. 1 bit is used to mode select which has less error energy.

B. The Harmonic Track Extraction

The harmonic track is determined using the difference signal $D(k)$ from the original MDCT coefficient $M_{32}(k)$ and the synthesized MDCT coefficient $\widehat{M}_{32}(k)$ by:

$$D(k) = |\widehat{M}_{32}(k) - M_{32}(k)|, k = 280, ..., 560 \qquad (1)$$

Because Layer 6 is the first SWB layer, the synthesized MDCT Coefficient is 0.

$$P_i(m) = \sum_{n=280}^{560-m} (|M_{32}(n)| \times |M_{32}(n+m)|), m = 20, ..., 27, i = 1, 2 \qquad (2)$$

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No.2011-0013830) .

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Equation (2) is the autocorrelation function used to calculate the fundamental frequency P_i of track D_j . The possible fundamental frequency ranges from 20 to 27 in the MDCT domain.

$$PS_i(2m-1) = \sum_{n=1}^{280/P_i} |M_{32}[(2m-1)+P_i \times n]|, m = 1, ..., 16 \quad (3)$$

$$PS_i(2m) = \sum_{n=1}^{280/P_i} |M_{32}[(2m)+P_i \times n]|, m = 1, ..., 16 \quad (4)$$

Equations (3) and (4) are used for calculating the starting position of harmonic track. For one pitch component we extract the two harmonic tracks that contain the maximum energy. Position is determined by the fundamental frequency. The two harmonic tracks are determined using both the fundamental frequency and the starting position, considering the pitch feature in the MDCT domain.

C. The Harmonic Track Quantization

Four harmonic tracks will contain a maximum of 44 pulses, which contain both amplitude and sign information. 3 bits are used to extract the pulses with the maximum energy, and 8 bits are used to vector the quantization of four pulses in 2 harmonic tracks (VQ). 24 bits are used in total for the 8 out of 44 pulses in the four harmonic tracks [4]. The remaining 36 pulses are placed in one harmonic track. The DCT transform is applied to this new track. The transformed harmonic track is quantized using 19 bits, instead of the original amplitude and sign information from the harmonic tracks. A general description of the procedure is shown in TABLE I.

TABLE I
THE SINUSOIDAL HARMONIC LAYERS QUANTIZATION

Track	Highest pulse (bit)	Pulse size (bit)	Pulse sign (bit)	DCT (bit)
1	3	8	4	19
2	3		4	
3	3	8	4	
4	3		4	

III. THE PERFORMANCE EVALUATION

A. Spectrum Evaluation

The part of dotted line covering in Fig. 3 corresponding to coding part of SWB. Here we can see the G.718-SWB failed to represent all of the sinusoidal of harmonic structure. On the other hand, the sinusoidal harmonic coding method represent harmonic much better than G.718-SWB.

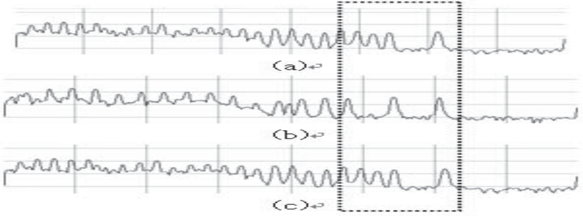


Fig. 3. The harmonic signal spectrum comparison: (a) the original sound, (b) the Sinusoidal mode, and (c) the Sinusoid harmonic mode

B. The Mushra Test

We invited twelve people to take MUSHRA test for a subjective evaluation [5]. For the evaluation we compared the modified audio listed in TABLE II; the results shown in Fig. 4 reveal that the proposed algorithm performed much better than the original G.718-SWB at the same bit rate.

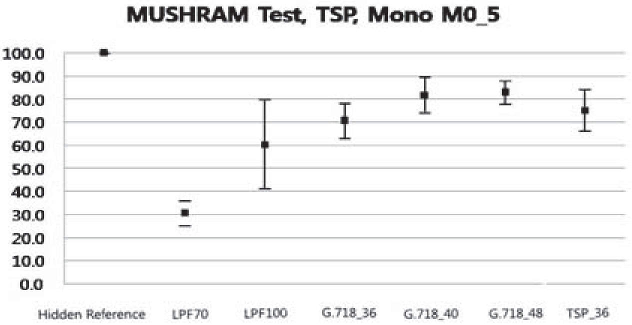


Fig. 4. The MUSHRA test results

TABLE II
THE MUSHRA TEST SAMPLES

No.	An official name	Explanation
1	Original sound	The original sound
2	LPF70	Low pass filter(fc=7kHz) Passing the original signal
3	LPF100	Low pass filter (fc=10kHz) Passing the original signal
4	G.718_36	G.718-superwideband 6 Layer
5	G.718_40	G.718- super-wideband 6 7 Layer
6	G.718_48	G.718- super-wideband 6 8 Layer
7	TSP_36	Considering the characteristics of audio signals super-wideband codec 6Layer

IV. CONCLUSION

In this paper, we present a harmonic coding algorithm for the G.718-SWB super wide band speech codec. The algorithm represents sinusoidal position information in a more efficient manner than the existing super wide band speech codec by utilizing the fundamental frequency of the harmonic structure. The proposed algorithm has been evaluated using the G.718-SWB codec as an additional coding mode. The resulted in a 0.9dB better SNR in the objective evaluation. In the subjective MUSHA test, the proposed algorithm, which used 36 bits per frame, performed in an equivalent manner to the G.718 SWB 40bit per frame codec.

REFERENCE

[1] Makinen, "AMR-WB+: a new audio coding standard for 3rd generation mobile audio services", IEEE International Conference, 2, 2005.

[2] ETRI(2009), "High-level Description of ETRI Candidate for G.722/G.711 SWB Extension", ITU-T WP3/SG16 AC-0907-Q10-08.

[3] Daudet, L., M. Sandler, "MDCT Analysis of sinusoids : exact results and applications to coding artifacts reduction", IEEE Transactions, 12, 302-312, MAY 2004.

[4] Fraunhofer IIS, "Unified speech and audio coding scheme for high quality at low bitrates", *IEEE Internation Conference,* 2009.

[5] ITU-R recommendation BS.1534, "Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)", 2001.

Adaptive Playout Scheduling[1]and Packet Loss Concealment Technique for Enhancing VoIP Speech Quality

Hyoung-Gook Kim[1], Kwangduk Seo[1], JunSeong Hong[2]
[1]Kwangwoon University and [2]Samsung Electronics

Abstract— **This paper presents an adaptive playout scheduling and packet loss concealment technique for enhancing VoIP speech quality. The proposed technique delivers high voice quality by pursuing an optimal trade-off between buffering delay and packet loss.**

I. INTRODUCTION

Voice over Internet Protocol (VoIP) has quickly become one of the fastest-growing technologies worldwide. As a result of the steady growth in VoIP usage, providing reliable services with satisfactory voice quality is now a high priority for Internet and VoIP service providers. However, excessive delay, packet loss, and high delay jitter may affect the quality of voice transmitted through an IP network [1], [2]. Therefore, improving the quality of service in IP networks is a major challenge for real-time voice communications.

In this paper, we propose a new receiver-based playout scheduling and packet loss concealment (PSLC) technique to deliver high voice quality by pursuing an optimal trade-off between average buffering delay and packet loss rate.

The proposed playout scheduling method copes with the effect of transmission jitter by compressing each packet according to the predicted network delay and variations. To recover loss packets and remove the metallic artifacts due to pitch synchronous period repetition, the linear prediction-based packet loss concealment with adaptive muting factor is implemented.

II. PROPOSED STRUCTURE OF PLAYOUT SCHEDULING AND LOSS CONCEALMENT

The proposed structure of the receiving portion of a mobile Internet phone is illustrated in Fig. 1.

The receiving system employs combined playout scheduling and packet loss concealment (PSLC) on decoded signal frames. On the receiver side of mobile VoIP, when a packet arrives at the receiver, the receiver strips the packet information and places the packet in the jitter buffer. The jitter buffer holds incoming packets, rearranges the arriving packets due to the time when the arriving packets are generated at the sending host, and releases them for decoding at a regulated speed. Next, the arriving packet information is passed on to network jitter estimation that incorporates rapid detection of delay spikes to react to changes in network conditions. The network jitter is more accurately estimated by using dynamic weighting factor of the network jitter variance, mean and variation of interarrival jitter than conventional algorithms.

This research was also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0001941).

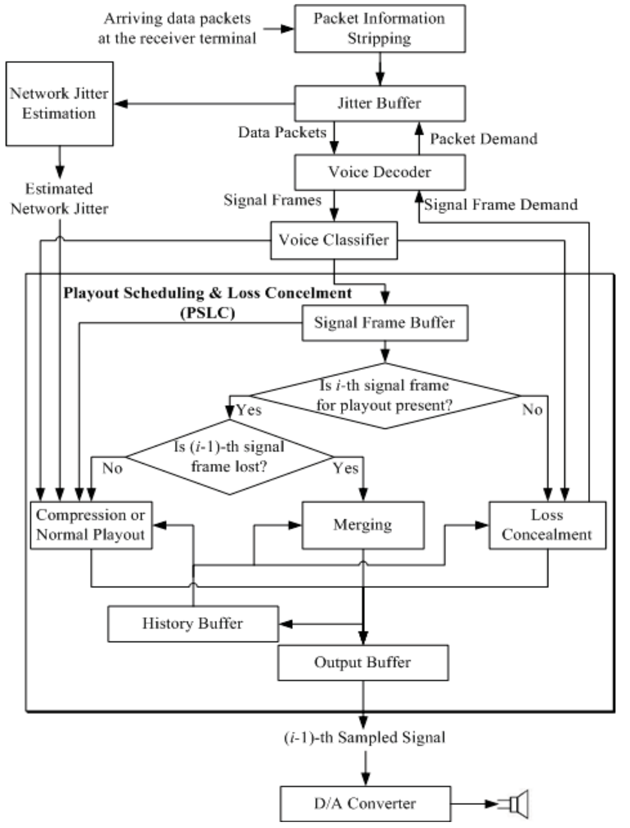

Arriving data packets at the receiver terminal
Packet Information Stripping
Network Jitter Estimation
Jitter Buffer
Data Packets
Packet Demand
Voice Decoder
Estimated Network Jitter
Signal Frames
Signal Frame Demand
Voice Classifier
Playout Scheduling & Loss Concelment (PSLC)
Signal Frame Buffer
Is *i*-th signal frame for playout present?
Yes
No
Is (*i*-1)-th signal frame lost?
No
Yes
Compression or Normal Playout
Merging
Loss Concealment
History Buffer
Output Buffer
(*i*-1)-th Sampled Signal
D/A Converter

Fig. 1. Structure of the receiving portion of a mobile Internet phone

To play out the arriving packets at a regular interval, the receiving system needs to maintain a jitter buffer, a signal frame buffer, a history buffer, and an output buffer. Each signal frame decoded from the jitter buffer is stored in a signal frame buffer. The currently used voice codecs are G.711, G.722.2, G.726, G.728, G.729AB, and G.729E.

Using normalized auto correlation and peak detection, the decoded signal frames are classified into one of five classes (silence, background, unvoiced, voiced and transient signal frames) and then input to the PSLC module, which decides on one of three processing modes: loss concealment, merging, or timing recovery using compression or normal process. The digital-to-analog (D/A) converter regularly converts the sampled signal frame from PSLC into an analog signal. Finally, the user hears the analog voice signal through a speaker.

The proposed decision logic for signal processing is one of the key contributions for maintaining a balance between conversational interactivity and speech quality and performs as follows:

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

• *Loss concealment mode*: If i^{th} signal frame is absent in the signal frame buffer, a packet is declared as lost, and “loss concealment mode” is entered. The algorithm extracts the residual signal of the previously received packet by linear prediction analysis, uses periodic replication to generate the excitation signal of missing signal frame, and generate synthesized signal frame using the excitation. To remove the metallic artifacts and improve the voice quality, the synthesized signals are multiplied by an adaptive muting factor, and gradually muted for the duration of the loss period. The adaptive muting factor is updated sample by sample. The decreasing speed depends on the signal class to reconstruct the lost frame.

• *Merging mode*: If i^{th} signal frame is present in the signal frame buffer and $(i\text{-}1)^{th}$ signal frame was lost, discontinuity between i^{th} signal frame and $(i\text{-}1)^{th}$ substituted signal frame occurs, and “merging mode” is entered. By merging, two signal frames in a transition region are smoothly interpolated to alleviate discontinuity of the transitions from a signal frame to the substitute frame or from the substitute frame to the following signal frame.

• *Timing recovery using compression or normal process mode*: If i^{th} signal frame is present in the signal frame buffer and $(i\text{-}1)^{th}$ signal frame was not lost, “timing recovery mode” is entered. The classified results of the i^{th} signal frame are then used in the timing recovery process. If the ratio *Cr* between total length of the remaining signal frames in the signal frame buffer to the active network jitter estimated from the jitter buffer is larger than the hard compression threshold, the time compression is performed in the region including signal frames classified as voiced/unvoiced/silence/background. However, *Cr* exists between soft compression threshold and hard compression threshold, the compression is performed only for the signal frame classified as silence or background.

For recovering lost packets, the PSLC module often makes a subsequent frame demand from the voice decoder. This causes the voice decoder to make a packet demand from the jitter buffer.

III. RESULTS

For the performance comparisons of total loss rate vs. average buffering delay, and average buffering delay vs. Perceptual Evaluation of Speech Quality (PESQ) [3], two network delay traces obtained from the Internet links of the testbed are used and are listed in Table I. The speech samples used for the experiments are sampled and digitized at 8 kHz. Each trace lasts for approximately five minutes, and each packet consists of 20 ms of speech content.

Fig. 2 shows the late loss rate vs. average buffering delay using different algorithms for the two traces. The performance comparison related to average buffering delay vs. PESQ is illustrated in Fig.3. M1 is based on an adaptive normalized least mean squares playout algorithm with delay spike detection [4]. In M2, a timing recovery and loss substitution method [5] is combined with modeling the statistics of the interarrival times with the K-Erlang distribution [6]. Although our simulations are based on G.711 voice codec, the algorithms are audio codec independent and can be implemented on the receiver alone.

TABLE I
STATISTICS OF NETWORK TRACES

Trace	End-to-End Network Delay (ms)	STD of Network Delay (ms)	Maximum Jitter (ms)	Network Packet Loss (%)
1	32.07	11.88	146	2
2	78.22	31.22	371	4

STD = standard deviation

The proposed technique (PM) is very suitable for operating at a low buffering delay, handling various loss patterns and maximum jitters more effectively compared to other methods.

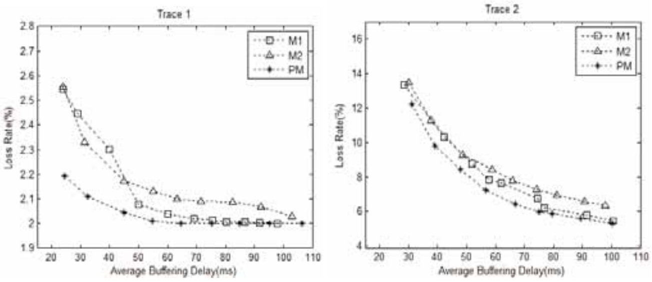


Fig. 2. Performance comparisons of total loss rate vs. average buffering delay.

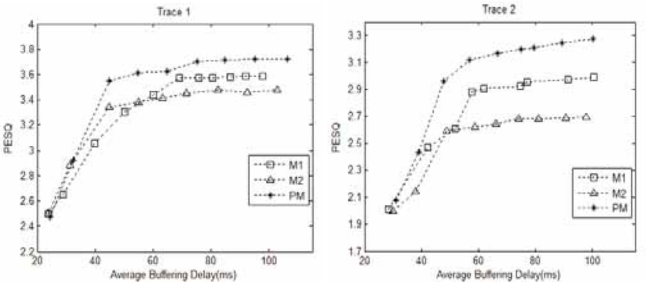


Fig. 3. Performance comparisions of PESQ score vs. average buffering delay

REFERENCE

[1] M. Gidlund, and J. Ekling, "VoIP and IPTV distribution over wireless mesh networks in indoor environment," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 1665-1671, Nov. 2008.

[2] Kuo-Kun Tseng, Yuan-Cheng Lai, and Ying-Dar Lin, "Perceptual codec and interaction aware playout algorithms and quality measurements for VoIP systems," *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 297-305, Feb. 2004.

[3] ITU-T Rec. P.862, "Perceptual Evaluation Of Speech Quality (PESQ), An Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs," *International Telecommunication Union*, Feb. 2001.

[4] A. Shallwani and P. Kabal, "An adaptive playout algorithm with delay spike detection for real-time VoIP," *IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 2, pp. 997-1000, May. 2003.

[5] S. V. Andrsen, W. B. Kleijn, and P. Sorqvist, "Method and arrangement in a communication system," *U.S. Patent 7 321 851*, 2008.

[6] H. Li, G. Zhang, and W. Kleijn, "Adaptive playout scheduling for VoIP using the K-Erlang distribution," *The 2010 European Signal Processing Conference*, pp. 1494-1498, Aug. 2010.

The Feasibility Study on the 4K-UHD Satellite Broadcasting Service in Ka-band

Min-Su SHIN*, JoonGyu RYU*, DeokGil OH* and Yong-Goo KIM†
*Electronics and Telecommunications Research Institute (ETRI), KOREA
†Korean-German Institute of Technology (KGIT), KOREA

Abstract—**The goal of the paper is to present an efficient way for providing satellite UHD broadcasting services via Ka frequency band and, for the purpose, to develop a service adaptation and stabilization technology which is adaptive to the channel conditions including rain attenuation.**

I. Introduction

Many advanced countries have developed various UHD (Ultra High Definition) broadcasting technologies, opening a new horizon on the possibility of commercial UHD broadcasting service. With private enterprises, Japanese government pushes forward trial UHD broadcasting in the year of around 2015 and expected to launch commercial UHD service in the year of about 2020. In case of Korea, the government plans experimental UHD broadcasting in the year of around 2014 through the satellite having Ka-band communication capabilities. In the initial stage of such a service, satellites using 21GHz frequency bands are expected to be the major media for that service because such frequency band is new to broadcasting services and thus no backward compatibility concerns with the existing HD services are considered. Hence, many countries are expected to be in competition with each other in the development of technologies for the various essential system technologies involved in UHD satellite broadcasting. Especially, since weather in Korea is going against the satellite broadcasting, it is very pressing to develop technologies to overcome the rain attenuation effects in Ka frequency band and to investigate the pertinence and possibility of commercial broadcasting services using Ka frequency band for immersive media including UHD video. To the end, the paper presents rain attenuation reference for UHD broadcasting via satellite using Ka frequency band in Section II, and analysis of link margin and effective bitrates for each code rate and modulation method of DVB-S2 at a certain rain rate in Section III. Next, we provide channel-adaptive UHD satellite broadcasting scenarios based on H.264/AVC SVC and DVB-S2 ACM in Section IV. Finally, in Section V, we offer a summary conclusion of this paper.

II. Rain Attenuation Model Analysis in Ka-band

To provide UHDTV service via Ka-band satellite in a reasonable way, we need to investigate rain attenuation reference over Ka-band frequency in Korea. This has been conducted by analyzing existing rain attenuation models and by expecting the rain attenuation effect of Ka-band using ITU-R P.618-5 model, which is known better for domestic Ku-band rain attenuation expectation, and ITU-R P.618-9, which is a recent update of the P.618-5 model. The estimated attenuation exceeded for p% of an average year by ITU-R methods can be determined from (1)

$$A_p = A_{0.01} \times I(p) \tag{1}$$

where $A_{0.01}$ is the estimated attenuation exceed for 0.01% of an average year and I(p) is the interpolation formula to have the attenuation value to be exceeded for other percentages of an average year. $I(p)$ is given differently in the ITU-R P.618-5 and P.618-9, respectively as follow.

$$I(p) = \begin{cases} 0.12 \times p^{-(0.54+0.043\ \log(p)} \\ \left(\frac{p}{0.01}\right)^{-(0.655+0.033\ \ln(p)-0.045\ \ln(A_{0.01}))} \end{cases} \tag{2}$$

For setting the rain attenuation reference, the operating frequency was set as the center frequency of Chollian satellites channel number 3, and selected vertical polarization and worst case between the two models. For the simulation, the following specifications of satellite data are used : Longitude = 128.2°E, Height = 35,8578km, Elevation Angle = 46.4°, Down-link Freq. = 20.13GHz, EIRP = 60dBW, and Channel Bandwidth = 100MHz.

To take into account domestic rain distribution trends, we take 66.65mm/h as the rainfall rate, $R_{0.01}$, exceeded for 0.01% of an average year which is the measured value for Seoul and a little higher than the reference value of 50.6 mm/h in ITU-R P.837.5. For calculating the slant-path length, L_s, and the horizontal projection, L_G, of the slant-path length, the rain height, h_R, is calculated using ITU-R model[3] with the location data of Seoul(126.58°E, 37.33°N) to have 3.815km. With these values, we can get the estimated attenuation due to rain for the regional use in Ka-band satellite communications as in Table I.

As mentioned above, we will use P.618-9 results with vertical polarization as the rain attenuation reference for the paper since they imposes worst-case effect on the channel.

III. Link Analysis for Ka-band Satellite

In the section, the analysis of DVB-S2 system performance has been conducted by finding the required C/N to achieve packet error rate of TS packet less than 10^{-7} for each coding and modulation technique of DVB-S2 and by evaluating the link margin and available bitrate of the Chollian satellite for each MODCOD of DVB-S2 system at each rain attenuation

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

TABLE I
THE ESTIMATED ATTENUATION IN TERMS OF POLARIZATION AND AVAILABILITY

Rec.	Pol.	99.9%	99.7%	99.5%
P.618-5	Horizontal	12.1504	7.1668	5.5215
	Vertical	11.3045	6.6679	5.1371
	Circular	11.7152	6.9101	5.3237
P.618-9	Horizontal	14.2328	7.9630	5.9159
	Vertical	12.5390	6.9756	5.1687
	Circular	13.3468	7.4458	5.5243

reference. The required C/N of DVB-S2 and corresponding available bitrate for each coding and modulation can be obtained by simple calculation and are not listed in the paper due to the limited space. The link margins of the Chollian satellite downlink which can used to define possible transmission method at certain rainfall rate, are given in Table II. For more realistic analysis, the required C/N contains the non-linear power loss under the assumption of using pre-distortion technique at earth station.

TABLE II
DOWN-LINK PERFORMANCE OF CHOLLIAN-SAT

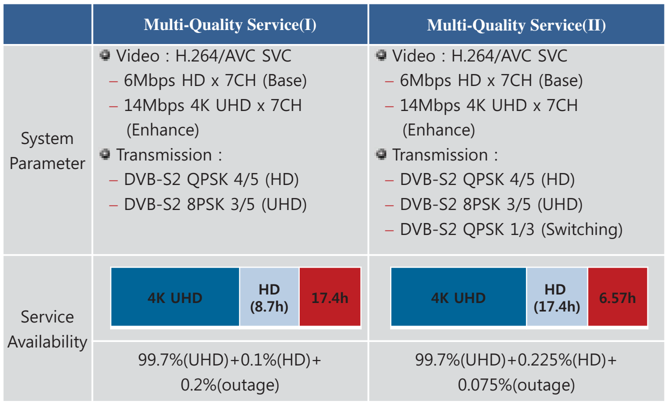
D/L (20.13GHz)	Clear	Rain (0.1%)	Rain (0.3%)	Rain (0.5%)	Rain (0.6%)
EIRP	60	60	60	60	60
Propagation Loss	-209.91	-209.91	-209.91	-209.91	-209.91
Additional Losses	-1.0	-1.0	-1.0	-1.0	-1.0
Rain Attenuation	0	-12.54	-6.98	-5.17	-4.67
Earth G/T	14.78	14.78	14.78	14.78	14.78
Noise Bandwidth	-79.21	-79.21	-79.21	-79.21	-79.21
Noise(Temp/Rain)	0	-0.36	0	0	0
Boltzman's constant	228.60	228.60	228.60	228.60	228.60
D/L C/N[dB]	13.26	0.36	6.28	8.09	8.59

*Earth G/T : Assuming that Ant. diameter = 45cm(38dBi) and Noise Temp. = 210K, Additional Losses : including all losses by atmospheric gas, clouds and mis-pointing. etc.

IV. EXPECTATION OF UHD BROADCASTING SCENARIOS

Based on the above rain attenuation reference, required C/N of DVB-S2, link margin, and effective bitrate of the Chollian satellite, we considered a few channel-adaptive UHD broadcasting test scenarios and analyzed their service performances. In this section, 3 scenarios of UHD satellite broadcasting using Ka-band have been compared : 1) multi-channel 4K UHD service based on H.264/AVC video coding at 20 Mbps, 2) the same number of channels of 4K UHD service based on the combination of H.264/AVC SVC and DVB-S2 VCM, and 3) system parameter switching when heavy rain over the country. The performance of each case in terms of service availability and service quality can be summarized as follows.

For all service scenarios, the bitrates are configured into 20Mbps and 6Mbps which are thought to be reasonable when H.264/AVC HP@5.2 is considered for one 4K UHD channel of 3840x2160@60i and one HD channel, respectively. As the simplest case, 7 channels of UHD service can be provided through the single quality scenario with 8PSK 3/5 transmission method and this results in 99.7% of service availability with 26.28 hours of outage over the year. To enhance the availability of UHD service while the same number of UHD channels are provided, more sophisticated scenarios are presented using combinational configurations of H.264/SVC and DVB VCM transmission method as in Fig. 1. In the first multi-quality scenario, the service availability can be extended into 99.8% by keeping service provision with lower quality of HD when the channel condition becomes worse by rain. And with another multi-quality scenario, the availability can be extended into above 99.9% even in the Ka-band if more resources are allocated for HD service when it rains as much as UHD service is impossible.



	Multi-Quality Service(I)	Multi-Quality Service(II)
System Parameter	● Video : H.264/AVC SVC – 6Mbps HD x 7CH (Base) – 14Mbps 4K UHD x 7CH (Enhance) ● Transmission : – DVB-S2 QPSK 4/5 (HD) – DVB-S2 8PSK 3/5 (UHD)	● Video : H.264/AVC SVC – 6Mbps HD x 7CH (Base) – 14Mbps 4K UHD x 7CH (Enhance) ● Transmission : – DVB-S2 QPSK 4/5 (HD) – DVB-S2 8PSK 3/5 (UHD) – DVB-S2 QPSK 1/3 (Switching)
Service Availability	4K UHD | HD (8.7h) | 17.4h	4K UHD | HD (17.4h) | 6.57h
	99.7%(UHD)+0.1%(HD)+ 0.2%(outage)	99.7%(UHD)+0.225%(HD)+ 0.075%(outage)

Fig. 1. Multi-Quality UHD Service Scenarios with H.264/SVC and DVB-S2 VCM via Ka-band Satellite

V. CONCLUSION

This paper presents a variety of deep analyses on the channel adaptive source-level transmission techniques for the broadcasting service of immersive media using the satellites of Ka frequency bands, providing the technological backgrounds for the verification of satellite UHD commercialization. Hence, the main results of the paper are expected to be essential for deciding the direction of future technical developments for Ka frequency band satellite UHD broadcasting services, and thereby enhancing the service quality and the economic feasibility of satellite UHD services. The results presented in the paper can be directly utilized in the trial broadcasting system provisioning for the verification of Ka band communication systems practical use in the satellite.

REFERENCES

[1] ITU-R, Propagation data and prediction methods required for design of earth-space telecommunication systems, ITU-R Rec. P.618-5, 1997.
[2] ITU-R, Propagation data and prediction methods required for design of earth-space telecommunication systems, ITU-R Rec. P.618-9, 2007.
[3] ITU-R, Rain height model for prediction methods, ITU-R Rec. P.839-3, 2001.
[4] B. Y. Kim et.al, A Study on Feasibility of Dual-Channel 3DTV Service via ATSC-M/H, *ETRI Journal*, vol. 34, no. 1, pp. 17–23, Feb. 2012.

The Design and Implementation of T-DMB Filecasting Terminal

Ji Hoon Choi and Jihun Cha

Broadcasting & Telecommunications Convergence Media Dept., ETRI

Abstract — the purpose of DMB filecasting service is to transmit multimedia content and additional information to the user through DMB data channel. In This paper, we show how to design T-DMB filecasting terminal compliant with DMB-ECG and DMB-AF specification.

I. Introduction

With the emerging of mobile IP service era, the legacy T-DMB market has entered a phase of stagnation in Korea.

Therefore AT-DMB(Advance Terrestrial-Digital Multimedia Broadcasting) system, which can support high-quality AV and high-capacity additional data, has been developed by the industry-university-institute collaboration to overcome technical limitations, i.e. low bandwidth, definition of file types, etc. And we designed and implemented T-DMB filecasting terminal compliant with DMB-ECG (Electronic Content Guide) and DMB-AF(Application Format) specification[1][2] to test and verify DMB filecasting service as a part of AT-DMB data service.

In this paper, we will explain about an implemented DMB filecasting processing module including update manager, filecasting decoder and metadata engine on the legacy T-DMB device. And then, we will examine main functions of filecasting decoder module for signaling and transmission. Finally, we will show an example of T-DMB filecasting service.

II. T-DMB Filecasting Terminal System

There are two T-DMB filecasting service methods for to transfer signalling information through T-DMB channel. Firstly, the most ideal method is to use DMB-ECG metadata to define and transfer signalling information, program title, and genre. Secondly, the other method is to use the user-application-type field of FIG(Fast Information Group) 0/13 on FICs(Fast Information Channels)[3].

The proposed T-DMB filecasting terminal is designed and implemented using the first method as mentioned above. Because the first method can reduce overhead of DMB channel information and transfer many images and much more additional information systematically.

In Figure 1, T-DMB filecasting terminal consists of DAB data block and DMB filecasting block. DAB data block is a basic function module of the legacy DMB device implemented by DMB specification. Output data of DAB data block are metadata containers based on DMB-ECG and multimedia contents compliant with DMB-AF. DMB filecasting block is comprised of update manager module to manage version of filecasting contents and metadata containers, filecasting decoder module to extract filecasting contents from containers and DMB-AF data, and metadata engine to parse ECG metadata.

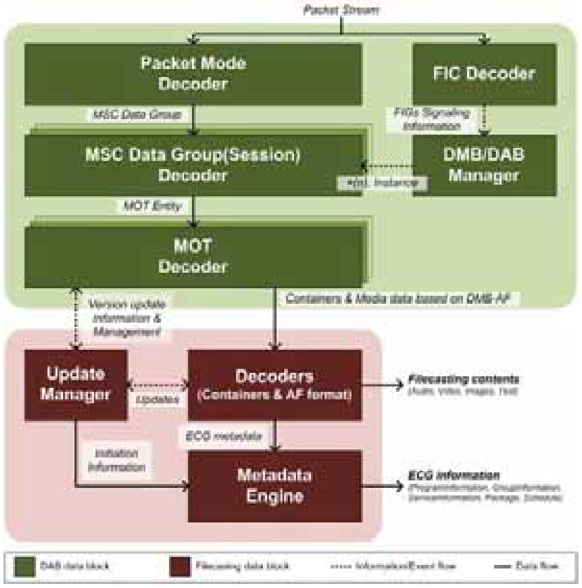


Figure 1. Structure of DMB filecasting terminal

III. Signaling and transmission

We will explain method of signaling and transmission for filecasting service in this chapter

A. Signaling of filecasting contents

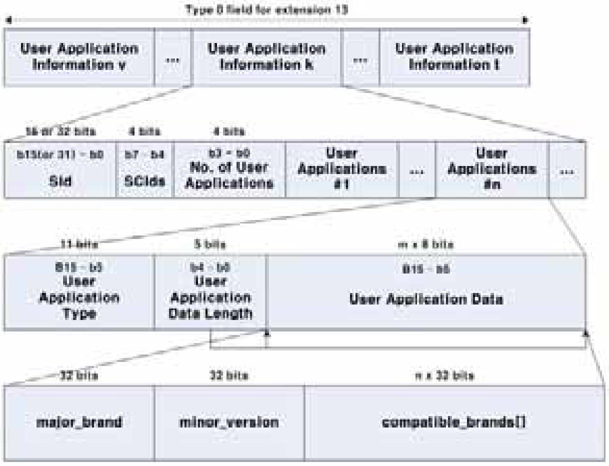


Figure 2. Structure of FIG 0/13 user-application-type

In this section, we explain about signaling method with user-application-type and brands.

Two kinds of brands are set in the File Type Box(ftyp box) of filecasting contents based on DMB-AF. One is the major_brand

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

that identifies the specification of the best use for the file. Second is the compatible_brands, which can identify multiple specifications to which the file complies. If more than one brand is present in the list of the compatible_brands, and one or more brands are supported by the player, the player shall play those aspects of the file that comply with those specifications[2].

Therefore T-DMB terminal can recognize kinds of data services or applications by FIG 0/13 user-application-type including brand information.

For transmission and signaling simulation, we set FIG 0/13 user-application-type value as ‘0x7DF (Filecasting Service)’ and ‘0x7FF (ECG Service)’ as shown in Figure 2. Then we set major_brand value as ‘dv4v’, compatible_brands value as ‘dv3b’, and minor_version value as ‘0x00’ extracted from the ftyp box of filecasting contents.

B. Transmission of ECG metadata

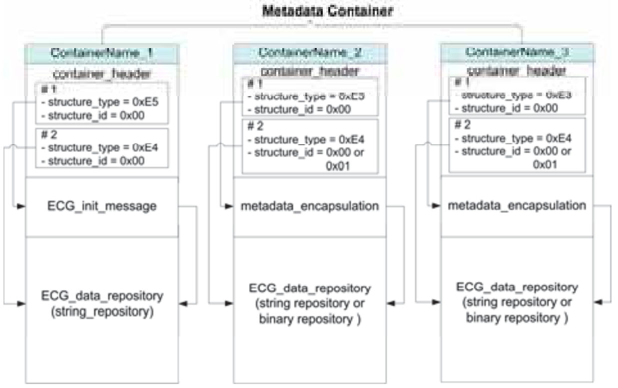


Figure 3. Structure of ECG metadata container

In this section, we explain about transmission method of ECG metadata container including ECG metadata. In Figure 3, ECG metadata container can take information of ECG metadata encapsulation, textual/binary encapsulation, and ECG initialization message to extract metadata[3].

ECG metadata is stored as a form of meaningful fragment in data repository of container. Then metadata container is encapsulated and transferred by using MOT(Multimedia Object Transfer) transmission protocol[3].

C. Transmission of multimedia data

Multimedia data are transferred as MOT objects of MOT directory mode through filecasting channel.

The following extensions and restrictions apply for filecasting channel[4].

- MOT Directory must be re-broadcast at a minimum repetition rate of 60 seconds.
- Compression of the MOT Directory is permitted.
- As a minimum, the MOT Directory shall list the file or files currently in the MOT carousel.
- Directory entries of the MOT Directory shall be sorted in ascending order of the ContentName parameter.
- MOT directory shall have a maximum uncompressed size of 16 Kbytes (16,384 bytes).

VI. Results

The T-DMB filecasting service is personalized and interactive DMB service because filecasting contents can be stored and distributed. In addition, various types of multimedia can be supported easily by additional definition of DMB-AF components without modification and extension of T-DMB transmission specification.

As shown Figure 4, T-DMB filecasting player received DMB-ECG metadata can provide us with filecasting contents information such as title, context, schedule information, and so on. If selected filecasting content is on air, player downloads filecasting contents at once. If it is not, player waits until filecasting content is transferred. After downloading filecasting contents, player can play all contents by using content-map information which is hierarchical structure and relation information among many sub-contents in a filecasting content.

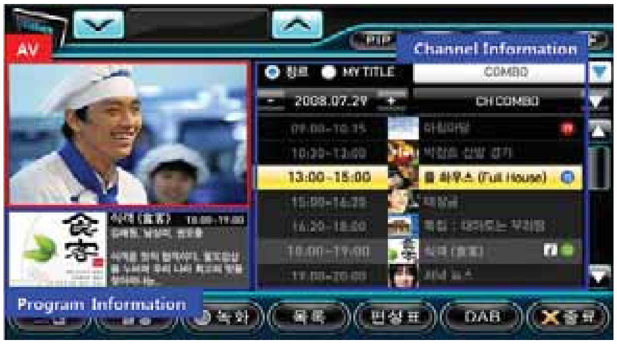


Figure 4. T-DMB filecasting player

V. Conclusion

We designed and implemented T-DMB filecasting terminal with minor modification of DMB-AF components and T-DMB transmission specification to consider backward compatibility with the legacy T-DMB data service.

The T-DMB filecasting technology can be applied to AT-DMB data services such as realtime auto-update services of navigation information, e-commerce services, personalized education services and etc.

In future, it will be necessary to discuss how to transfer huge amounts of contents through T-DMB filecasting channel and how to recover transmission error for efficient data download on MOT protocol level. Moreover, T-DMB bi-directional filecasting technologies will be needed for convergence of broadcasting and communication.

References

[1] TTAK.KO-07.0066, “Encoding and Transportation of Electronic Content Guide (ECG) for Terrestrial Digital Multimedia Broadcasting (DMB),” 2008.

[2] ISO/IEC 23000-9:2008 “Information technology — Multimedia application format (MPEG-A) — Part 9: Digital Multimedia Broadcasting application format.”

[3] Ji Hoon Choi and Han-kyu Lee, “The implementation of terminal for a personalized DMB service,” ICACT 2010.

[4] Ji Hoon Choi, Kyutae Yang, and Jihun Cha, “DMB Filecasting Service Technology,” Journal of Broadcast Engineering, Jan 2012.

An UHDTV Cable Television Distribution in Combinations of Multiple 64 and 256 QAM Channels

Yoshitaka Hakamada*1, Naoyoshi Nakamura*1, Kimiyuki Oyamada*1, Takuya Kurakake*2,
and Takeshi Kusakabe*3
*1NHK Science & Technology Research Laboratories *2NHK Sendai Broadcasting Station
*3NHK Matsuyama Broadcasting Station

Abstract-- High speed transmission for UHDTV cable broadcasting is achieved by adopting MPEG-2 TS basis channel bonding technology. A newly designed TDM frame format based on ITU-T J.183 is proposed. A 181.2 Mbps signal transmitted by a 64 QAM and four 256 QAM channels is received error-free at our prototype set-top box.

I. INTRODUCTION

An extremely high-resolution video system called Super Hi-Vision (SHV) that can provide an increased sense of reality and presence to viewers is being developed [1]. SHV is an ultra high-definition television (UHDTV), having 16 times more pixels than a standard HDTV. The development target is to start experimental broadcasts in 2020 by using a broadcasting satellite in the 21-GHz band.

We are developing an UHDTV transmission technology for cable television systems. The present transmission scheme for digital cable television in Japan is 64 QAM and 256 QAM that have been standardized in Annex C of ITU-T J.83 [2]. The physical bit rate per channel is 31.644 Mbps with 64 QAM and 42.192 Mbps with 256 QAM. A cable television channel can usually provide one or two HDTV services using MPEG-2 coding.

UHDTV needs a much larger transmission capacity for broadcasting than HDTV. We decided to utilize channel bonding technology for the UHDTV cable distribution and developed extended modifications to the frame structure based on ITU-T J.183.

II. PROPOSAL FOR SUPER FRAMES

Cable television systems operate with 64 QAM and 256 QAM signals. The framing structure of the transport streams multiplexing frames (TSMF) [3] is used in current digital cable television systems in Japan. TSMF bundles multiple MPEG-2 TSs into a single stream to transmit them by using a conventional cable QAM modulator. TSMF is composed of 53 slots. The length of each slot is 188 bytes, which is the same length as an MPEG-2 TS packet. The information identifying bundled TS streams and other additional information are stored in the TSMF_header located in the first slot of TSMF.

We propose a super frame consisting of multiple TSMFs to transmit large capacity MPEG-2 TS such as UHDTV using channel bonding technology. Figure 1 outlines the structure of a super frame. The number of TSMFs in a super frame is determined to make the periods of super frames identical regardless of the modulation format. A super frame has three TSMFs for a 64 QAM and four TSMFs for a 256 QAM according to the bit-rate ratio.

The large capacity of MPEG-2 TS is divided at the cable television headend and multiplexed into super frames. The modulation scheme of either channel, 64 QAM or 256 QAM, is determined depending on its transmission characteristics.

Signals of different frequencies may propagate at different speeds. A receiver has to temporally align all signals demodulated from bonded carriers. The TSMF_header of the first TSMF in every super frame is utilized as a marker to synchronize channels. The maximum acceptable time difference between channels is equal to the period of super frames, i.e., 8.2 ms. After all signals have been aligned, the receiver restores the split signals to the original MPEG-2 TS.

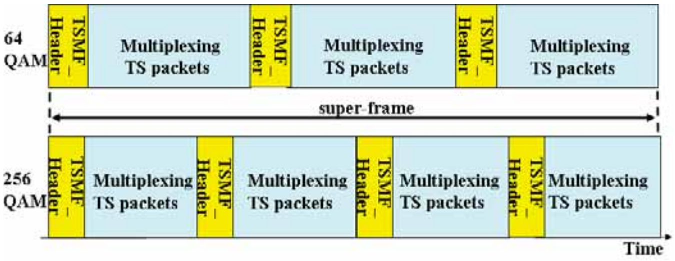


Fig. 1. Structure of super-frame for 64 QAM and 256 QAM

III. EVALUATION OF PERFORMANCE

We evaluated the performance of BER and the functionality of multiple-channel bonding with our prototype to find out how well our proposed method of transmission worked. As can be seen in Fig. 2, 181.2 Mbps MPEG-2 TS is transmitted with five channels. One is a 64 QAM channel and the other four are 256 QAM channels. Table 1 summarizes the parameters of the transmitted signals. The input power at each tuner is -43.7 dBm. Additive white Gaussian noise (AWGN) is added to all channels.

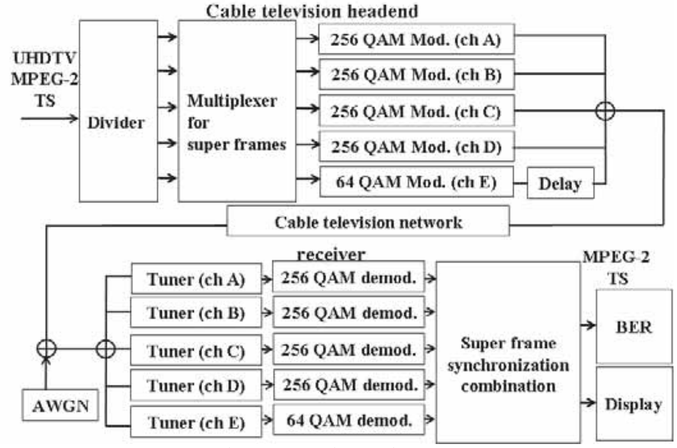


Fig. 2. Experimental setup

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

TABLE I
PARAMETERS OF TRANSMITTED SIGNALS

MPEG-2 TS rate	181.1 Mbps
No. of channels	Four channels with 256 QAM and a channel with 64 QAM
Bandwidth per channel	6 MHz
Symbol rate	5.274 Mbaud
Bit rate per channel	256 QAM : 42.192 Mbps 64 QAM : 31.644 Mbps
FEC framing	Reed Solomon (204,188)

A. *Performance of BER*

The theoretical BER of 2^n-QAM with rotational symmetry and gray coding is calculated by:

$$BER = \frac{\sqrt{2^n} + \left(\frac{n}{2}\right) - 1}{\sqrt{2^n} \cdot \left(\frac{n}{2}\right)} \cdot erfc\left(\frac{\delta}{\sqrt{2}\,\sigma}\right),. \qquad (1)$$

where n is the number of bits per symbol, δ is half the minimum distance between coded symbols, and σ^2 is the variance in AWGN. The BER_{UHDTV}, which is the BER of a restored MPEG-2 TS at a receiver, is derived as:

$$BER_{UHDTV} = \frac{1}{5}\left\{\frac{10}{24} erfc\left(\frac{CNR_{64QAM}}{42}\right) + \frac{19}{64}\sum_{i=1}^{4} erfc\left(\frac{CNR_{256QAM,i}}{170}\right)\right\},. \qquad (2)$$

where CNR_{64QAM} and $CNR_{256QAM,i}$ are the carrier-to-noise ratios (CNRs) of the 64 QAM carrier and 256 QAM carriers.

Fig. 3 plots the measured performance of BER. The required CNR is 30.9 dB for the BER of 2×10^{-4} before forward effort correction to achieve quasi-error free performance using Reed Solomon (204,188) coding. This demonstrates that the divided signals are restored in the original MPEG-2 TS signal at the receiver.

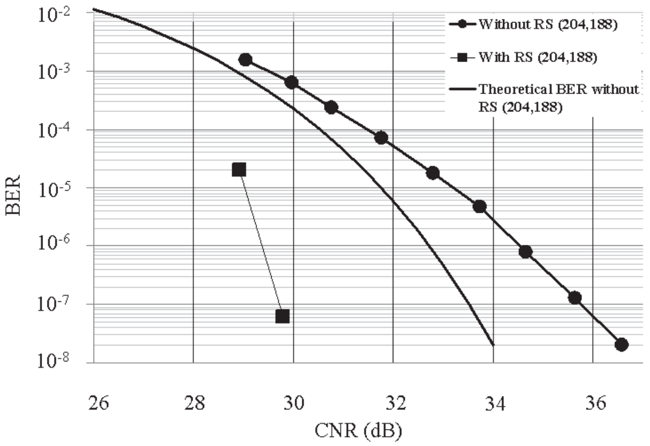


Fig. 3. BER performance of UHDTV

B. *Time alignment between channels*

The functionality to adjust delays between multiple channels was evaluated. An optical fiber was applied to a 64 QAM channel transmission as a delay line. The delay time therefore corresponded to the length of the fiber. The length of a 500-km long optical fiber causes about 2.5 ms of delay. The received CNR of the signals was adjusted to a constant value of 31.5 dB, independently of the length of the fiber. Fig. 4 plots the BER performance of signals versus the delay time of a 64 QAM. The experimental results indicate that the measured BER_{UHDTV} is almost constant and our prototype receiver can successfully compensate for delays and synchronize the super frames of multiple QAM carriers.

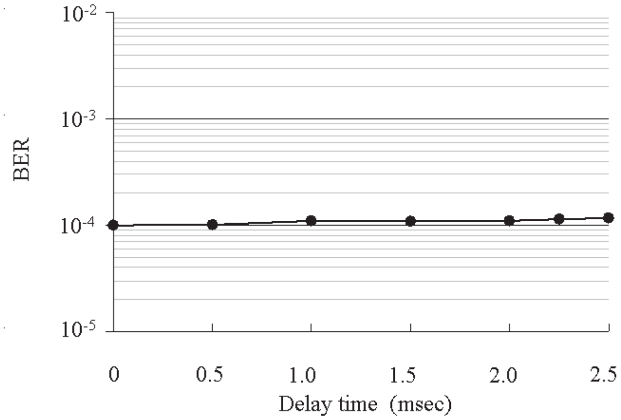


Fig. 4. BER performance of a MPEG-2 TS versus delay time of a 64QAM

The MPEG-2 TS of an UHDTV that was divided into a 64 QAM channel and four 256 QAM channels was transmitted with the prototype in the experiment. We also observed an UHDTV video that was stably displayed.

IV. CONCLUSION

High speed transmission of cable television in combination with differently modulated multiple signals was achieved. Our proposed super frames made it possible to deliver a UHDTV signal via existing cable television networks by dividing the large capacity of MPEG-2 TS into multiple channels with 64 QAMs and 256 QAMs in any combinations. We evaluated the BER performance of UHDTV and the functionality of multiple-channel bonding with a prototype. The experimental results demonstrated the feasibility of UHDTV transmission in existing cable television networks.

REFERENCES

[1] E.Nakasu, “Super Hi-vision on the Horizon,” IEEE Consumer Electronics Magazine, vol. 1, no. 2, pp. 36–42, April 2012.

[2] *Digital multi-programme systems for television, sound and data services for cable distribution*, ITU-T J.83, ITU, ITU-T, 2007.

[3] *Time-division multiplexing of multiple MPEG-2 transport streams over cable television systems*, ITU-T J.183, ITU, ITU-T, 2001.

Performance Analysis on Cooperative Reception of ISDB-T One-Segment Service Against the Number of Terminals

Ryo Araki, Akira Nakamura, Kohei Ohno, Makoto Itami
Department of Applied Electronics, Facility of Industrial Science and Technology, Tokyo University of Science

***Abstract*—This paper discusses a cooperative reception technique for Japanese digital terrestrial television broadcasting "One-segment service" using multiple terminals. In this paper, received signals and the channel information are transmitted by using one-way or two-way communication. Maximum ratio combining scheme are performed in the own terminal using the exchanged channel information. The error rate performance can be improved by using multiple receiving terminals in the One-Segment Service.**

I. INTRODUCTION

The Japanese DTTB (Digital Terrestrial Television Broadcasting) is standardized as the ISDB-T (Integrated Services Digital Broadcasting- Terrestrial). In the system, OFDM (Orthogonal Frequency Division Multiplexing) modulation is used. One-Segment Service is a broadcasting service for mobile terminals like a mobile phone [1].

However, reception quality of One-Segment Service is degraded by dips because the bandwidth is narrower than ordinal 13-Segment ISDB-T. The dip is a drop of received power in a multipath fading channel. The antenna diversity is one of the effective techniques to solve the multipath fading problem. However, it is difficult to install multiple antennas in a small mobile terminal. Therefore, the cooperative reception was proposed [2].

In cooperative reception, received signals and the channel information are exchanged by a radio link to obtain the antenna diversity effects. However, the data rate of the radio link is limited. Therefore, the reduction scheme of amount of sharing data using scattered pilot symbols was proposed [3][4].

In this paper, we discuss cooperative reception technique using multiple receiving terminals in order to improve the receiving characteristic. Received signals and channel information are exchanged by using one-way or two-way communication. The exchanging schemes are proposed and the performances are evaluated.

II. COOPERATIVE RECEPTION TECHNIQUE

In this paper, cooperative reception technique using multiple receiving terminals is proposed. The received signal and the channel information are shared by using Bluetooth between own terminal and other terminals.

However, all received signals and the channel information in receiving terminals cannot be shared because the data rate of Bluetooth is limited. Therefore, the sharing data decreasing scheme exchanging scattered pilot symbols was proposed [3]. In ISDB-T, 37 scattered pilot symbols are inserted in 1 OFDM symbol. They are quantized and exchanged. The channel information is estimated from the scattered pilot symbols.

A. One-way Communication method

The One-way communication method is that other terminals transmit the received signal and the channel information in order of more powerful carriers one-sidedly. Thus, the own terminal only receives them. The model of cooperative reception by the one-way communication is shown in Figure 1.

The procedure is explained as follows:

1. Other terminals transmit scattered pilot symbols and received signals in order of more powerful carriers. The transmitted information is quantized.
2. Own terminal estimates the channel information of other terminals from received scattered pilot symbols.
3. In own terminal, maximum ratio combining with received signals of own and other terminals is performed using scatter pilot symbols.

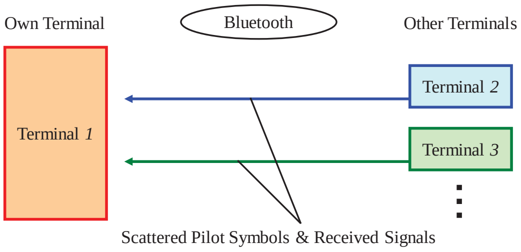


Figure 1. Model of Cooperative Reception using One-way Communication

B. Two-way Communication method

In the Two-way communication, the own terminal requests the received signal of the weaker carriers to the other terminals. The model of the cooperative reception by using two-way communication is shown in Figure 2.

The procedure is explained as follows:

1. Own terminal transmits scattered pilot symbols that are quantized.
2. Other terminals estimate the channel information of own terminal from received scattered pilot symbols.
3. Other terminals transmit scattered pilot symbols and received signals of other terminals in order of weaker carriers in own terminal.
4. Own terminal estimates the channel information of other terminals from received scattered pilot symbols.
5. In own terminal, maximum ratio combining with received signals of own and other terminals is performed using scatter pilot symbols.

III. SIMULATION

In this paper, we proposed the cooperative reception by using more than two terminals for ISDB-T:mode3. Parameters are set the same as the ISDB-T One-Segment Service [1].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

In this consideration, Bluetooth 2.0+EDR that the data rate is 1306.9 kbps is considered. When scattered pilot symbols are transmitted, 37 scattered pilot symbols per one OFDM signal are quantized by 2bit [4]. Then, the number of sharable carriers per one terminal using one-way or two-way communication is shown in Table 1. When the number of terminals for the cooperative reception is increased, the number of shareable carriers is decreased. In this simulation, three paths model is used [4].

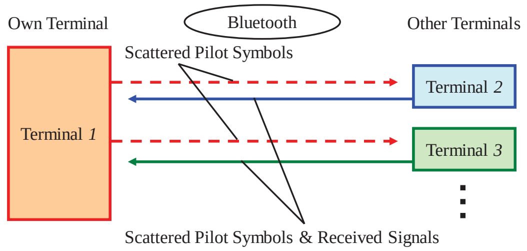


Figure 2. Model of Cooperative Reception using Two-way Communication

Table 1. Sharable Carriers per one Terminal

The Number of Own and Other Terminals	Sharable Carriers per one Terminal	
	One-way Communication	Two-way Communication
2 terminals	432 carriers	432 carriers
3 terminals	261 carriers	228 carriers
4 terminals	152 carriers	130 carriers
5 terminals	98 carriers	81 carriers

A. *Cooperative Reception using One-way Communication*

In Figure 3, the BER (Bit Error Rate) performance of MRC using the one-way communication is shown.

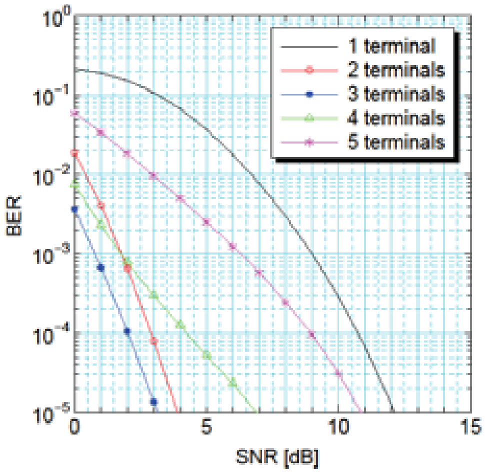


Figure 3. The BER Performance of Cooperative Reception with One-way Communication

When the number of terminals is three, the BER performance becomes the best. The performance of the proposed scheme using three terminals becomes better about 9.0 dB than the performance not to adopt the cooperative reception. When the number of terminals is increased, the diversity gain can be obtained, however the number of sharable carriers are decreased.

B. *Cooperative Reception using Two-way Communication*

In Figure 4, the BER performance of MRC using the two-way communication is shown.

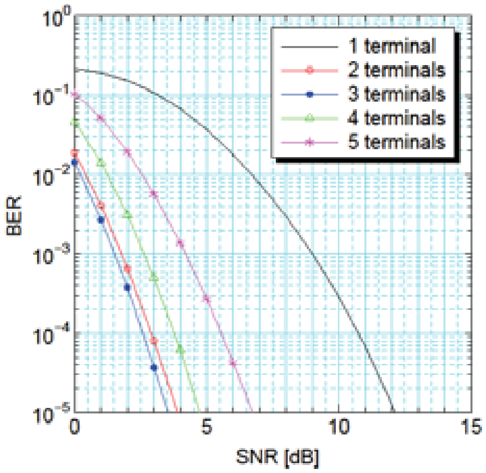


Figure 4. The BER Performance of Cooperative Reception Using with Two-way Communication

When the number of terminals is three, the BER performance also becomes the best. The performance of the proposed scheme using three terminals becomes better about 8.5 dB than the performance not to adopt the cooperative reception.

Comparing the results of Figure 3 and 4, the BER of using one-way communication becomes better about 0.5 dB than the BER of using two-way communication in case of the cooperative reception using 3 terminals. In two-way communication scheme, they have potential that the sharing carrier numbers are overlapped among the terminals. Then, the diversity gains become less than the case of using one-way communication.

IV. CONCLUSION

In this paper, the cooperative reception technique using multiple received terminals is considered in order to improve the receiving characteristic. Received signals and channel information is transmitted by the one-way or two-way communication.

The BER performances of the proposed scheme using three terminals become the best. Therefore, it is able to improve the error rate characteristic of the One-Segment Service using multiple receiving terminals.

REFERENCE

[1] "Transmission system for Digital Terrestrial Television Broadcasting", ARIB STD-31, v2.0, Mar. 2011

[2] T. Okubo, M. Fujii, M. Itami, "A study on cooperative reception of one segment ISDB-T", ITE Technical Report Vol. 31, No. 36, Jul. 2007.

[3] N. Kobayashi, et. al., "Cooperative Reception using Scattered Pilot Symbol for ISDB-T One- Segment Service", ICCE2011, Jan. 2011.

[4] R. Araki, N. Kobayashi, K. Ohno and M. Itami, "Cooperative Reception Scheme Using Multiple Terminals for Digital Terrestrial Television Broadcasting One-Segment Service", ITE Technical Report, Feb. 2012

Software-defined DTV Platform

Jungpil Yu, Hyun-Yong Lee, Chang Hoon Choi, and Jaehun Chung
DMC R&D Center, SAMSUNG Electronics, Suwon, Republic of Korea

***Abstract*— Since the first appearance of the digital TV standard in mid-1990, the number of digital Television standard increases up to a dozen until now. In this circumstance, it is no wonder for a modern DTV to support several standards simultaneously. For example, almost all DTVs sold in Europe support DVB-T standard as well as DVB-C standard. From the engineering standpoints, this multi-standard DTV is a good candidate for software-defined radio concept which allows the DTV to adapt to various standards by only affecting the software modules. In this paper, a software-defined DTV platform which has been developed and successfully tested in SAMSUNG electronics R&D center is described. This platform is equipped with a reconfigurable processor and common hardware. As an example, the 2nd generation terrestrial broadcasting, namely DVB-T2 is successfully incorporated into the platform of FPGA (field-programmable gate arrays).**

I. Introduction

The introduction of digital TV has attracted a lot of interest in the attainable broadcasting data rate limit. Recent DVB standards such as DVB-T2 and DVB-C2 are sophisticated enough to be regarded as those approaching this limit [1][2]. They are the 2nd generation version of DVB-T and DVB-C, respectively and characterized by a higher spectral efficiency than the former standards [3][4]. On the other hand, as the standards are getting more complex, the receiver implementation is getting harder. To mitigate this implementation burden, a software-defined DTV platform can be conceivable.

The software-defined radio (SDR) concept is as follows [5]. The different receiver algorithms are implemented in software based on the same hardware. The resulting receiving function is achieved in a cost effective fashion.

This paper is organized as follows. After this brief introduction, Section II gives an overview over the concept of the Software-Defined DTV platform and implementation aspects. In section III, the performance of the platform is addressed and finally followed by some remarks as a conclusion.

II. Software-Defined DTV Platform

A. Concept

The DTV platform presented in this paper faithfully follows the common SDR concept depicted by the figure 1.

Each software module alters the behavior of reconfigurable processor which results in desired receiving functions. Some common hardware logic circuit is included to assist the reconfigurable processor working on cycle demanding functions.

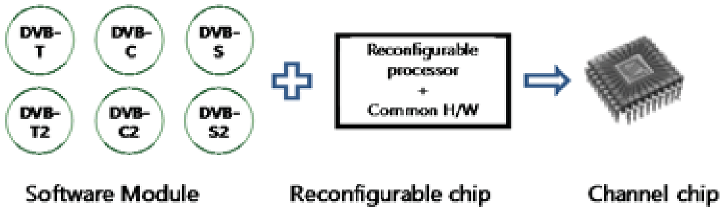


Fig. 1. Software-defined DTV concept

B. DTV Platform

The developed platform is comprised of several blocks which are shown in figure 2. Basically, 4 different blocks are identified. They are a digital front end, a reconfigurable processor, common H/W, and a control processor.

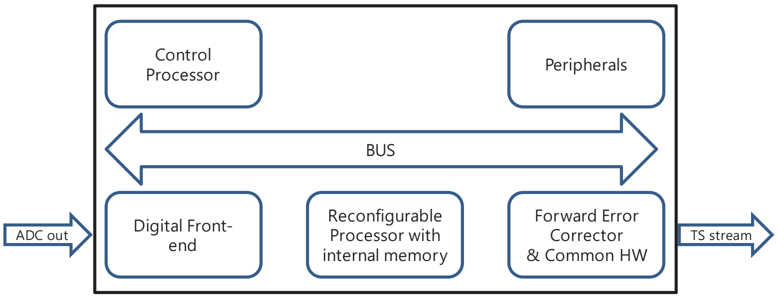

Control Processor
Peripherals
BUS
ADC out
Digital Front-end
Reconfigurable Processor with internal memory
Forward Error Corrector & Common HW
TS stream

Fig. 2. DTV Platform Structure

1) Digital Front-end

Digital Front-end receives the RF (Radio Frequency) signal and the signal is periodically sampled to digital data in ADC (Analog to Digital Converter). Its role is divided into two. One is sampling rate conversion from the ADC sampling rate into the symbol clock rate applying symbol timing compensation. The other is compensation of the carrier frequency and phase offset estimated in the next block (Reconfigurable Processor). In order to do so, it is essential to recover the carrier and symbol timing correctly. The automatic gain control of input signal level is also done in the digital front-end.

2) Reconfigurable Processor

The Reconfigurable Processor (RP) used in the platform is the in-house digital signal processor having a coarse-grained reconfigurable architecture [6]. It consists of a coarse-grained array of functional units (FU), global and local register files, instruction, data, and configuration memory, and several bus connections for off-chip data transfers. The FUs are arranged on a 4x4 grid. The core can operate in two modes: VLIW (Very Long Instruction Word) and CGA (Coarse-Grained Array). Switching between the two modes happens by special

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

operations.

In VLIW mode, the RP core behaves like a general purpose 4-issue VLIW processor with several pipeline stages. Only the FUs that were assigned to the different VLIW slots are active. VLIW mode is used for sequential or infrequently executed code.

In CGA mode, the RP core operates in data flow mode. In other words, there is no control flow (i.e. no branches). All FUs are active. All entities (FUs, local and global register files, muxes) are programmed via bits in so-called configuration memories. CGA mode can exploit loop level parallelism (LLP) and is thus used exclusively for loops.

3) Common Hardware

The common hardware described in figure 2 is mainly the forward error correction (FEC) block. The FEC of DVB-T2 is a concatenation of LDPC codes and BCH codes with a codeword length of 16200 and 64800 bits. The long codeword block size is favorable in the decoding performance but entails a long latency as a cost. It requires a huge number of clock cycles for a processor to decode it. This is the reason why the FEC is executed by common hardware.

4) Control Processor

The all blocks are started and synchronized by the control processor. Besides, this processor takes part in de-jittering in stream adaptation before the final Transport Stream (TS) output.

C. Platform Prototype

All blocks are incorporated into one FPGA system for a prototyping. The FPGA system consists of 4 FPGA devices and runs at 35MHz bus frequency. This is not enough to decode DVB-T2 signal of 8MHz bandwidth in real time. However, the DVB-T2 signal of 1.7 MHz bandwidth is decoded and used for the platform tests.

III. Performance

A. Platform Test Setup

The DTV Platform has been set up as shown in figure 3. A DVB-T2 modulator feeds the modulated signal into the RF tuner and ADC board connected into the FPGA board. The DVB-T2 demodulation happens in the FPGA board and retrieves the transport stream packets. These are finally fed to a TV set for a visual display.

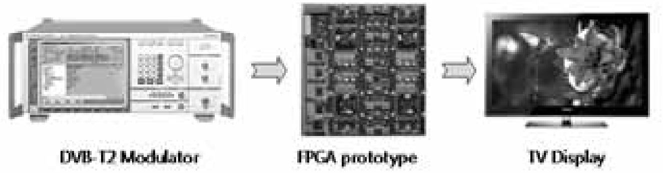


Fig. 3. DTV Platform Test Configuration

B. Experiments

The industry association for digital television in the United Kingdom, the Digital TV Group (DTG) publishes and maintains the technical specification (the D-Book) for the Digital TV. The association also serves as the digital television test center. The D-Book stipulates a set of DVB-T2 functional test stream parameters which are identified by dtg *x* stream where *x* is numerals. The single PLP streams from dtg 0 to dtg 161 are candidates for a real time display and they are tested. Some of functional tests have passed without any particular care but the most have passed with the reduced FEC blocks per interleaving frame. This is due to the slow FPGA operating frequency. The reduced FEC blocks per interleaving frame alleviates the computational burden and renders the DVB-T2 signal decodable. The carrier-to-noise (C/N) for single PLP streams are well within the 4dB + (required C/N for reference BER listed in section 9.13.2.1, DVB-T2 carrier-to-noise in the D-Book).

IV. Conclusion

In this paper, the authors presented the software-defined DTV platform developed and tested in DVB-T2 broadcasting environment. The developed FPGA-based platform prototype shows limited decoding ability due to the current FPGA operation frequency. However, this platform is scheduled to be made as an integrated circuit and is expected to be powerful enough for real time decoding.

The authors consider that this receiver solution is fairly attractive in reducing the total DTV set cost. In addition, the flexible nature given by software-defined platform is quite well suited to the current overwhelming DTV standard circumstance.

Acknowledgment

The authors wish to thank many colleagues collaborating in developing the DTV platform and preparing this paper. They are Jeongrok PARK, Sung il PARK, Seong-Hee PARK, Youngho OH, Soon Chan KWON, Jun-Seok KANG, Jisung OH in DMC R&D Center, and Ho YANG in SAIT.

ReferenceS

[1] *Digital Video Broadcasting (DVB); frame structure, channel coding and modulation for a second generation digital terrestrial television broadcasting system* (DVB-T2), ETSI EN 302 755 V1.1.1, Sep. 2009.

[2] *Digital Video Broadcasting (DVB); frame structure, channel coding and modulation for a second generation digital transmission system for cable sysmtes* (DVB-C2), ETSI EN 302 769 V1.2.1, Apr. 2011.

[3] *Digital Video Broadcasting (DVB); frame structure, channel coding and modulation for digital terrestrial television* (DVB-T), ETSI EN 300 744 V1.2.1, July. 1999.

[4] *Digital Video Broadcasting (DVB); frame structure, channel coding and modulation for cable systems* (DVB-C), ETSI EN 300 429 V1.2.1, Apr. 1998.

[5] W. Tuttlebee, *Software Defined Radio: Enabling Technologies*. New York, USA: John Wiley & Sons, 2002.

[6] F. Bouwens, M. Berekovic, B. De Sutter, and G. Gaydadjiev, *Architecture Enhancements for the ADRES Coarse-Grained Reconfigurable Array*. pp.66-81, High Performance Embedded Architectures and Compilers: Springer Berlin Heidelberg, 2008.

Quality Assessment of Compressed Video Sequences Having Blocking Artifacts by Cepstrum Analysis

Yuta YAMAMURA[†], Shinya IWASAKI[†], Yasutaka MATSUO[†‡], *Member, IEEE*, and Jiro KATTO[†], *Member, IEEE*
[†]Department of Computer Science, Waseda University, Tokyo, Japan
[‡]NHK Science & Technology Research Laboratories, Tokyo, Japan

Abstract--Objective picture quality measures cannot estimate the effect of blocking artifacts caused by video compression sufficiently. In this paper, we apply cepstrum analysis to quantify the blocking artifacts. We show experimental results for some test sequences using different coding schemes and prove effectiveness of our approach.

I. INTRODUCTION

It has been long pointed out that PSNR (peak-to-peak SNR) does not match subjective quality assessment of compressed video sequences sufficiently. Various objective quality assessment methods had been developed to overcome this problem, especially in VQEG (Video Quality Experts Group). Some of them focus on blocking artifacts, which become remarkable in low bitrates, but are devoted to specific coding algorithm such as JPEG [1].

On the other hand, cepstrum analysis, which had been originally evolved in speech signal analysis, has been applied to video analysis tasks such as evaluation of blocking artifacts, image restoration and motion analysis [2-5]. This paper focuses on an application of the cepstrum analysis to objective assessment of blocking artifacts in compressed video sequences, and shows experimental results for some test sequences using various video coding standards.

II. CEPSTRUM ANALYSIS OF IMAGE SIGNALS

Cepstrum analysis has been mainly used in speech and audio analysis to extract pitch frequency and spectral envelope. In image processing, it can be used to detect periodic patterns in images such as blocking artifacts [2,3] and to separate blurring kernel from the image component [4,5]. Figure 1 illustrates cepstrum analysis, in which FFTs (Fast Fourier Transforms) and logarithmic transformation are applied to a difference image between an original image and a decoded image.

III. OBJECTIVE ASSESSMENT MEASURE BASED ON CEPSTRUM ANALYSIS

Cepstrum analysis enables detection of periodic gaps caused by blocking artifacts. As a pilot study, we apply cepstrum analysis to test sequences encoded by JPEG with very high compression ratio, such that most encoding bits are assigned only to the DC component. We also shrink/expand decoded images, by which blocking artifacts appear per 4, 8 and 16 pixels, and calculate their cepstrums. Figure 2 shows a relationship between a quefrency and its cepstrum amplitude.

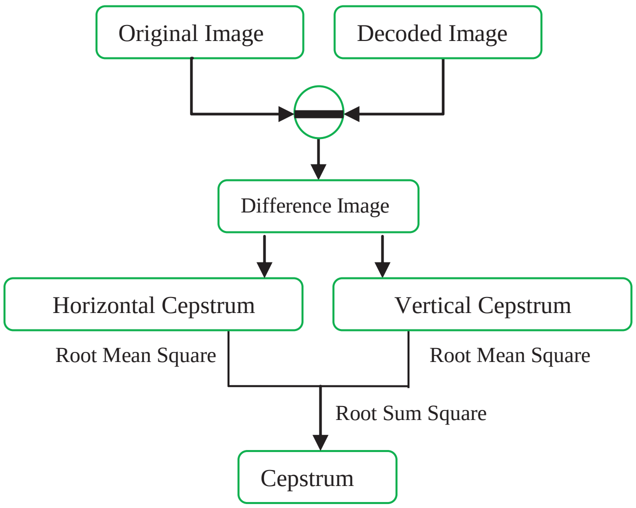


Fig.1: Flow chart of cepstrum analysis [2].

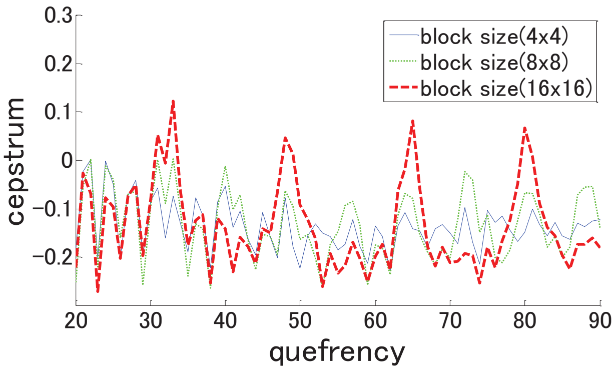


Fig. 2: A pilot study result of cepstrum analysis.

From this result, we can confirm that locations of the cepstrum peak values are associated with the block sizes. We then define our objective assessment measure of blocking artifacts by a mean peak value of cepstrum, of which quefrency period is equal to the DCT block size. We consider 8x8 block case only in this paper, though we evaluate HEVC and H.264/AVC which have variable block size DCT.

IV. EXPERIMENTAL RESULTS FOR VARIOUS VIDEO CODING STANDARDS

Figure 3 illustrates examples of original and decoded sequences of which size is CIF and which are encoded by 500kbps. In this experiment, we tried seven video coding standards, HEVC, H.264/AVC, MPEG-4 part2, H.263, MPEG-2, MPEG-1, H.261 and two still image coding standards (all-intra), JPEG-2000 and JPEG, and changed bitrates from 500kbps to 2Mbps. Figure 4 compares subjective

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

assessments corresponding to MOS (Mean Opinion Score) for total picture quality and blocking artifacts. In this comparison, we asked 6 subjects to score decoded sequences from two viewpoints of total picture quality and blocking artifacts. Figure 5 compares two objective assessment measures; PSNRs and mean peak values of cepstrum (note that vertical axis of the cepstrum peak is inversed for comprehensibility).

First, from Figure 3, we can observe that blocking artifacts are remarkable except HEVC, H.264/AVC and JPEG-2000, and total qualities are also bad except HEVC and H.264/AVC. Figures 4 and 5 explain this fact more quantitatively. As bitrates become large, MOS for blocking artifacts increases and mean peak values of cepstrum decreases with higher correlation, of which correlation coefficient, averaged over all bit rates, is 0.875 which is 0.069 larger than that of PSNR, 0.806.

Second, from Figures 4 and 5, within the tested bitrates, HEVC and H.264/AVC show higher quality scores against other standards in both subjective and objective assessments. Furthermore, it is observed that blocking artifacts are not clearly noticed in subjective and objective assessments. This is because the effect of a deblocking filter appears greatly, which had been applied since H.264/AVC.

On the other hand, in JPEG-2000, mean peak values of cepstrum stay almost constant. This fact aligns with theoretical expectation that blocking artifacts do not occur thanks to wavelet transform. In fact, subjective assessment suggests that subjects did not feel existence of blocking artifacts. In JPEG, both the subjective and objective assessments suggest heavy blocking artifacts as expected.

V.CONCLUSIONS

In this paper, we evaluated the influence of blocking artifacts based on cepstrum analysis, and compared nine video and image coding standards. Experimental results confirmed effectiveness of the proposed method. As further study, we will consider usage of spectral envelopes to improve the objective assessment measure and introduction of linear prediction coding instead of cepstrum analysis.

REFERENCES

[1] Zhou Wang, Hamid R. Sheikh and Alan C. Bovik, No-Reference Perceptual Quality Assessment of JPEG Compressed Images," IEEE ICIP 2002, September 2002

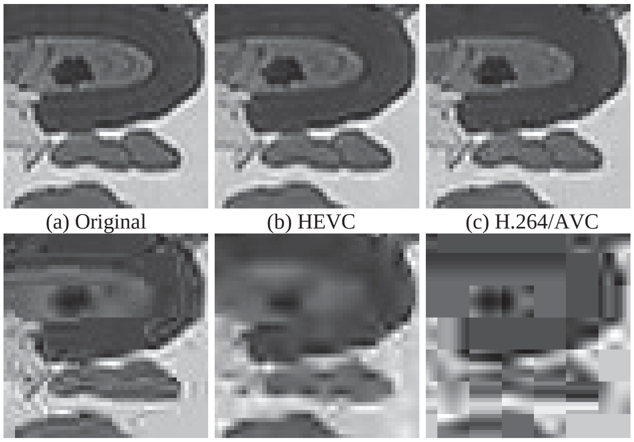
[2] H.Koda and H.Tanaka, A New Method of Measuring the Blocking Effects of Images Based on Cepstrral Information, IEICE Trans, Fundamental, Vol.E79 A, No.8, pp.1274-1282, August 1996

[3] H.Koda and H.Tanaka, Estimation of the Blocking Effects based on Cepstral Information," ISITA'94, pp. 173-176, November 1994.

[4] S. Wu, Z. Lu, E. P. Ong and W. Lin, Blind Image Blur Identification in Cepstrum Domain," ICCCN 2007.

[5] H. Ji and C. Liu: "Motion Blur Identification from Image Gradients", IEEE CVPR 2008, pp.1-8, 2008.

[6] Malver H.S. and Staelin D.H "The LOT: Transform coding without blocking effects", IEEE Trans. Accoust., Speech & Signal Process., ASSP-37, pp.553-559, April 1989.



(a) Original (b) HEVC (c) H.264/AVC

(d) MPEG-2 (e) JPEG2000 (f) JPEG

Fig. 3: Examples of original and decoded sequences from mobile & calendar.

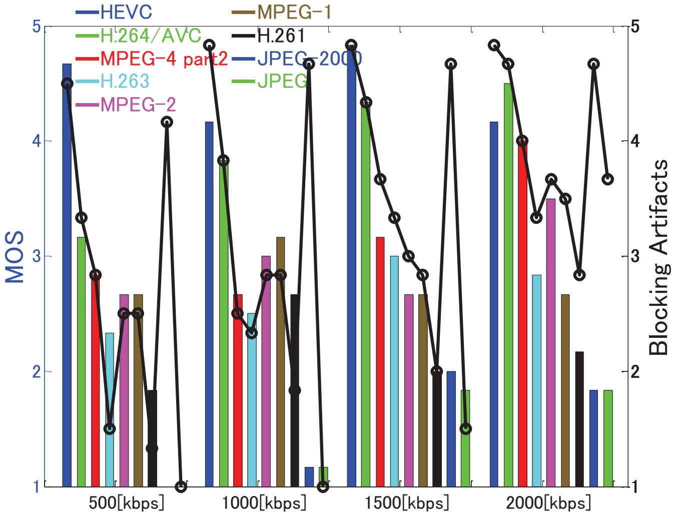


Fig. 4: Comparison of subjective assessments; MOS for total quality (bar) and blocking artifacts (circle).

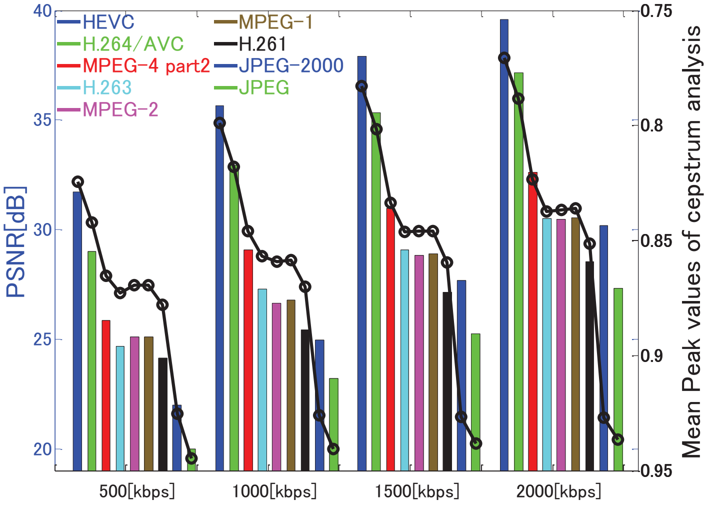


Fig.5: Comparison of objective assessments; PSNRs (bar) and mean peak values of cepstral (circle)

Perceptual Video Quality Assessment for Wireless Multimedia Applications

Yeong-Kang Lai, *Member, IEEE*, Yu-Fan Lai, *Student Member, IEEE,*
Cheng-Han Dai, and Thomas Schumann

***Abstract*—Perceptual video coding is more intuitive than traditional compression methods such as temporal, spatial, and statistical redundancy. With the popularity of wireless multiplication devices, the transmission of video is an important issue. In order to decrease the transmission time and get better performance, the proposed perceptual evaluation model can save not only bit-rates but also maintain video quality. From the experimental results, by adjusting the quantization parameters (QP) by associating visual properties, it can save about 6%~21% bit-rates without any noticeable difference in the visual qualities.**

I. Introduction

With the progress of technology, wireless multimedia devices such as PDA, smart phone, and tablet PC become more and more popular. For human visual system (HVS), the purpose of perceptual video coding is to keep the video quality and save more bit-rates. Usually we adopt mean square error (MSE) and peak signal to noise ratio (PSNR) to evaluate the video/image quality, but sometimes they are not absolutely correct [1]. Since the human eyes are most intuitive to receive the quality of the video or image, we use the concept of perceptual video coding [2][3] to estimate video/image quality. Using just-noticeable-distortion (JND) model [4] was a good method. However, their computations of algorithms are too complicated to accomplish in hardware implementation. H.264/AVC is the most widely coding standards in video applications, many researches [4]-[6] also adopt it as the video standard for perceptual video coding. In Fig. 1(a), we also add the perceptual evaluation model into H.264/AVC encoder.

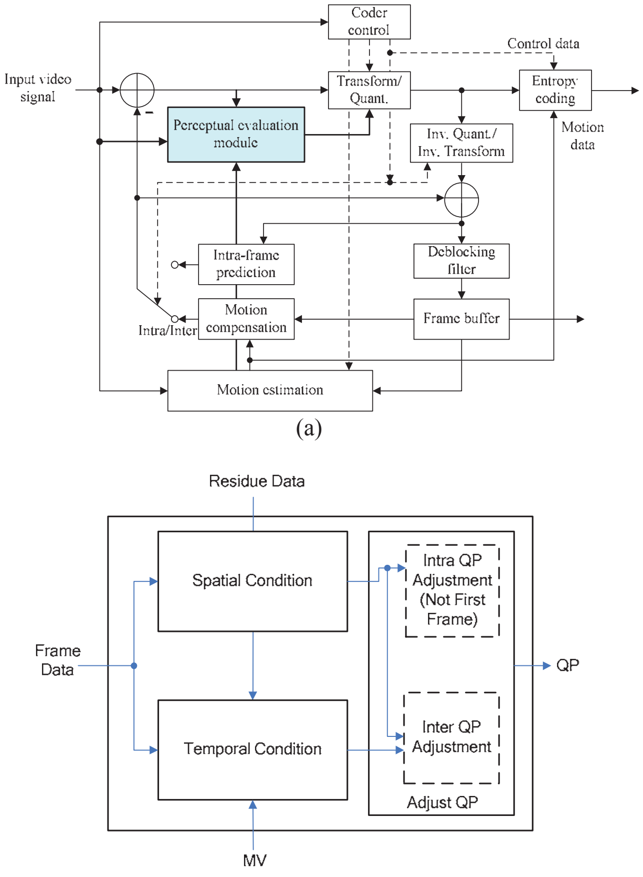
II. Proposed Algorithm

The trade-off between video qualities and bit-rates is determined by the quantization parameters (QP). Adjusting the QP value may affect not only the current macro block (MB) but also the others. The best way is to adopt the rate-distortion (R-D) cost.

$$J(s,c,mode \mid QP)=D(s,c,mode \mid QP)+\lambda R(s,c,mode \mid QP)$$

where s and c indicate the original and reconstructed image block, respectively. In Fig. 1(b), the proposed perceptual evaluation model which consists of two parts: spatial condition and temporal condition. The detailed descriptions are as follow,

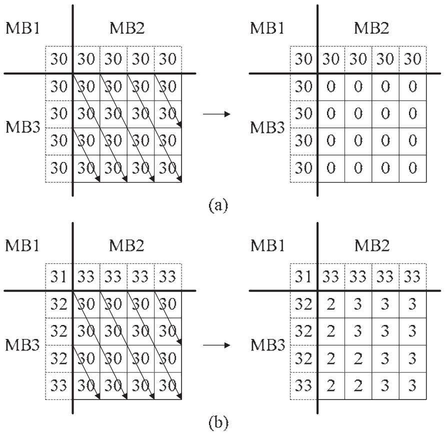
A. Spatial Condition



(a)

(b)

Fig. 1. (a) Block diagram of H.264/AVC encoder which contains the proposed perceptual evaluation module. (b) Proposed perceptual evaluation module.



30	30	30	30	30
30	30	30	30	30
30	30	30	30	30
30	30	30	30	30
30	30	30	30	30

30	30	30	30	30
30	0	0	0	0
30	0	0	0	0
30	0	0	0	0
30	0	0	0	0

31	33	33	33	33
32	30	30	30	30
32	30	30	30	30
32	30	30	30	30
33	30	30	30	30

31	33	33	33	33
32	2	3	3	3
32	2	3	3	3
32	2	2	3	3
33	2	2	3	3

Fig. 2. Error propagation example

The spatial condition effect which is caused by the neighbor image quality change that affects the current MB quality is generated in the intra mode. It is observed that the current MB pixels should minus the prediction pixels. However, these prediction pixels may contain distortion. Fig. 2 shows the error propagation example of a Macro block (4x4 pixels). Fig. 2(a) generates no residues in the original condition. However, Fig. 2(b) is the condition whose neighbor MBs' QP values have been adjusted. Under this condition, the prediction pixels may contain more distortion and the current

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

MB remains more residues. Therefore, even though the current MB's QP is not adjusted, the quality and bit-rate may also be affected by the neighbor block whose QP is adjusted.

B. *Temporal Condition*

The temporal condition effect, which is the change in the current MB quality caused by the reference frame quality, is generated in the inter mode. It is that motion estimation (ME) will find a block which is like the current MB in the reference frame. If the QPs are increased in the reference frame, the quality of reference frame may decrease. Therefore, the ME can exhaustively find the best result. Human eyes can not find details of a region which has many changes when the temporal frequency is high. Hence we can save more bit-rates. The concept is represented as follows.

The differences of MBs are calculated.

$$Diff(i,j) = | p_t(i,j) - p_{t-1}(i,j) |$$

Filter the difference of pixels and calculate the change quantities.

$$Diff_t(i,j) = \begin{cases} 0 & if\ Diff(i,j) < Threshold\ value\ ) \\ 1 & Otherwise \end{cases}$$

$$CQ = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} Diff_t(k,l)$$

Where the $p_t(i,j)$ is the current MB pixel at location (i,j) and $p_{t-1}(i,j)$ is the pixel at location (i,j) in the previous frame. $N \times N$ is the MB size.

III. Experimental Results

The experimental results are as follows. The test sequences are Stefan and Football. Tables I and II show the bit-rate reduction compared to the originals (JM code). In order to sustain the video/image quality, we adopt subjective test [7] to examine the proposed perceptual evaluation model. Ten participants are invited to take part in this experiment; the experimental environment is processed in a dark room. The sequence is selected randomly, and avoid the same content appear sequentially. From Figs. 3 and 4, we can't find obvious differences from human eyes although PSNR has some degradation. However, the bit-rates can save about 6%~21% in the same human visual system.

IV. Conclusion

In this paper, we propose a perceptual model which adjusted QP in accordance with the visual properties. In order to know the effect of the video quality and bit-rate reduction, some experiments are designed. It can save about 6%~21% bit-rates without any noticeable difference in human visual system. Hence the proposed perceptual model can save more transmission time for wireless multimedia application.

Acknowledgment

This work was supported by the National Science Council of the Republic of China under Contract NSC 101-2220-E-005-004.

References

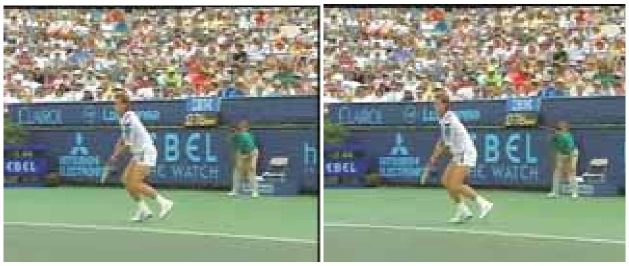
[1] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: from error visibility to structural similairty," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600−612, Apr. 2004

Table I Bit-rate reduction (Stefan sequence)

Stefan	H.264(JM Original)		Perceptual Evaluation model		
QP	Bit-rate	PSNR	Bit-rate	PSNR	Reduction
24	3102.368	39.869	2578.159	38.403	16.90%
26	2412.93	38.194	1954.967	36.667	18.98%
28	1883.011	36.596	1483.576	35.042	21.21%
30	1381.319	34.666	1094.248	33.288	20.78%
32	997.862	32.986	785.981	31.69	21.23%
34	737.246	31.481	606.57	30.232	17.72%
36	536.34	29.849	463.854	28.824	13.51%

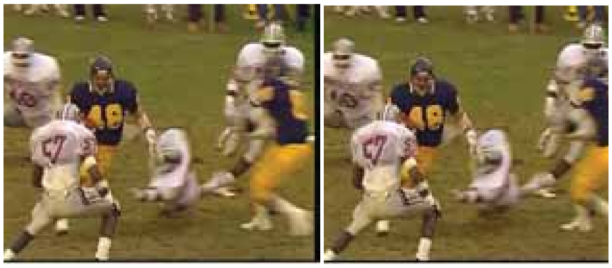
Table II Bit-rate reduction (Football sequence)

Football	H.264(JM Original)		Perceptual Evaluation model		
QP	Bit-rate	PSNR	Bit-rate	PSNR	Reduction
24	2912.886	39.982	2732.374	39.502	6.20%
26	2326.4	38.459	2172.942	37.976	6.60%
28	1885.835	37.099	1750.965	36.56	7.15%
30	1463.598	35.434	1355.935	34.943	7.36%
32	1145.418	33.92	1059.822	33.451	7.47%
34	902.496	32.619	823.108	32.103	8.80%
36	680.473	31.179	624.594	30.701	8.21%



(a) (b)

Fig. 3. Simulation result of perceptual evaluation model. (Stefan sequence) (a) The original frame. (b) Simulation result.



(a) (b)

Fig. 4. Simulation result of perceptual evaluation model. (Football sequence) (a) The original frame. (b) Simulation result.

[2] Hyeong-Min Nam, Keun-Yung Byun, Jae-Yun Jeong, Seung-Jin Baek, and Sung-Jea Ko, "Perceptual quality-complexity optimized video playback on handheld devices," in *Proc. IEEE International Conference on Consumer Electronics*, pp. 509-510, Jan. 2010.

[3] Zhenzhong Chen,Weisi Lin, and King Ngi Ngan, "Perceptual video coding: challenges and approaches," in *Proc. IEEE International Conference on Multimedia and Expo*, pp.784-789, Jul. 2010.

[4] C. H. Chou, Y. C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467-476, Dec. 1995.

[5] Jianwen Chen, Jianhua Zheng, and Yun He, "Macroblock-level adaptive frequency weighting for perceptual video coding," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, , pp.775-781, May. 2007.

[6] M. Naccari and F. Pereira, "Advanced H.264/AVC-based perceptual video coding: architecture, tools, and assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp.766-782, Jun. 2011.

[7] ISO/IEC JTC1/SC29/WG11, "Subjective test results for the CfP on Scalable Video Coding Technology," Doc.W6383, Munich, Mar. 2004.

Haptic Interaction with User Manipulation for Smartphone

Jong-uk Lee, Jeong-Mook Lim, Heesook Shin, and Ki-Uk Kyung
Electronics and Telecommunications Research Institute, Daejeon, Korea

Abstract—**This paper presents haptic interaction design and implementation for our designed smartphone bumper case [1] providing an interactive and a realistic physical feeling. The thin actuator is installed in the case to simulate a rapid realistic response. We designed a software structure guaranteeing a real-time physical response. The designed API can be used to provide realistic touch responses corresponding to an interactive physical feeling during gaming applications corresponding to sound effects.**

I. INTRODUCTION

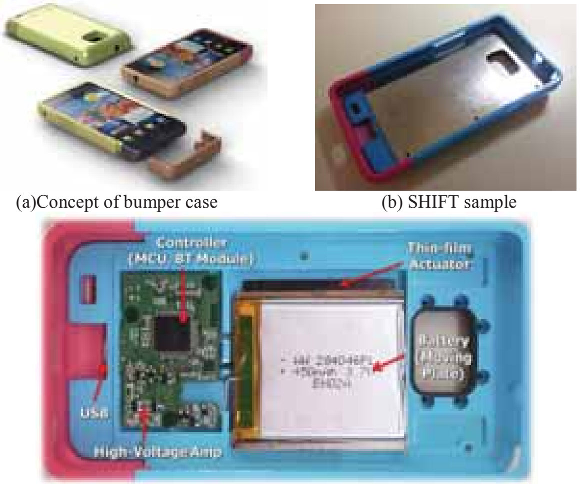
As touchscreen interfaces are widely adopted, users flexibly utilize the GUI on the touchscreen. However, the performance of the touchscreen input for precise or rapid tasks such as key typing is slower compared to the use of an actual typing due to the lack of sufficient feedback. Therefore, the effectiveness of a touchscreen input with additional sensory feedback such as auditory or haptic feedback has been investigated. It is showed that users can more immediately react to tactile feedback than to auditory feedback and that the accessibility, operability, and usefulness of tactile feedback are superior to these features when auditory and visual feedback are used [2]. Thus, tactile feedback from the touchscreen is useful, and if visual feedback and auditory feedback are combined, it provides reasonable user friendliness. Methodologies to provide haptic feedback for touchscreen devices have been suggested with a variety of features beyond the classical use of vibrating motors [3]-[5]. Poupyrev et al. designed a piezoelectric bimorph stack actuator and embedded tactile equipment in a PDA and simulated realistic clicking sensations [6]-[7]. Although their work could not be applied to a commercial device, it proved the value of tactile feedback. Another approach used audio haptic effects in a mobile phone [8]. It focused on only an analog audio signal with different materials and did not address GUI interactions, but it demonstrated that the conversion of audio into touch may be effective.

The present paper shows that haptic interaction for GUI manipulations in the smartphone is very effective. In this paper, we present to haptic interaction design.

II. SYSTEM DESIGN AND IMPLEMENTATION

A. Hardware Design and Implementation

As shown in Fig. 1, the proposed system is composed of a controller board, a thin-film actuator and a moving plate. The controller board includes a microprocessor, a Bluetooth communication module, and a high-voltage amplifier. The high-voltage amplifier allows 0~3.3V input and generates 0~1320V with current damage protection to supply the driving voltage of the film actuator.


(a)Concept of bumper case (b) SHIFT sample

(c) Inside structure
Fig 1. Bumper case prototype

B. Interaction Architecture

We designed a type of interaction architecture that synchronizes the haptic, auditory, and graphic responses. The manner in which the bumper case is actuated is determined by a touch or an application event primarily through a user action (touching the screen or pressing a button) or by messages.

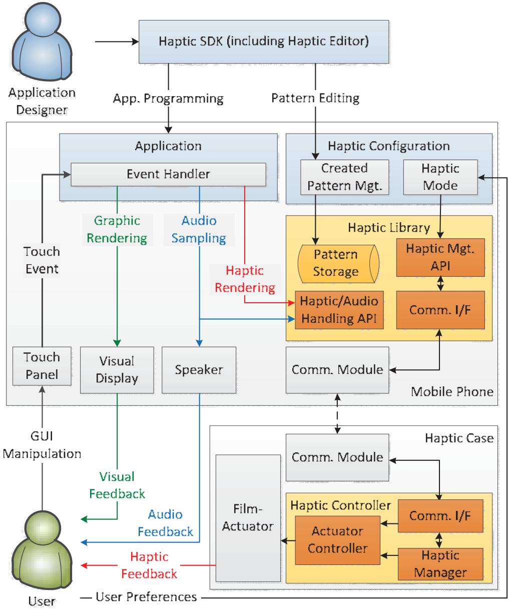


Fig 2. The interaction architecture for synchronized visio-haptic feedback

The haptic library consists of haptic event handling and event management. The haptic event handling is used to

This work has been supported by the S/W Computing R&D program of MKE/KEIT(10035360-2010-01, TAXEL: Visio-haptic Display and Rendering Engine).

control the bumper case. It performs to output real haptic feedback communicated with the bumper case. The haptic event management is used to add or remove haptic patterns in the bumper case. Adding haptic patterns is valuable to output the defined haptic patterns used frequently. Additionally, a pattern storage component maintains the defined patterns designed and commonly used by the provider as well as patterns created and designed by any designer.

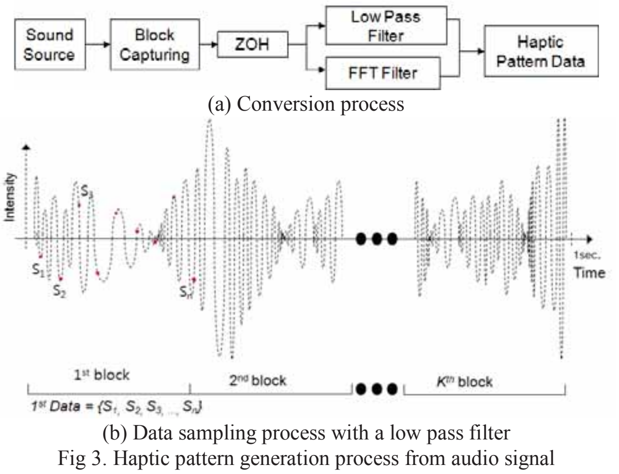
Based on the interaction architecture and the haptic library, synchronized visual-auditory-haptic feedback can be performed. They can sense haptic feedback while playing a game not developed by the haptic SDK using a method that converts sound information into haptic patterns in real time. The haptic library acquires the audio sampling information and generates a corresponding haptic pattern.

III. INTERACTION DESIGN

The haptic library supports the conversion of an audio source to tactile output. The conversion process is described in Fig. 3. To analyze the audio signal, the library captures the signal as a block. The block capturing rate is adjustable between 1~100 blocks/s. As the block capturing rate decreases, the time interval between the blocks increases and then causes a time delay. As the block capturing rate increases, there will be less of a time delay, but the conversion requires a high computation load in a smartphone. Because the allowable delay may depend on the application, we designed this so that the conversion library sets a variable k_c as the adjustable capturing rate.

When the audio signal is captured, the audio signal data is digitized by a zero-order-holder. The holding period T is also adjustable, and it determines the quality of the captured sound. The period is indirectly determined by setting a variable n as the number of divisions in a block. Therefore, T is determined as $1/(k_c \text{ x } n)$ seconds.

The frequency range of the audio signal is 20~20,000Hz and the haptic feedback frequency range is 0~300Hz. Because the frequency range of the audio signal is higher than that of the haptic feedback signal, it requires a frequency conversion. We utilize two conversion approaches.



(a) Conversion process

(b) Data sampling process with a low pass filter

Fig 3. Haptic pattern generation process from audio signal

A simple and effective means of doing this is to use a low-pass filter, as described in Fig. 3(b). Because the sampling frequency of the low-pass filter is lower than that of the audio signal, the sample data is sparsely saved in the form of S1, S2, …, Sn. The data are directly converted into the input signal for the haptic feedback. In this case, the haptic feedback focuses on the tendencies inherent in the sound changes rather than a precise reproduction. Practically, our actuator has a resonance frequency around 90~100Hz; the most efficient sampling frequency is 200Hz in terms of the output magnitude. This principle is useful for the conversion of arbitrary sounds containing rhythms.

If the audio effects mainly focus on a specific frequency range, a new conversion algorithm in the frequency domain becomes necessary. In this case, we can consider a FFT (Fast Fourier Transform) filter. For example, if we want to represent a bell ringing in the frequency range of 6~8 kHz, the intensity of the sound frequency from a FFT analysis is converted into 90~100Hz vibration, whereas a low-pass filter does not allow representation of such a high frequency. The haptic library allows the direct mapping of the sound frequency and the haptic feedback frequency.

IV. CONCLUSION

In this study, we proposed haptic interaction mechanism for mobile devices providing realistic and exciting haptic feedback. Our experimental results show that the proposed system performs better than a conventional haptic feedback method. The system is nearly ready to be used in consumer products. Moreover, the advanced and improved haptic library including several tactile patterns for user manipulation will improve the haptic interaction and help to develop applications using this accessory.

REFERENCES

[1] J.U. Lee, J.M. Lim, H.S. Shin, and K.U. Kyung, "SHIFT: Interactive Smartphone Bumper Case," *In the Proc. of the EuroHaptics 2012, Part II, LNCS 7283*, pp. 91-96, 2012.

[2] S.H. Kim and H.S. Kim, "The Haptic feedback of touch screen based mobile phone interface,*" In the Proc. of the 7th Intl. Conf. of Asia Digital Art and Design Association (ADADA 2009)*, pp. 108-111, 2008.

[3] P. Albinsson, and S. Zhai, "High precision touch screen interaction," *In the Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI 2003)*, pp.105 – 112, 2003.

[4] K.U. Kyung, and J.S. Park, "Ubi-Pen: Development of a Compact Tactile Display Module and Its Application to a Haptic Stylus," *In the Proc. of the World Haptics 2007: 2nd Joint Eurohaptics Conf. and Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp.109-114, 2007.

[5] L. Kim, W. Park, H. Cho, and S. Park, "A Universal Remote Control with Haptic Interface for Customer Electronic Devices," *IEEE Trans. on Consumer Electronics, Vol. 56, No. 2*, pp. 913-918, 2010.

[6] I. Poupyrev, M. Shigeaki, and J. Rekimoto, "TouchEngine: A Tactile Display for Handheld Devices," *In the Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI 2002)*, pp.644-645, 2002.

[7] I. Poupyrev, and S. Maruyama, "Tactile interfaces for small touch screens," *In the Proc. of the Symp. on User Interface Software and Technology (UIST 2003)*, pp.217-220, 2003.

[8] A. Chang, and C. O'Sullivan, "Audio-haptic feedback in mobile phones," *In the Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI 2005)*, pp.1264-1267, 2005.

Empirical Non-local Algorithm for Image and Video Denoising

Daehoon Kim, Byungjik Keum, Hyunchan Ahn and Hwangsoo Lee
Department of Electrinical Engineering, KAIST, Daejeon, South Korea

***Abstract*--This paper presents a new denoising algorithm capable of effectively removing noise without too much loss of original data. Based on the non-local means (NL-means) algorithm proposed by Buades et al., This approach uses a concept of thresholding weight of neighborhood pixels which have low similarity and applies the three-step search (TSS) technique in order to save computations.**

I. INTRODUCTION

With today's rapid advances in information technology, most people commonly use consumer electronic devices, such as web-cams and mobile phone cameras. Thus, stable and reliable noise reduction technologies for image and video have grown in importance [1]-[3]. To tackle edge preservation and aperture problem, non-local means (NL-means) denoising algorithm has been developed by Buades et al. [4]. The basic idea of this algorithm is to use properties that a lot of patterns are regularly repeated in natural images. Although this method is appropriate for region of repeated pattern, the details and fine structures are removed together with noise in the region of non-repeat pattern. In addition, this method often needs large amounts of computation time, which make it unsuitable to use in practical environments [3]. We present a modified denoising algorithm for image and video based on the NL-means method [5] by developing a weighting strategy that adjusts weights according to the window's similarity and introducing the three-step search (TSS) scheme to reduce unnecessary arithmetic operations with marginal weights.

II. PROPOSED DENOISING ALGORITHM

A. *NL-means denoising*

The NL-means algorithm can be represented as the following pixel intensity equation $I(\mathbf{x}, t)$ which has an integral formula in the space and time domain,

$$I(\mathbf{x},t) = \frac{1}{N(\mathbf{x},t)} \int_R \int_\Omega w(\mathbf{x},t,s) I(\mathbf{y},s) d\mathbf{y} ds. \tag{1}$$

where $N(\mathbf{x},t)$ is the normalizing factor, t is the current frame index, s is the neighborhood frame index, $\mathbf{x}$ is the current spatial location, $\mathbf{y}$ is the spatial location of the neighborhood pixels, Ω and R are the searching window in each spatial and temporal domain, respectively. The weight function w is represented as

$$w(\mathbf{x},t,s) = e^{-\frac{Q(\mathbf{x},t,s)}{h^2}}. \tag{2}$$

where h is the filtering parameter varying according to noise level. The exponential term, $Q(\mathbf{x},t,s)$ is expressed by the weighted sum of the Euclidean distance for each pixel between two neighborhood blocks as

$$Q(\mathbf{x},t,s) = \int_\lambda G_\rho \left| I(\mathbf{x}+\mathbf{u},t) - I(\mathbf{y}+\mathbf{u},s) \right|^2 d\mathbf{u}. \tag{3}$$

where G_ρ is the 2D Gauss kernel with standard deviation ρ , $\mathbf{u}$ is the spatial location, and λ is the neighborhood square.

B. Weighting strategy

In our proposed algorithm, the key is to limit the weight calculated by (2) and (3) for all neighborhood pixels in the search window. Our weighting strategy follows as

$$w(\mathbf{x},t,s) = \begin{cases} 1 & ,T_2 \le w(\mathbf{x},t,s) \\ w(\mathbf{x},t,s) & ,T_1 \le w(\mathbf{x},t,s) < T_2 \\ 0 & ,otherwise \end{cases}. \tag{4}$$

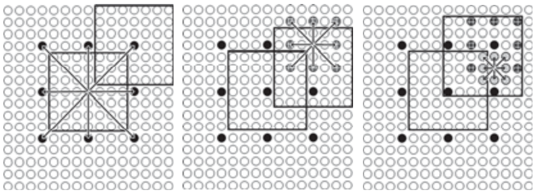
where T1 and T2 are the low and high threshold, respectively. Such constraint leads that weights being less than T1 convert into zero, which may be thus interpreted as not being involved in the weighted averaging while one being greater than T2 convert into the largest weight. Note that T1, T2 are fixed, h is controlled depending on the noise level.

The performance of our algorithm by using this strategy is similar to one of the NL-means method in the flat and edge region, but original information loss is reduced for details and fine structures in the non-repeat and unique region. Moreover, it is possible to improve the processing time because weights set to zero are not required to include in computation process. To skip such unnecessary operation, we are able to introduce the TSS [6] technique described in the next section

C. Three step search(TSS)

Fig. 1 shows an example to illustrate the procedure of TSS (W = 7, W is maximum distance in vertical and horizontal directions from center pixel) on condition that the size of the neighborhood square is 7 x 7, and the search window size is 15 x 15 in the spatial domain. In the first step, similarity between the neighborhood square around each of the nine points (including a center point) and one around a center pixel are computed, and then the process of weighting strategy is applied. Only the large points than T_1 are considered for search process in the next step. For the second and third steps, the number of checking points is eight excluding the location which is already checked in the previous step. Note that similarity between each of checking points is calculated by

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a) 1st step (b) 2nd step (c) 3rd step

Fig. 1. The TSS procedure. (a) An example for one direction in the 1st step. (b) An example for one direction in the 2nd step after (a). (c) An example for one direction in the 3rd step after (b).

comparing (7 x 7) patch around the points.

Notice that the number of steps can be changed according to W while the search window size is increased or decreased. In general, given W, the number of steps required is

$$M = [\log_2(W+1)]. \tag{5}$$

where $[x]$ is the ceiling function, which denotes the smallest integer not less than x. Consequently, the step-size (distance between points in a search step) for n-th step is given by

$$Ss(n) = 2^{M-n}. \tag{6}$$

It is shown that TSS exhibits simplicity and regularity by using a uniformly distributed search pattern in each step. For W = 7, while the total number of checking points for FS is 225, the speed up ratio for TSS depends on pattern's repetitiveness and regularity in sample data.

III. EXPERIMENTAL RESULT

All experiments are performed on the condition that T_1 and T_2 are set to 0.1 and 0.5 respectively in this section. Empirically, when T_2 is from 0.4 to 0.6, our algorithm gives remarkable visual quality compared with the NL-means method, and supports high PSNR value for most of the samples.

In simulation to compare performance, the NL-means algorithm uses conditions that the size of the neighborhood square is 7 x 7, and the search window size is 21 x 21 as suggested in Buades'paper [5]. Comparison with the NL-means method and the proposed algorithm for the Lena image is presented in Fig. 2. Although our result is slightly lower than the result of the NL-means method in PSNR as shown in Table I, the original information is preserved better in the non-repeat and unique region. In Fig. 3, the algorithm is applied to a video sample produced by the 7D DSLR camera working in high ISO mode and in low light conditions. A frame for the video denoising result shows that relatively strong noise is reduced, and block artifact due to compression is removed without losing too many details and fine structures. Table II, shows that the proposed algorithm using TSS is approximately 3 times faster than the original NL-means algorithm using FS, and yet produces better visual quality Improvement of computational complexity varies with the noise level. Experiments are done on a PC with 2.4GHz Core 2 CPU and 3.00G RAM using Visual studio 6.0 software.

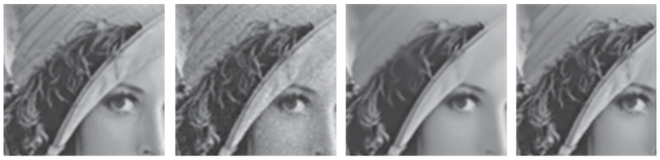


Fig. 2. Comparison with the NL-means method and proposed algorithm. From left to right: original image, noisy image (σ = 10), the result of NL-means (PSNR 35.41) and our result (PSNR 35.20).

TABLE I. Performance comparison of noise reduction in terms of the PSNR (UNIT : DB).

σ	5	10	15	20	30
NLM	39.53	35.41	32.80	31.88	29.85
proposed	39.34	35.20	32.53	31.72	29.64

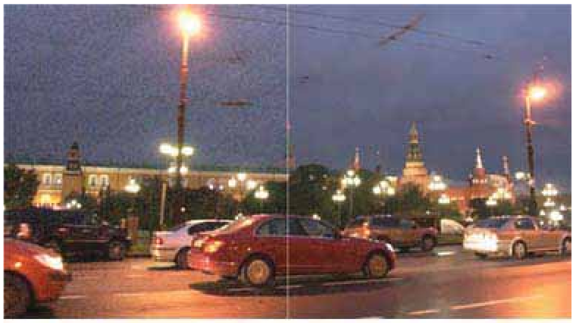


Fig. 3. Our denoising result for a real video (σ = 20). Block artifact can be removed duo to compression.

TABLE II. Comparison for processing time (σ = 10).

Image	size	NLM	Proposed	ratio
Boat	148×140	31 sec	6 sec	5.17
Lena	256×256	102 sec	31 sec	3.29
Camera man	512×512	142 sec	41 sec	2.94

IV. CONCLUSIONS

In this paper, we propose a new denoising algorithm based on the NL-means denoising by using the weighting strategy and TSS. On empirically selected parameters, experimental results demonstrate that the proposed method outperforms the NL-means method in the non-repeat and unique region with reduced computational complexity. The denoising results still satisfies with three principles (method noise, noise to noise, and statistical optimality) proposed in Buades'paper.

REFERENCE

[1] C. Liu and W. T. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," In Proc. *ECCV*, vol. 6313, pp. 706-719, 2010.

[2] H. Ji, C. Liu, Z. Shen and Y. Xu, "Robust video denoising using low rank matrix completion," In Proc. *IEEE CVPR*, pp. 1791-1798, June, 2010.

[3] V. Karnati, M. Uliyar and S. Dey, "Fast Non-Local Algorithm for Image Denoising," In Proc. *IEEE ICIP*, pp. 3873-3876, 2009..

[4] A. Buades, B. Coll and J.M. Morel, "A non-local algorithm for image denoising," In Proc. *IEEE CVPR*, vol. 2, pp. 60-65, 2005.

[5] A. Buades, B. Coll and J.M. Morel, "Nonlocal Image and movie denoising,*" Int. J. Comput. Vis.*, vol. 76, no. 2, pp. 123-139, 2008.

[6] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, pp. 438-442, Aug. 1994.

Using Remote Update Controller of FPGA's as Built-in Self Test for Embedded Systems

Christian Stoerte
Astro Strobel, Bergisch Gladbach, Germany

***Abstract*—This paper is an approach of a reliable built-in self test (BIST) on FPGA based embedded systems (e.g. digital TV receivers, compact head-ends, etc.), which could easily be executed by consumers. The proposed technology is quite similar to boundary scan testing via JTAG-Interface, which is a well-established method during production to find hardware errors on a circuit design at an early stage during the manufacturing process. In contrast to boundary scan the BIST is implemented as a part of factory image in state of the art FPGA's and could be activated by graphical user interface. Since the factory image part of the firmware will never be updated by consumers, the BIST is always available even if a previous application firmware update has been failed. This offers a comfortable customer service, because in case of any support questions hardware errors could be excluded before complex assistance for configurations and firmware updates begins.**

Keywords— Embedded System; Self Test; Remote Update

Introduction

Due to the downwards price trend over the last years for consumer electronic devices, circuit designs must be developed as low-cost as possible. International business competition has caused commercial launch even if the development of new products has not been completely finished. This trend not only requires the use of low-cost IC's on the design, even the possibility of subsequent firmware updates for consumers must been given. To find a solution for both requests lots of embedded designs for signal processing are based on FPGA's. The use of these devices brings the idea of new cost savings, which are not based on the product itself, but rather on additional support. In case of customer service we would like to find out quickly if the consumer uses a mismatched firmware or any wrong configuration or if he really has a hardware failure, which requires a replacement. It would be a good idea to check the complete hardware design with BIST, which could be executed by the customer.

Self Test During Production

During production this kind of hardware-self-tests are well-known to check the circuit design at an early stage during the manufacturing process. Two kinds of technologies are state of the art: Visual inspection and boundary scan testing. Visual inspection is an easy method to locate wrongly placed parts comparing a snapshot with a reference design. A video analysis displays all faults automatically. Nevertheless, this kind of hardware testing is not feasible for customers and also not necessary, because hardware errors on a once working design are electrical failures, which cannot be found by visual inspection. But how about boundary scan test? There are lots of commercial software available to check hardware with a boundary scan test via JTAG-Interface during production. Due to the fact that consumers normally do not have JTAG-Connectors to check their devices other solutions similar to boundary scan testing must be found. The best method could be an additional processor on the circuit, which could be invoked by the user interface. If the processor finishes the BIST, it sends the result back to the main processor and with it to the user interface. One way of cost-effective solution for an additional processor is to use application logic of FPGA's, which is reprogrammed during self testing. This paper shows how remote update controllers in state of the art FPGA's could be used to find an efficient solution for BIST, which could easily be executed by customers via graphical user interface.

Implementation of Self Test

The following chapter gives an approach of how to implement self-testing methods without losing performance for the customary application. Using Altera® FPGA's considering device family support a factory configuration in remote mode could be used [1]. On power-up the device always loads a factory file from a specific flash position, which is typically located at the beginning of the flash content. This image is usually programmed during the manufacturing process and will never be changed by the customer. After a couple of seconds the remote controller starts a watchdog counter and reconfigures the FPGA. The application configuration for the customer is loaded from another predefined flash memory location. If the customer deadlocks the application processor with a mismatched firmware update, the factory image could be used as fallback solution. In case of a watchdog timeout during the reconfiguration, a corrupted firmware is detected and the factory image is reloaded. This enables the customer to re-load firmware even if a previous update has been failed. In summary the factory image includes a very small integrated processor with on

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

chip memory, the interface to the user logic and a connection to the flash memory device. All other functionality for the whole customer design with its corresponding software is located in the application image, which gives us enough FPGA logic in factory configuration to realize any BIST. Additionally all user inputs and outputs (I/O's) could be defined different from application image. The following chapter gives an example of a kind of factory image build-in self test which must always be fitted to the used hardware.

Build-In Self Test Example

An example for a BIST of a complete system with its connection to the user interface is given in Fig. 1. One of the most important devices to check during self test is an external memory, which is usually connected to the FPGA. Memory faults are divided into simple and linked faults. Algorithms to detect linked faults are very complex since the error pattern depends on the appearance of other faults. Due to the limited time while factory image with its self test is in progress, it is impossible to find these faults. Unlike simple faults, which could be found more quickly. In case of consumer self test a precise localization of the error pattern is not necessary. A good compromise between test time and error detection could be March algorithm[2] with the following read (*r*) or write (*w*) operations in upwards (↑) or downwards (↓) address order:

$$\{\updownarrow(w0);\uparrow(r0,w1);\downarrow(r1,w0);\ \updownarrow(r0)\}$$

This kind of self test is able to detect, but not to locate the following faults: address decoder faults (AF), stuck-at faults (SAF), coupling faults (CF) and transition fault (TF). In case of word-oriented memories are used, it is possible to improve test time by use of special march test [2]. Additionally, improvements could be made for self tests of memories with burst-oriented access (e.g. DDR SDRAM)[3]. Another important part of BIST is the detection of SAFs, SOFs and CFs at any I/O pin of the FPGA. Using programmable features of I/O pins, such as pull-up resistors and bidirectional data paths, a pattern generator could be used to output test sequences. After this, all outputs are set to high-impedance and the inputs are monitored [4][5]. In this way, a response analyzer is even able to find some SOFs since a missing input capacitance of normally connected IC's could be detected. All other peripheral devices could be self tested by polling the corresponding ports and monitoring their response.

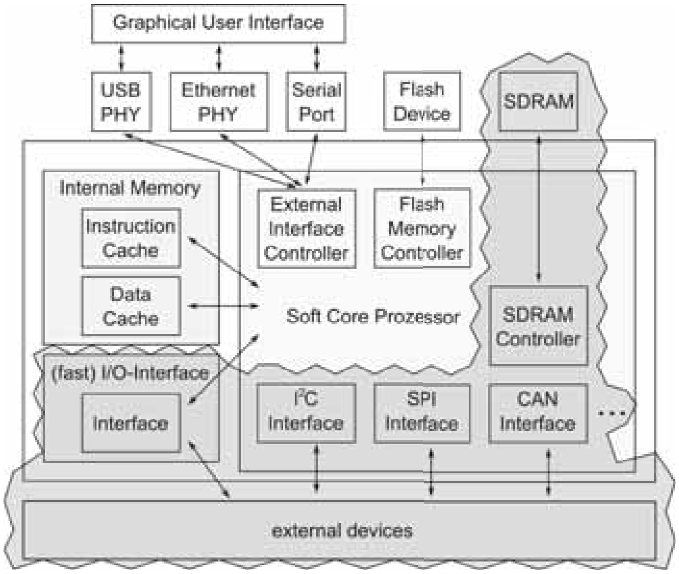


Fig. 1: BIST of embedded system and its peripheral

It is even recommended to toggle I/O connection lines between management FPGA and connected IC's, if the corresponding device supports general purpose I/O's (GPIOs). If the state of any GPIO of peripheral devices could be set via commands over their serial interface, this is a very useful BIST to find open netlines on the circuit design.

Conclusion

State of the art consumer products (e.g. digital TV receivers, compact head-ends, etc.) are always featured with interfaces allowing subsequent firmware updates of their devices. To prevent customers from a deadlocked system due to any mismatched firmware update, soft-core processors are equipped with special controllers. This paper has shown how factory configuration of these remote controllers could additionally be used as BIST for embedded systems. An approach for one kind of self test application has been given, which always must be adapted by developers to the used hardware design.

References

[1] Altera® Corporation, "Remote System Upgrade", San Jose, 2012

[2] A.J. van de Goor and I.B.S. Tlili,"March tests for word-oriented memories" Proceedings of the conference on Design, automation and test in Europe, pp.501-509, 1989

[3] André B. Soares, Alexandro C. Bonatto and Altamiro A. Susin, "A new March Sequence to Fit DDR SRAM Test in Burst Mode", ACM Press, pp. 28-33, 2008

[4] Stroud, C., "Built-in self-test of FPGA interconnect", Proceedings of the Test Conference, pp. 404-411, 1998

[5] Sudheer Vemula and Charles Stroud, "Built-In Self-Test for Programmable I/O Buffers in FPGAs and SoCs, 2006

Influence of Clock Structure on EMI Level and Audio Clock Jitter During High-speed Serial Transmission

Tsuyoshi Ikushima[1], Tsutomu Niiho[1], Osamu Shibata[1], Syuji Kato[2], Naoshi Usuki[3]
[1] Material & Process Development Center, Panasonic Corporation, Osaka, Japan
[2] Industrial Device Company, Panasonic Corporation, Kyoto, Japan
[3] AVC Networks Company, Panasonic Corporation, Osaka, Japan

Abstract - **In this study, we investigated the influence of the separate-clock structure and embedded-clock structure on the characteristics of electromagnetic interference (EMI) and audio clock jitter during high-speed serial transmission. We found that the EMI peak spectrum of a separate clock can be reduced by 11 dB if its frequency is decreased. The peak spectrum of transmission when using a separate clock is much lower than that of transmission when using an embedded clock, and results in less audio-clock jitter than transmission using an embedded clock.**

I. INTRODUCTION

The resolution of audio/video signals for TV sets is continuing to increase, and consequently the bit rate necessity for audio/video signal transmission is also certain to increase[1]. High-speed serial transmission of uncompressed audio/video data, such as via HDMI (High-Definition Multimedia Interface), will therefore require a higher bit rate.

For serial transmission, there are two transmission structures for the video clock [2] [3]. The first is a structure that transmits a video clock signal separately from audio/video data such as HDMI. This is called the “separate clock” structure. The other structure does not transmit a video clock signal directly, since it is embedded in the audio/video data, such as DisplayPort. This is called the “embedded clock” structure.

One of the aims of the embedded-clock structure is to reduce the EMI level by avoiding the use of a clock line [4]. On the other hand, decreasing the clock amplitude reduces the EMI level attributable to the clock signal [5]. A clock generates many harmonics, so EMI noise is likely to appear in wider and higher frequency ranges, such as the GHz wireless band, if the clock frequency is high.

Audio clock jitter is another major concern in these clock transmission structures, when audio clock is recovered from video clock in a receiver. Different video clock structures show different jitter values. The audio clock signal is not separately transmitted: it is recovered by dividing the video clock signal in a receiver circuit [3]. This means that audio clock jitter also varies according to the clock structure.

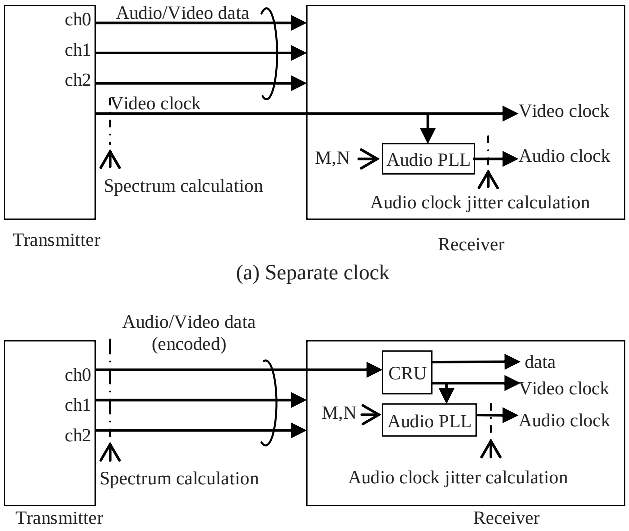
In this paper, the influences of clock structure and clock parameters on both EMI level and audio clock jitter are studied in a high-speed serial transmission environment. First, clock structures and simulation models are explained. Next, for the purpose of estimating EMI level, the spectrum of transmitted signals is simulated. Finally, simulation of audio clock jitter is examined.

II. CLOCK TRANSMISSION STRUCTURE

Fig. 1 shows two structures for transmitting a video clock signal. In both structures, audio/video data is transmitted via multiple channels. As shown in Fig. 1 (a), the separate clock transmits a video clock signal separately from the audio/video data. On the other hand, the embedded clock sends a video clock signal that is embedded in the audio/video data using specific encoding technology, as shown in Fig. 1 (b). At the receiver circuit shown in Fig. 1(b), the video clock signal is recovered from the video data using a clock recovery unit (CRU). In both structures, an audio clock is generated from the video clock using an audio phase-locked loop (PLL) in the receiver LSI. The audio clock frequency f_s and the video clock frequency f_{video_clk} are related as follows [2]:

$$128 * f_s = \frac{N}{M} \cdot f_{video\_clk} \quad , \tag{1}$$

where N is natural numbers and M is averaged sampling value.



(a) Separate clock

(b) Embedded clock

Fig. 1. Clock transmission structures for high-speed serial transmission.

III. SIMULATION MODEL

In this paper, we assume the resolution of the video data to be 4K2K, the frame rate 60p, the color depth 24 bits, and the number of audio/video data channels three. The bit rate for full HD video resolution requires approximately 1.5 Gbps/channel in HDMI format. Since the transmission speed for 4K2K video resolution is needed four times faster than that for full HD video resolution, the bit rate of the audio/video

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

data will be approximately 6 Gbps/channel for 4K2K resolution. In the separate-clock structure, to synchronize the video clock and the video data, the video clock frequency is usually set to 600 MHz, which is 1/10 of the transmission rate because one transmission character is consisted by 10 bit data. Additionally 150 MHz, which is 1/40 of the transmission rate, is evaluated for improvement. In the embedded-clock structure, the frequency of the video clock is set at 600 MHz.

The simulation model parameters are shown in Table 1. The amplitude of the video clock is normalized relative to the amplitude of the audio/video data.

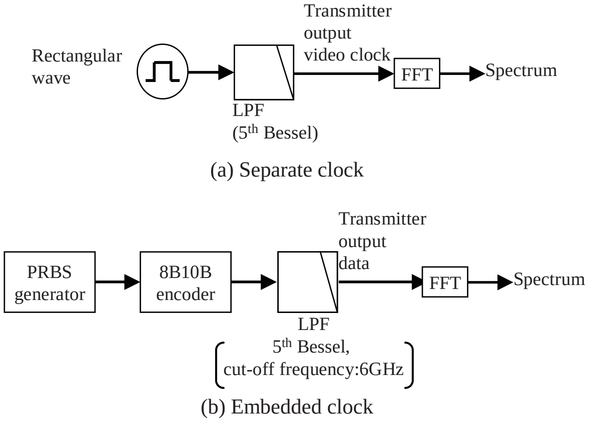
Table 1. Simulation model parameters

	Video clock transmission structure	Video clock parameters		
		Frequency	Amplitude	Rise time
1a	Separate clock	150 MHz	1	100 ps, 200 ps, 300 ps
1b		600 MHz	1	100 ps, 200 ps, 300 ps
1c			1/3	
2	Embedded clock	600 MHz	—	—

A. Influence on EMI

The EMI level for each spectrum is correlated with the amplitude of the transmitted signal current. We therefore simulated the spectrum of the signals at the output of the transmitter for the video clock for the separate-clock structure and for the audio/video data for the embedded-clock structure.

Fig. 2 shows models of the spectrum simulation. Fig. 2 (a) shows the model for the separate-clock structure. The input rectangular wave is filtered by a 5th-Bessel low-pass filter (LPF), whose cutoff frequency is determined by the rise time shown in Table 1. The model for the embedded-clock structure is shown in Fig. 2 (b). In this case, it is used for 8B10B encoding, and the Pseudo Random Bit Sequence (PRBS) pattern is inputted into an 8B/10B encoder. The length of the PRBS pattern is 2^{16}–1. The encoded data is filtered with a 5th-Bessel LPF with a cutoff frequency of 6 GHz. In both structures, the spectrum is calculated by Fast Fourier Transform (FFT) of the output time-domain signal.



(a) Separate clock

(b) Embedded clock

Fig. 2. Spectrum simulation model

B. Influence on audio clock jitter

Random jitter, included in the audio/video data or the video clock at the receiver circuit, causes audio clock jitter. The deviation of the division ratio (M) in Equation (1) may also cause significant jitter [6]. However, this problem can be reduced by improved implementation of the circuit. Therefore, in this paper, we ignore deviation of the division ratio and use averaged value of the division ratio. Fig. 3 shows video clock jitter models input to the audio PLL: 3 (a) shows the video clock which is input into the audio PLL for the separate-clock structure. It is generated as follows.

- As a clock, a rectangular wave is input into LPF, whose cutoff frequency is determined by the rise time.
- As transmitter jitter, random-frequency noise is added to the wave sent from the LPF. The amplitude of the frequency noise is equivalent to 41.7-ps jitter, which is 0.25 Unit intervals of one bit of the audio/video data.
- The wave transmits to a filter whose loss value is the same as that of a HDMI cable. It is given by

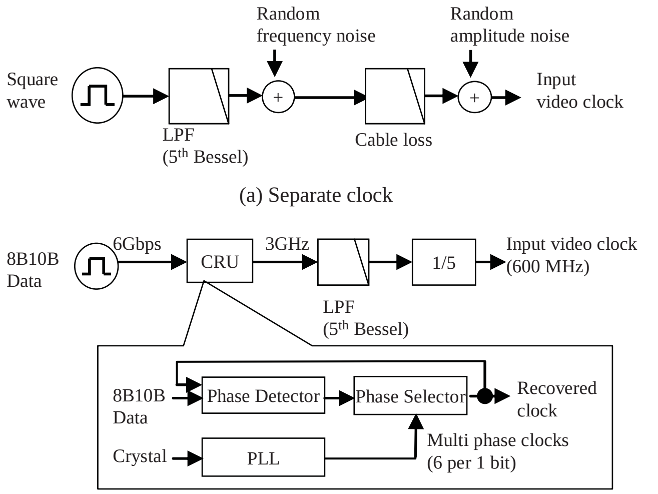
$$Loss = -4.85\times10^{-9}\cdot f\ \ (\mathrm{dB}), \qquad (2)$$

  where f is frequency.
- Finally, random-amplitude noise is added as jitter to a transmission line. When the rise time of the clock wave is 100 ps and the amplitude of the clock wave is the same as that of the data, the amplitude of the frequency noise is equal to 83.3-ps jitter, which is 0.5 Unit intervals of one bit of the audio/video data.

For the embedded-clock structure, the video clock input into the audio PLL is shown in Fig. 3 (b). It is generated as follows.

- 8B/10B data is input into a clock recovery unit (CRU). The 8B/10B data includes 83.3-ps jitter, which is 0.5 Unit interval of one bit of the audio/video data.
- The CRU uses a multi-phase PLL. The CRU selects a clock from the 6-phase clocks whose frequency is 3 GHz. The clock edge is nearest to the center of the eye diagram of the input data. The CRU then outputs the selected clock signal as a recovered clock.
- An LPF eliminates the high-frequency jitter included in the recovered clock signal.
- To make a 600-MHz clock, the recovered clock frequency is divided by 5.

Fig. 4 shows a simulation model for audio clock jitter. In this model, the jitter of the input clock is calculated first. The input jitter is represented in UIs (Unit Intervals), which express the ratio of jitter to clock period. Next, the jitter spectrum, X(s), calculated using FFT, is multiplied by the transfer function of the audio PLL. Finally, the output jitter is calculated using inverse FFT (IFFT). The audio PLL includes an input divider whose division ratio is M, a phase frequency detector (PFD), a second order LPF, a voltage-controlled oscillator (VCO), and a feedback divider whose division ratio is N. The cutoff frequency of the audio PLL is 100 Hz.



(a) Separate clock

(b) Embedded clock

Fig. 3. Video clock jitter model

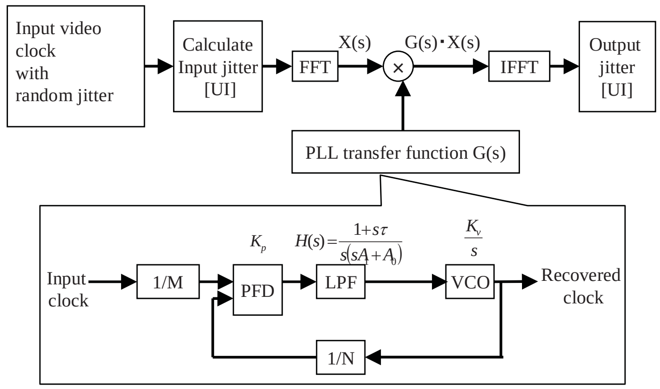

Input video clock with random jitter
Calculate Input jitter [UI]
FFT
X(s)
G(s)・X(s)
IFFT
Output jitter [UI]
PLL transfer function G(s)
Input clock
1/M
K_p
PFD
$H(s)=\frac{1+s\tau}{s(sA+A_0)}$
LPF
$\frac{K_v}{s}$
VCO
Recovered clock
1/N

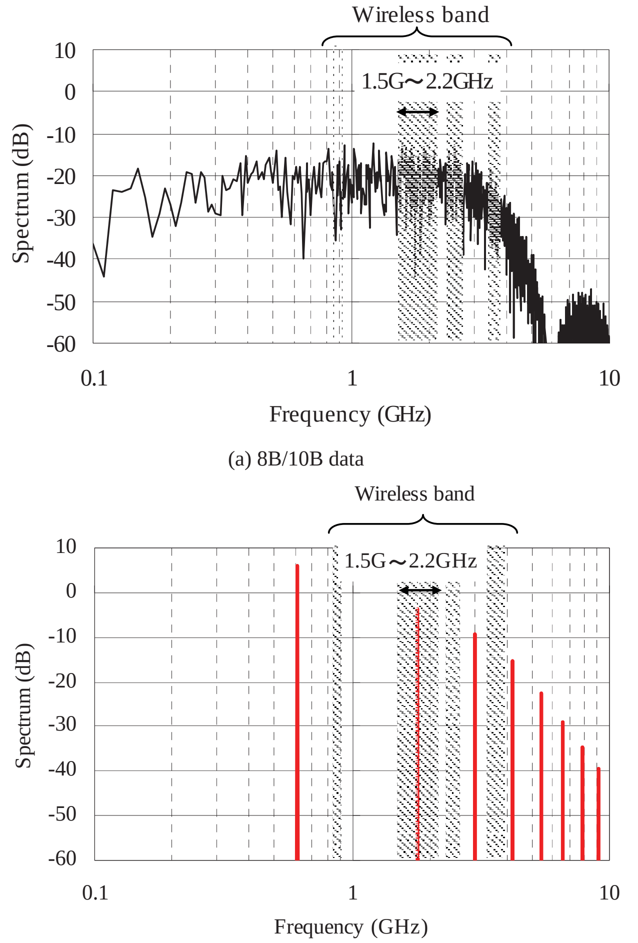
Fig. 4. Simulation model for audio clock jitter

IV. SIMULATION RESULTS

A. EMI Level

Fig. 5 (a) and (b) show the calculated spectrum of 8B/10B data, and that of the video clock where the clock frequency is 600 MHz, and the rise time is 100 ps.

Since interference causes problems chiefly in wireless signals, we focus on the peak spectrum in the wireless band, shown as shaded portions in Fig. 5. The spectrum of the video clock has a noticeably large peak between 1.5 GHz and 2.2 GHz. We therefore simulated the peak spectrum in the range from 1.5 - 2.2 GHz using the parameters shown in Table 1. Fig. 6 illustrates the results. The peak level of spectrum for 8B/10B data is –12.6 dB. The 600-MHz video clock, whose amplitude is 1 and rise time is 100 ps, has a larger peak spectrum than that of 8B/10B data. Although lengthening the rise time reduces the spectrum peak, the spectrum peak is still higher than that for 8B/10B data. Decreasing the clock amplitude and decreasing the clock frequency are more effective. For example, when the rise time is 100 ps, decreasing the clock frequency lowers the spectrum peak by 11 dB. When the clock frequency is decreased to 150 MHz or the clock amplitude is decreased to 1/3, the spectrum peak falls below that of the 8B/10B data.



(a) 8B/10B data

(b) Video clock (clock frequency 600 MHz, rise time 100 ps)

Fig. 5. Calculated spectrum of 8B/10B data and video clock

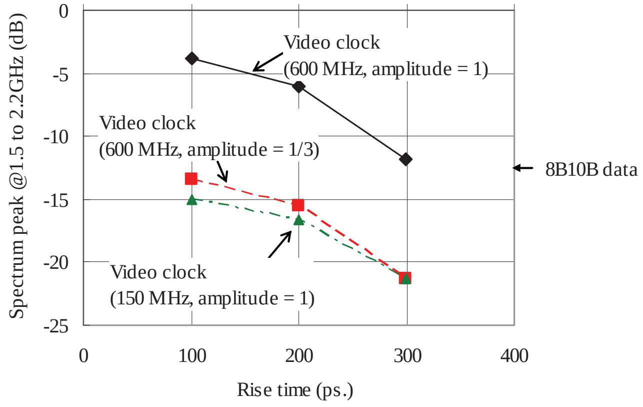
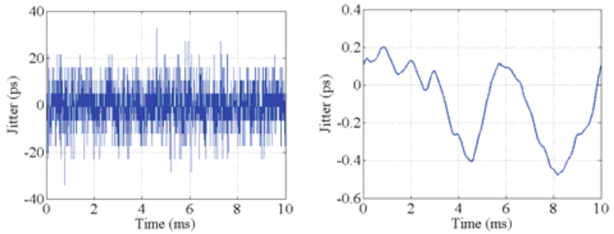


Fig. 6. Spectrum peak of 8B/10B data and video clock (1.5 - 2.2 GHz)

B. Audio clock jitter

Fig. 7 shows an example of calculated input and output jitter of the audio PLL for the separate-clock structure. In Fig. 7, the clock frequency is 600 MHz, amplitude is 1, and rise time is 200 ps. We simulated the peak-to-peak value of output jitter, which is equal to audio clock jitter, using the parameters shown in Table 1. Fig. 8 shows the results. The jitter in the embedded clock is 1.71 ps. For the separate-clock structure, although the increased rise time and decreased clock amplitude effectively reduce the EMI level, both increase audio clock jitter. However, the decreased clock frequency has the effect of limiting audio clock jitter, allowing it to keep jitter below that from the embedded clock.



(a) Input jitter (b) Output jitter

Fig. 7. Audio PLL input and output jitter
(600 MHz video clock, amplitude 1, rise time 200 ps)

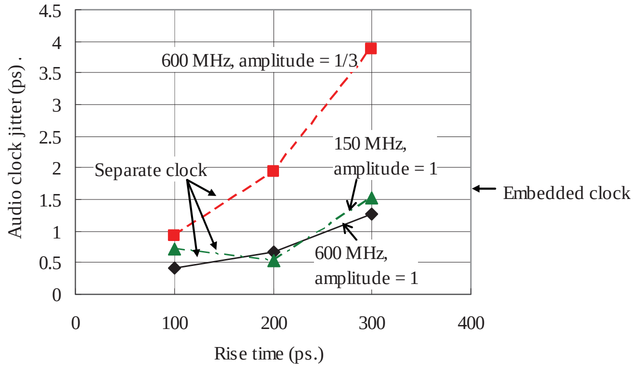


Fig. 8. Simulation results of audio clock jitter

V. CONCLUSION

In this paper, we investigated the influence of clock transmission structure and video clock parameters on EMI level and audio clock jitter during high-speed serial transmission.

We compared the separate clock and the embedded clock structures by selecting several clock parameters such as frequency, amplitude and rise time. First, to estimate the EMI level in the wireless band, we simulated the transmitter output spectrum for 8B/10B data and a video clock. Our results showed that decreasing the video clock amplitude or decreasing the video clock frequency can effectively reduce peak spectrum to 11dB. The peak is smaller than that for 8B/10B data used in the embedded-clock structure.

Next, audio clock jitter was simulated for both structures. Audio clock jitter was assumed to be caused by random jitter present in the received signal. Since the audio clock signal is recovered from the video clock signal, audio clock jitter is influenced by the video clock transmission structure. In the separate-clock structure, it was shown that audio clock jitter increases with increased rise time and decreased video clock signal amplitude. However, decreasing the clock frequency only slightly increases audio clock jitter.

In conclusion, after optimizing the parameters of the video clock, the separate-clock structure can show a better EMI level and audio clock jitter performance than the embedded-clock structure.

REFERENCES

[1] S. Namiki, T. Kurosu, K. Tanizawa, J. Kurumida, T. Hasama, H. Ishikawa, T. Nakatogawa, M. Nakamura, K. Oyamada, “Ultrahigh-Definition Video Transmission and Extremely Green Optical Networks for the Future”, IEEE J. Sel. Top. Quantum Electron., vol.17, pp.446-457, 2011

[2] HDMI Specification 1.3, June 2006

[3] X. Zhang, S. Zhai, Y. Wang, “Stream Clock Recovery in High Definition Multimedia Digital Systems”, United States Patent Application No. US2011/0075782, 2011

[4] Y. Kim, J. Song, W. Heo, C. Kim, “An Efficient Architecture of Encoder and Decoder for DisplayPort Physical Layer,” ICCE 2009, No.10.2-3, 2009

[5] K. Lee, Y. Shin, S. Kim, D. Jeong, G. Kim, B. Kim, V. Costa, “1.04 GBd Low EMI Digital Video Interface System Using Small Swing Serial Link Technique,” IEEE J. Solid-State Circuits, vol.33, pp.816-823, 1998

[6] H. Wang, “Audio Clock Regenerator with Precise Parameter Transformer,” United States Patent Application No. US2009/0167366, 2009

Wirelessly controlled LED Fixture with Heat sink – Design and Implementation

G Dhivya, K Subaashini, J Shamshudeen, G Rekha and R Pitchiah
Centre for Development of Advanced Computing (C-DAC), Chennai, India

Abstract-- **This paper presents the design, implementation and deployment of thermally stable LED fixtures that can be controlled wirelessly using ZigBee. To improve the longevity of LEDs, parameters like LED disconnection, driver feedback voltage and temperature of the designed LED fixtures have been monitored and controlled remotely. Energy saving results shows that the LED fixtures designed can be used as a replacement for T8 fluorescent lamps.**

I. INTRODUCTION

The future of lighting will be based on the use of energy efficient LED lamps. Heat sink [4] helps to keep devices at a temperature below its specified maximum operating temperature. So, the design of heat sink is essential because the life of LEDs depend on how efficiently the heat is being dissipated. We have designed and developed an illumination system using LEDs which is wirelessly monitored and controlled by using ZigBee [5]. Our contributions differ from [1], [2] as follows:

- A ZigBee controlled dimmable LED driver circuit and heat sink for LEDs has been designed, implemented and deployed in C-DAC, Chennai.
- Developed TinyOS 2.x [6] components for continuous wireless monitoring of the LED fixtures

II. DESIGN

A. LED Fixture - Hardware Design

A dimmable LED fixture with wireless control has been developed for smart building environment. A LED driver circuit has been designed in order to provide proper starting voltage and to regulate the current flow through the LED and also to protect it from voltage fluctuations. Dimming of LEDs is done by varying the width of pulses (PWM). The advantage of PWM dimming is that it enables dimming with minimal color shift in LED output. The overall block diagram and specifications of LED fixture designed by us are shown in Fig 1 and Table I respectively.

Each LED fixture has a ZigBee End Device (ZED) to control its operation. The LED lamps are connected to the ZED through the LED Driver. ZED in the fixture operates with 5V DC power supply which has been designed by us. 3 LED drivers are used to drive 42 discrete LEDs. Each LED driver drives 14 LEDs. The viewing angle of a LED is 170°. Prismatic lens has been used to distribute the illumination further. A ZigBee Coordinator (ZC) connected to the server controls the operation of LED fixture based on the decision taken by illumination control algorithm. A wireless sensor mote with CC2430 Radio, 8051microcontroller core and onboard temperature sensor has been used as ZC and ZED.

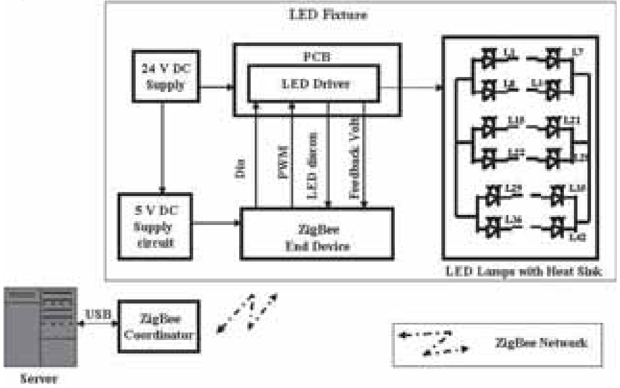


Fig. 1 Overall block diagram

TABLE I LED FIXTURE SPECIFICATIONS

Wattage	56 W
Light output	4267 lumens
Operating Voltage	24 V DC
Color	Cool white
Heat sink type	Extruded type (Natural Convection)
Fixture Size	56.5 x 56.5 cm

Thermal resistance (R_{th}) is the measure of the heat dissipation capability of a surface. Based on the details from the LED data sheet the heat sink has been designed.

$$R_{th\ (j\text{-}a)} = R_{th\ (j\text{-}c)} + R_{th\ (c\text{-}a)}$$

$R_{th\ (j\text{-}c)}$ is Junction to Casing Thermal Resistance, $R_{th\ (c\text{-}a)}$ is Casing to Ambient Thermal Resistance (Thermal Resistance of Heat Sink) and $R_{th\ (j\text{-}a)}$ is Junction to Ambient Thermal Resistance.

$$R_{th\ (j\text{-}a)} = (T_j - T_a) / (V_f * I_f * \text{No. of LEDs})$$
$$= (135 - 30)/ (3.0*0.35*14) = 7.14\ ^{o}C/W$$
$$R_{th\ (c\text{-}a)} = 7.14 - 6.5 = 0.64\ ^{o}C/W$$

$R_{th\ (heat\ sink)}$ should be less than 0.64 ^{o}C/W. Convective Heat Transfer Coefficient of air (h_c) is 10 W/ (m²K). Q_1 is the heat transferred by Natural Convection.

Surface Area of Fins (A_s) = fin count*length * height

$$Q_1 = h_c * A_s * (T_j - T_a) = 183.75\ W$$
$$R_{th\ (heat\ sink)} = (T_j - T_a)/ Q_1 = 0.57\ ^{o}C/W$$

Since, thermal resistance of the heat sink is less than 0.64 ^{o}C/W the heat sink design is optimal for the LEDs.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Wireless Monitoring and Control of LED Fixture

Top level configuration of LightingC component developed in TinyOS 2.x is shown in Fig 2. The interfaces "Provided" and "Used" are shown above and below the LightingM component. Downward and upward arrows depict commands and events. ZED in the fixture is programmed for 20 different duty cycles. By varying the width of the pulses, the current flowing through the LED is varied. EnableM component has been developed to control LED driver operation. The parameters like fixture temperature (TempC component), LED disconnection (LEDDisconC) and driver feedback voltage (AdcC component) are being continuously transmitted to the ZC for every 2 minutes. Heat sink has been designed for an ambient temperature of 30°C. If the measured temperature is above 30°C, then the illumination control algorithm will issue a dimming command through ZC to reduce the current flow in LED which in-turn reduces the heat generation.

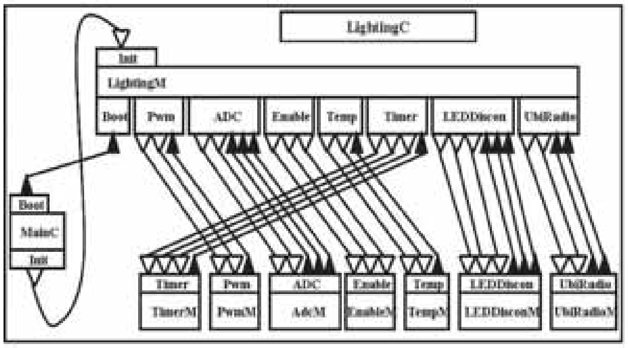


Fig. 2 Component diagram of Ligting control in ZED

The working status of LEDs in the fixture is also monitored. Feedback voltage of the LED driver is measured to keep the operating voltage and current within the rated values.

III. IMPLEMENTATION AND RESULTS

The T8 Fluorescent lamps (T8 FL) in Ubicomp lab (Size: 15x3x2.6 m) of CDAC Chennai have been replaced with 7 LED Fixtures (with ZEDs). Star topology with a ZC and 7 ZEDs has been established. The ZEDs have been programmed in Non-Beacon enabled mode with auto acknowledgement. To compensate for packet losses ZED has been programmed to retransmit for 3 times if it does not receive any acknowledgement from the ZC. To avoid collision, ZED and ZC has been programmed to use CSMA-CA. Our lab was modeled using simulation software tools by setting the geometry and surface properties. Before deployment, the placement of fixtures has been simulated using an optical simulation tool to provide an average luminance of 500 lux (office lighting as per IESNA [3]) in work plane. The deployment and internal view of the LED fixture is shown in Fig.3. A 76 W T8 FL has been compared with 56 W LED fixtures. Actual Light Output (AOL) is the product of Designed Light Output (DLO) and Light Output Ratio (LOR). LOR of T8 FL is obtained from data sheet and for LED fixture it is found using lighting design and simulation software.

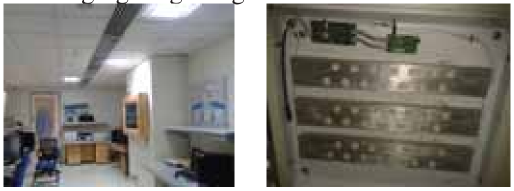


Fig. 3 Deployment of fixtures in ubicom lab at CDAC chennai

Lumen output and wattage comparison of both the lamps in shown in Table II. Three, 24V DC SMPS have been used to drive 7 LED fixtures.

TABLE II LUMEN OUTPUT COMPARISON OF T8 FL AND LED FIXTURE

Lamps	Power / lamp (w)	LOR (%)	DLO (lumens)	ALO (lumens)
T8 FL	76	69	4850	3347
LED	56	80	5334	4267

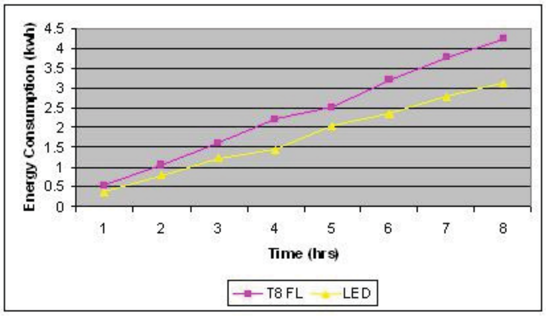


Fig. 4 Time vs. Energy consumed (kwh) by T8 FL and LED fixture

IV. CONCLUSION AND FUTURE WORK

Energy consumed by (7 lamps) T8 FL and LED fixtures has been measured for 8 hrs at maximum illumination (500 lux). Results (Fig 4) show that without including wireless overheads and by replacing T8 FL lamp with LED fixtures we are able to save up to 26.59% of the energy consumed. In future, to save the energy further the illumination system developed by us will be controlled based on user's activity and location.

REFERENCES

[1] Yin Jun, and Wang Wei, "LED lighting control system based on the Zigbee wireless network", ICDMA 2010, vol 1, pp. 892-895

[2] Maoheng Sun, Qian Liu, and Min Jiang, "An implementation of remote lighting control system based on Zigbee technology and Soc solution", ICALIP2008, pp. 629-632

[3] http://www.iesna.org/

[4] http://www.altera.com/literature/an/archives/an185.pdf

[5] http://www.zigbee.org/

[6] Philip Levis and David Gay, "TinyOS Programming". Cambridge University Press, 2009

Effect of Physical and Virtual Carrier Sensing on the AODV Routing Protocol in noisy MANETs

Haitham Y. Adarbah, Scott Linfoot, *Senior Member*, IEEE, Bassel Arafeh, *Member*, IEEE, and Alistair Duffy, *Senior Member*, IEEE

Abstract--**In cellular consumer devices today, one of the limiting factors behind efficiency is that of battery life. The challenge facing cellular consumer device designers is that the discovery phase of the routing process when attempting to establish a mobile ad hoc network tends to put the highest strain on the battery of the device.**

Such routing algorithms tend to be affected by carrier sensing which has lead to increasing packet loss within the network environment. This paper is concerned with studying the performance of Ad hoc On demand Distance Vector (AODV) based on physical and virtual carrier sensing as well as the effect of those carriers on the route discovery.

I. INTRODUCTION

A Mobile Ad hoc NETwork (MANET) is a self-configuring infrastructure-less network of low-power mobile devices connected by wireless links. The advantage to the consumer device is that in such an environment, due to the limited radio range of the wireless link, it may be possible for one node to enlist the aid of other nodes in forwarding data to a destination node not within the radio transmission range of the source. Thus, each node in the network operates not only as a host but also as a router. There are two phases in reactive protocols: route discovery and route maintenance. The earliest mechanism that has been proposed in the literature for route discovery in MANETs is pure flooding. Reactive (dynamic routing) protocols, such as Dynamic Source Routing (DSR) [1], Ad hoc On demand Distance Vector (AODV) [2] are the more widely used routing protocols in MANETs.

Physical Carrier Sense is used when a mobile consumer device seeking to transmit first assesses the channel. If the energy detected on the channel is above a certain threshold (the carrier sense threshold), the channel is deemed busy, and the node must wait. Otherwise, the channel is assumed idle, and the node is free to transmit. A Virtual Carrier Sense uses a special handshake to "reserve" the channel, called the RequestToSend (RTS)/ ClearToSend (CTS) mechanism.

In reality, communication channels in MANETs are unreliable due to channel impairments such as noise, distortion, signal attenuation, atmospheric absorption, free space loss, etc. Rapidly changing topologies caused by node movement and node count variability also contribute. The SINR is considered as a common way to measure the performance of the wireless connection. The definition of the Signal to Interference plus Noise Ratio (SINR) model has been used as described by Adarbah, *et al.*. [3]

II. RELATED WORK

Several studies of routing protocol based on the physical and MAC layers have been done in the literature. Jing Deng, *et al*. [4] and Kim *et al*. [5] argued that the carrier sensing range is a tunable parameter that can significantly affect the MAC performance in multihop ad hoc networks.

Mustapha, *et al.* [6] investigated the impact of sensing range on the throughput of MANET by taking into consideration two essential issues in MAC they are concurrent transmission, which is referred to spatial reuse, and collision in terms of transmission range persistent probability and back-off time.

Vaidya [7] investigated the impact of choosing an optimal carrier sense range by using an analytical model as well as simulation results. Their results reveal that the average of throughput will be affected unless the optimal carrier sense range is determined properly.

III. CARRIER SENSING RANGE

Carrier sensing is an essential mechanism in Carrier sense multiple access with collision avoidance (CSMA/CA) protocols. It consists of physical and virtual carrier sensing, which is called The RTS/CTS mechanism in 802.11.

To summarize RTS/CTS, the sender first sends an RTS message, and the destination replies with a CTS. Then the actual DATA/ACK exchange will be done. Neighboring nodes that receive either the RTS or CTS set their Network Allocation Vector (NAV) so as to reserve the channel for the coming DATA/ACK transmission.

When a node needs to transmit, the node first must sense the channel before transmission. If it senses a busy channel, it needs to abort the transmission to avoid or reduce collision. A busy channel is detected when the sensed power of the signal exceeds a specific threshold referred to as the Carrier Sense Threshold (CST). If the signal power is lower than this threshold, the channel is deemed to be an idle channel [5][9].

The CST value decides the sensing range and has an impact on the collision possibility as well as concurrent transmission in the MANET. When the value of CST is small, that's mean the signal can be sensed in a larger range, and vice versa.

IV. AODV

The Ad hoc On-demand Distance Vector (AODV) routing protocol is used by mobile consumer devices in an ad hoc network. It supports dynamic route conditions; has a minimized memory overhead; requires low processing and low network utilization; and is able to determine unicast routes to destinations within the ad hoc network.

The algorithm's primary objectives are:

1) Run route discovery packet only when it is needed.

2) Differentiate between general topology maintenance and local connectivity management (neighborhood detection).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

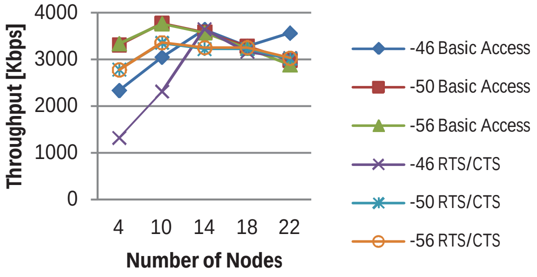


Figure 1: Throughput vs Number of Nodes

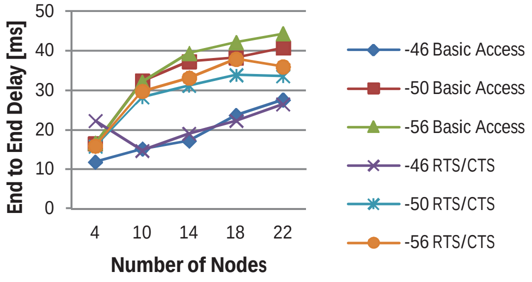


Figure 2: End to End Delay vs Number of Nodes

3) Broadcast information about changes in local connectivity to the neighborhood node devices which will also need the information. Route Discovery is initiated by broadcasting a Route REQuest (RREQ) packet to its neighbors, who then forward the request to their neighbors. This continues until the expiration of this RREQ. Each RREQ packet identifies the source and the destination of the route discovery, and also contains a unique request sequence number or identification number (ID), determined by the source of the request. If the destination node is located and successfully receives the RREQ, the destination node sends back a unicast Route REPly (RREP) packet to the source node back through the route from which it first received that particular RREQ [1].

V. PERFORMANCE EVALUATION

In this section the performance evaluation of AODV is presented in terms of the throughput and the average of the end-to-end delay. Packet Error Rate (PER) is used to determine random packet losses, as well as the successful receptions. PER is defined as the number of incorrectly received data packets divided by the total number of received packets. A packet is declared incorrect if at least one bit is erroneous. To be more realistic, the PER is calculated using the default pre-determined curve (PER vs SINR and packet size) for 802.11g standards, and SINR is calculated using received signal strength, noise plus interference as described by Baldo, *et al.* [8]. In this work, the numbers of nodes considered are 4, 10, 14, 18, and 22.

Simulation is applied for physical and virtual carrier sensing. For both modes of operations, the Carrier Sensing Threshold (CST) considered are -46, -50 and -56dBm. Here are main simulation parameters which have been used: simulation ime 800 sec, transmission power 10dBm, noise power 71dBm, TCP Traffic, 802.11g MAC protocol. The average end-to-end delay includes all possible delays caused by buffering during route discovery latency, queuing at the interface queue, retransmission delays at the MAC, and propagation and transfer times.

Figure 1 shows the Throughput against the number of nodes for different value of CST with ON/OFF for RTS/CTS It should be noted that if the number of nodes is greater than 10, there will be interference between the nodes. As after this number, the throughput has been decreased by the interference. Figure 2 shows the End to End delay against the number of nodes. It can be seen that the average end to end delay, when the RTS/CTS ON is less than when it is OFF, because the average throughput is small. The average end-to-end delay increases with lower carrier Sensing threshold values and increasing number of nodes. In conclusion, route discovery is affected by large carrier sensing and with the density of the MANETs.

VI. CONCLUSION

This research work focuses on the Physical and MAC modules of the MANETs protocol stack. The effect of physical and virtual carrier sensing on the AODV routing protocol, and how those types of carrier sensing are affecting the Packet Error Rate and signal to interference and noise ratio (SINR) are presented. Throughput and end-to-end delay, packet delivery ratio are considered as metrics for measuring a performance of AODV protocol. The route discovery mechanism has been affected by the interference when the number of nodes increases and tuning the carrier sensing range.

REFERENCES

[1] D. B. J. and D. A. Maltz, “Dynamic source routing in ad hoc wireless networks,” *Mobile computing*, pp. 153–181, 1996.

[2] C. E. Perkins, “Ad-hoc on-demand distance vector routing,” in *Proceedings. WMCSA ’99. Second IEEE Workshop on*, 1999, pp. 90–100.

[3] H. Y. Adarbah, S. Linfoot, B. Arafeh, and A. Duffy, “Impact of the Noise Level on the Route Discovery Mechanism in Noisy MANETs,” in *The 1st IEEE Global Conference on Consumer Electronics 2012*, 2012, pp. 710–714.

[4] P. . Jing Deng; Ben Liang; Varshney, “Tuning the Carrier Sensing Range of IEEE 802.11 MAC,” in *Global Telecommunications Conference, GLOBECOM ’04. IEEE*, 2004, pp. 2987–2991.

[5] T.-S. Kim, J. C. Hou, and H. Lim, “Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks,” in *Proceedings of the 12th annual international conference on Mobile computing and networking - MobiCom ’06*, 2006, p. 366.

[6] I. Mustapha, J. . Jiya, and M. Abbagana, “Effect of Carrier Sensing Range on the Throughput of Multi-hop Wireless Ad-Hoc Network,” in *Proceedings of the 1 International Technology, Education and Environment Conference (c) African Society for Scientific Research (ASSR)*, 2011, no. c, pp. 509–518.

[7] N. Vaidya, “On physical carrier sensing in wireless ad hoc networks,” *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, vol. 4, pp. 2525–2535.

[8] N. Baldo, F. Maguolo, and M. Miozzo, “A new approach to simulating PHY, MAC and routing,” in *WNS2 2008 The Second International Workshop on NS-2 (In technical cooperation with ACM)*, 2008.

A Practical MAC Protocol Supporting Discontinuous Channel Bonding

Wenzhu Zhang[1,2], Kyung Sup Kwak[2], *Member, IEEE*, Hongxiang Wang[1], and Jaedoo Huh[3]
[1]National Key Lab. of Integrated Service Networks, Xidian University, Xi'an, China
[2]School of Information and Communication Engineering, Inha University, Incheon, Korea
[3]Electronics and Telecommunications Research Institute, ETRI, Daejeon, Korea

***Abstract*--To enhance the wireless spectrum efficiency, a practical MAC protocol which supports discontinuous channel bonding is proposed. The protocol modifies the traditional frame structure and the access mode of MAC to bear the new idea. It adopts the traditional channel bonding scheme in case of that there are continuous narrow channels; while it adopts discontinuous channel bonding scheme in case of that narrow channels are discontinuous, thus to achieve a parallel transmission. Simulation results indicate that the MAC protocol focused can get obvious advantages on effective bandwidth compared with the continuous channel bonding MAC.**

I. INTRODUCTION

At present, there are two main schemes in using multiple narrow-band channels. One is to adopt parallel transmission, and the other is to bond narrow-band channels into a single wide-band channel, which is named channel bonding technology[1]. The parallel transmission can reduce data losses caused by data collision in some channels; however, it can not satisfy the increasing demand of high rate transmission. On the other hand, the channel bonding technology is a effective method in increasing bandwidth and the transmission rate, which has become a hot research topic[2][3].

In the draft of 802.11ac, researchers have put forward bonding four 20MHz channels to construct an 80MHz channel. To a large extent, the traditional channel bonding technology can indeed increase bandwidth, so to improve the information transfer rate. But it requests that the bonded channels must be rigorous continuous, therefore the defects are obvious. The requirement of channels to be continuous got plenty waste of frequency resource because it is hard to find multiple continuously free channels.

In this paper, we present a practical MAC protocol which supports discontinuous channel bonding by improving the traditional channel bonding scheme.

II. ALGORITHM

The implementation of the protocol presented is based on the new "Control Wrapper frame" proposed in the draft of 802.11ac. For the scheme in this paper, "Control Wrapper frame" is used to control RTS frame and CTS frame, and obtain corresponding "Control Wrapper RTS frame" or "Control Wrapper CTS frame", which is signed as RTS-Packaged and CTS-Packaged respectively. To get the sequence diagram of transmission, we just discuss one possible case of the channel groups. Fig. 1 describes that channel #2 separates four channels into channel group #1 and channel group #2 when the channel #2 is detected as busy. The sequence diagram of this case is depicted in Fig. 2.

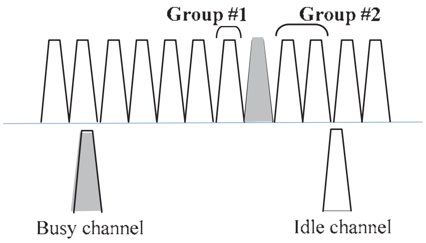


Fig. 1. Grouping of the channels

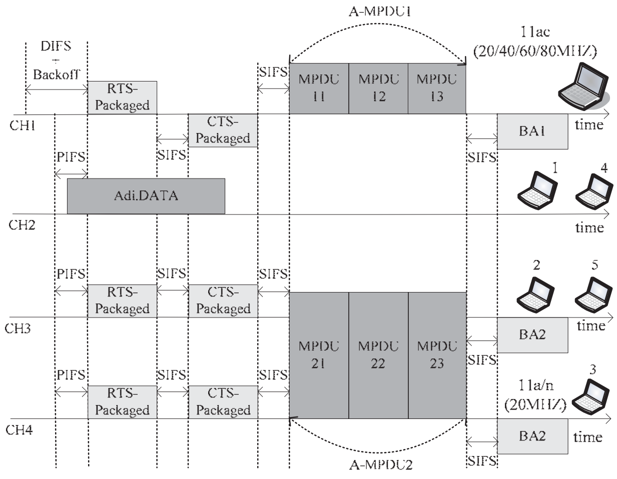


Fig. 2. Timing sequence diagram when channel #2 is busy

III. SIMULATION AND ANALYSIS

We evaluate the performance of the MAC protocol proposed by NS-2 simulation. The evaluating indicator, equivalent bandwidth is adopted to compare the performance of the MAC protocol we proposed with the MAC protocol that doesn't support discontinuous channel bonding. We assume that the wireless scenario contains one 802.11ac BSS (Basic Service Set, BSS) and two 802.11a BSSes; each BSS owns its AP and corresponding workstation. It is also assumed that each channel is exactly the same in frequency bandwidth,

This work was supported in part by the National Natural Science Foundation of China under grant No.61072068, and in part by the National Research Foundation of Korea grant funded by the Korea government (MEST) No.2010-0018116.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

and any different BSSes are likely to share one channel. Table I lists the simulation parameters related.

TABLEL I
SIMULATION PARAMETERS

Parameters	**Value**
Bandwidth of a channel	20MHz
Bit rate in 20MHz channel	130Mbps
Bit rate in 40MHz channel	270Mbps
Bit rate in 80MHz channel	540Mbps
Controlled bit rate	54Mbps
Delay of CCA	4.2 μs
Contention window	CWmin = 7 μs
	CWmax = 63 μs
Max length of A-MPDU	128KBytes
Interframe space	SIFS=16 μs
	PIFS=25 μs
	DIFS=34 μs
Length of RTS-packaged frame	25Bytes
Length of CTS-packaged frame	19Bytes
Length of BA frame	32Bytes

Equivalent bandwidth is the average value of a channel's bandwidth used by a node in the transmission, which can intuitively reflect the throughput performance.

Suppose that the idle probability of each channel is the same, which can be expressed as $P_i \ (i=1,2,3,4)$. Without loss of generality, it can be assumed that the idle probability is equal for each channel, so we get $P_1 = P_2 = P_3 = P_4 = P$. Now we calculate the idle probability of continuous bonding and discontinuous bonding transmission scheme under different bandwidths.

The idle probabilities of continuous bonding transmission scheme are given by the formulas from (1) to (4).

$$P_{80M} = P_1P_2P_3P_4 = (P)^4 \tag{1}$$

$$P_{60M} = P_1P_2P_3(1-P_4) = (P)^3(1-P) \tag{2}$$

$$P_{40M} = P_1P_2(1-P_3) = (P)^2(1-P) \tag{3}$$

$$P_{20M} = P_1(1-P_2) = P(1-P) \tag{4}$$

The idle probabilities of discontinuous bonding transmission scheme are given by the formulas from (5) to (8).

$$P_{80M} = P_1P_2P_3P_4 = (P)^4 \tag{5}$$

$$\begin{aligned} P_{60M} &= P_1P_2P_3(1-P_4) + P_1P_2P_4(1-P_3) \\ &+ P_1P_3P_4(1-P_2) = 3P^3(1-P) \end{aligned} \tag{6}$$

$$\begin{aligned} P_{40M} &= P_1P_2(1-P_3)(1-P_4) + P_1P_3(1-P_2)(1-P_4) \\ &+ P_1P_4(1-P_2)(1-P_3) = 3P^2(1-P)^2 \end{aligned} \tag{7}$$

$$P_{20M} = P_1(1-P_2)(1-P_3)(1-P_4) = P(1-P)^3 \tag{8}$$

The formula for equivalent bandwidth is as follow:

$$\begin{aligned} B_{eq} &= \sum_B P_B B_B = P_{20M}\cdot B_{20M} + P_{40M}\cdot B_{40M} \\ &+ P_{60M}\cdot B_{60M} + P_{80M}\cdot B_{80M} \end{aligned} \tag{9}$$

From formula (1) to (9), we can get the equivalent bandwidth for continuous channel bonding and discontinuous channel bonding scheme, which can be denoted as B_{eq0} and B_{eq1} respectively:

$$\begin{aligned} B_{eq0} &= 20\cdot\left[P(1-P)\right] + 40\cdot\left[(P)^2(1-P)\right] \\ &+ 60\cdot\left[(P)^3(1-P)\right] + 80\cdot\left[(P)^4\right] \end{aligned} \tag{10}$$

$$\begin{aligned} B_{eq1} &= 20\cdot\left[P(1-P)^3\right] + 40\cdot\left[3(P)^2(1-P)\right] \\ &+ 60\cdot\left[3(P)^3(1-P)\right] + 80\cdot\left[(P)^4\right] \end{aligned} \tag{11}$$

By using formula (10), formula (11) and the parameters in Table I, we compare the equivalent bandwidth for continuous channel bonding and discontinuous channel bonding transmission scheme under different channel's idle probability, as can be seen in Fig. 3.

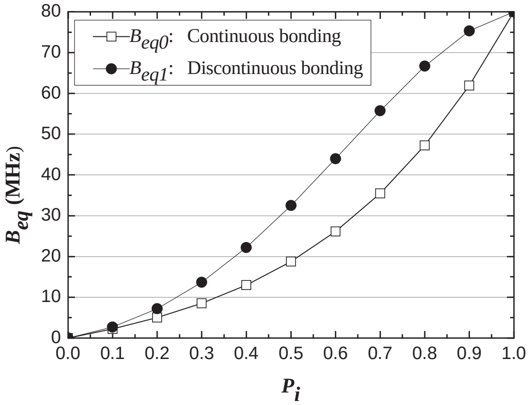


Fig. 3. Equivalent bandwidth vs. Channels' idle probability

Fig. 3 shows that the scheme of discontinuous bonding can increase the equivalent bandwidth significantly.

IV. CONCLUSION

A practical MAC protocol is proposed which supports discontinuous channel bonding. By sensing the channels' states, it divides the channels into groups. After that, it takes advantage of the improved Control Wrapped frame to carry the channel grouping information, and allocates the aggregated frames on different channel's groups and parallels the transmission. Simulation results show that the MAC protocol presented can improve the equivalent bandwidth.

REFERENCE

[1] L. Xu, K. Yamamoto, and S. Yoshida, "Performance Comparison Between Channel-Bonding and Multi-Channel CSMA," *IEEE Wireless Communications and Networking Conference 2007*, Hong Kong, pp.406-410, Mar. 2007.

[2] M. Park, "IEEE 802.11ac: Dynamic Bandwidth Channel Access," *2011 IEEE International Conference on Communications*, pp.1-5, June 2011.

[3] E. Perahia and R. Stacey, *Next generation wireless LANs: throughput, robustness & reliability in 802.11n*, Cambridge Press, Sept. 2008.

Time-based Interest Protocol for Real-Time Content Streaming in Content-Centric Networking (CCN)

Joonghong Park, Jaehoon Kim, Myeong-wuk Jang and Byoung-Joon (BJ) Lee
SAIT, Samsung Electronics, Korea

***Abstract*— Internet increasingly suffers congestion on the server side, particularly due to ever-increasing demand for high-quality audio/video streaming. CCN is considered as an important networking paradigm which can efficiently address such traffic explosion issue. Although early implementation of CCN protocol is available through CCNx open source project, current version lacks efficient support for real-time streaming applications. In this paper, we propose an enhanced mechanism for the support of real-time content streaming service in CCN, and present experimental results demonstrating its effectiveness.**

I. INTRODUCTION

The Internet was originally designed to connect end-host devices, but nowadays it is mainly utilized for mass distribution of high-quality streaming of video contents. Because of this significant change of usage pattern of the Internet, the possibility of congestion on the side of content servers has continuously been increasing. Content-Centric Networking (CCN) has received a lot of attention in recent years, and is being considered as one of the important network paradigms to address such fundamental problem due to explosive growth of data traffic [1,2]. In CCN, a user requests content by its name without having to know its location. Since the content being delivered can explicitly be stored on caches of intermediate CCN routers on its route with its name, other requests for identical contents sent from other users can be responded from cached contents at intermediate CCN routers. Thus, along with the request aggregation feature, CCN can maximize the effect of multicasting without additional multicast membership management [1,2].

In CCN, the granularity of content delivery is not a file but a segment. Therefore, content segments can be delivered from multiple nodes and/or through different routes. It also enables the transfer of each segment to easily adapt to timely changes of congested networks. Therefore, a content request message, called Interest, is sent to get a corresponding segment. It means that many Interests should be sent to access the entire content, and the number of Interest sent is equal to or bigger than the total number of segments of a content. In the basic CCN protocol, an Interest to retrieve a segment is sent after receiving its previous segment. This approach introduces inefficiency in data throughput, because a requester should wait until the previous segment is delivered. Pipelining can enhance the performance [3], where multiple Interests are transmitted within the pipeline window before receiving the previous segments. However, this mechanism still requires sending an Interest for each segment, causing processing overheads of sending and forwarding multiple Interests for retrieving contents.

To reduce potentially large number of Interests generated for high-quality streaming video content, this paper proposes a time-based Interest protocol; an Interest is sent with time duration of the intended request, and corresponding content segments are transferred during the specified time duration. This paper describes the proposed mechanism with experimental performance results.

II. TIME-BASED INTEREST PROTOCOL

2.1 Interest and content segment exchange flow

The time-based Interest protocol sends an Interest with information about time duration, and this time duration can be specified based on the performance requirements of the intended service, type of devices supported, and/or the condition of network environment in use, etc. During this specified time duration, all content segments generated or transferred from other nodes can be delivered to the requester.

It should be noted that, however, the content name carried in the proposed time-based Interest can only specify the name prefix part without the segment number, while the corresponding content segments carry the full content name including the segment numbers. Therefore, the content name in a time-based Interest should be checked by partial matching instead of exact matching. In order to distinguish time-based Interest from other basic Interests, a specific name component, %TIMEBASED%, is used. The last segment of the flow is also marked with a special component, END_OF_CONTENT, instead of segment number.

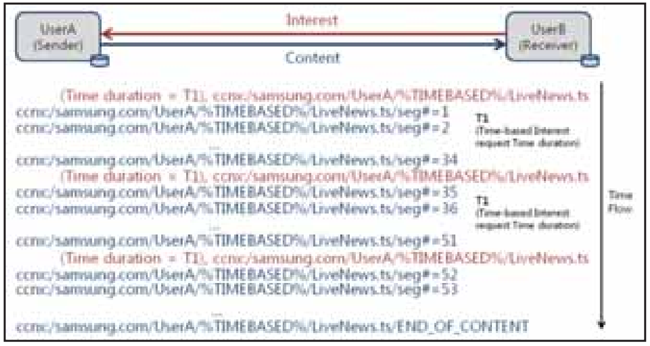

UserA (Sender)
Interest
Content
UserB (Receiver)
(Time duration = T1), ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/seg#=1
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/seg#=2
T1 (Time-based Interest request Time duration)
...
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/seg#=34
(Time duration = T1), ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/seg#=35
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/seg#=36
T1 (Time-based Interest request Time duration)
...
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/seg#=51
(Time duration = T1), ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/seg#=52
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/seg#=53
...
ccnx:/samsung.com/UserA/%TIMEBASED%/LiveNews.ts/END_OF_CONTENT
Time Flow

Fig. 1 Time-based Interest protocol Interest/Content segment exchange flow

Fig. 1 depicts the format of content names, and the Interest/Content segment exchange flow of the proposed time-based Interest protocol. When the time-based Interest carries the content name ccnx:/Samsung.com/UserA/%TIMEBASED%/LiveNews.ts, with time duration of T1, multiple content segments which can be sent during T1 are delivered to the content requestor. For the definition and structure of hierarchical CCN naming convention, refer to [1] and [3]

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

2.2 *Real-time cache management*

The most important requirement of real-time content streaming is to guarantee end-to-end delay constraints. In CCN, intermediate network nodes cache content segments, in order to maximize the multicast effect and available network capacity. In both unidirectional and interactive/bidirectional real-time streaming case, however, it is necessary to maintain only the minimum number of recent content segment, so that the possibility of excessive end-to-end latency and also the out-of-order segments can be avoided.

Fig. 2 depicts a cache management scenario where the proposed scheme utilizes separate cache space to support real-time streaming application, e.g., *real-time content1*. Note that, in the time-based Interest protocol, each node only keeps one most recent segment copy received from the upstream node. In the case of non-real-time content delivery, however, multiple segment copies may need to be cached for high data throughput performance.

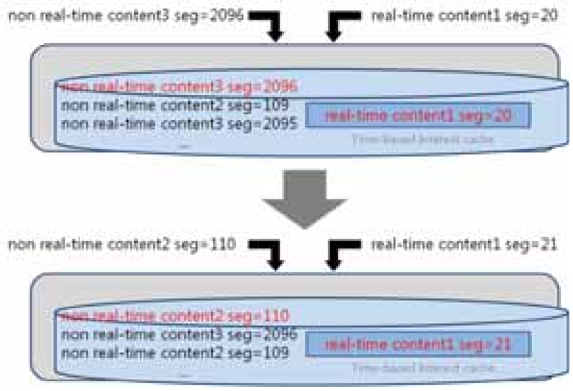

non real-time content3 seg=2096
real-time content1 seg=20
non real-time content3 seg=2096
non real-time content2 seg=109
non real-time content3 seg=2095
real-time content1 seg=20
Time-based Interest cache
non real-time content2 seg=110
real-time content1 seg=21
non real-time content2 seg=110
non real-time content3 seg=2096
non real-time content2 seg=109
real-time content1 seg=21
Time-based Interest cache

Fig. 2 Cache management scenario with separate minimum cache allocation for time-based Interest protocol supporting real-time streaming service

III. Performance Evaluation - Experimental Results

Our experiments are conducted with prototype implementation based on an enhanced version of CCNx open source code[3]. The testbed consists of a set of Android phones (Galaxy S2 with Android 2.3) with 1.2GHz dual-core processor, and connected through IEEE 802.11n Wi-Fi. We developed a real-time video streaming application running on a software codec. One phone captures and distributes the captured video to be re-played on the other 5 phones.

To compare the performance of proposed time-based Interest protocol with that of the simple pipelining mechanism, two factors are considered: (1) quality of service measured by end-to-end delay jitter, and (2) the average data rate measured by receiving phones. Both performance parameters are measured at 5 encoding rates: 300k, 500k, 700k, 1000k, and 1200kb. Each test is run for 30 seconds, and four pipelining sizes are considered, i.e., Pipe size= 1, 5, 10, and 20).

3.1 *Delay Jitter Performance*

Delay jitter is an important performance parameter, since the quality of service perceived by end users is greatly affected by play buffer under/over-run caused by large delay jitter. Fig. 3 shows the average result of delay jitter measurement on each receiver side.

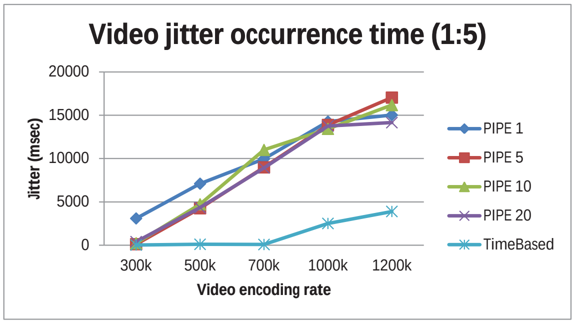


Fig. 3 Performance comparison: Delay Jitter

As the video encoding rate increases, delay jitter of the pipelining mechanism also significantly increases, while that of the proposed time-based Interest protocol marginally increases.

3.2 *Received Data Rate Performance*

The data rate measure at the receiving end is another important performance parameter of the network. Fig. 4 demonstrates the performance benefit of the proposed scheme as compared to that of the simple pipelining mechanism. As the encoding bit rate increases, the performance gap between the pipelining mechanism and time-based Interest protocol also becomes larger.

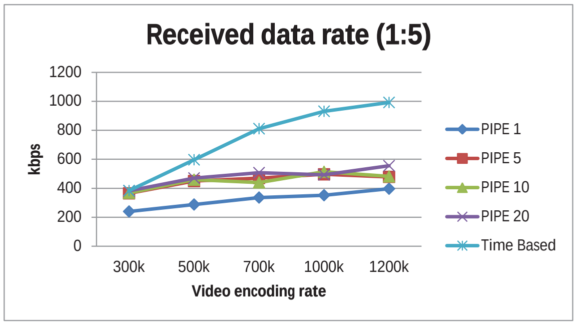


Fig. 4 Performance comparison: Received data rate

IV. Conclusion

The proposed time-based Interest protocol reduces the number of Interests and also the processing overhead of sending and forwarding of Interest packets, while allowing all the corresponding content segments in the specified time duration of an Interest to be delivered. Our experimental results demonstrated that the proposed scheme provides better quality of service in terms of end-to-end delay jitter, and higher data throughput than the simple pipelining mechanism. Future research plan also includes the consideration of multi-hop network environment where the time duration for the time-based Interest may be adjusted hop-by-hop in order to accommodate the various network conditions.

Reference

[1] Van Jacobson, Diana. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, Networking Named Content, In Proceedings of CoNEXT’09, Rome, Italy, Dec. 2009.

[2] Lixia Zhang, et al, “Named Data Networking (NDN) Project,” PARC Technical Report NDN-0001, October 2010.

[3] Project CCNx. http://www.ccnx.org, Sep. 2009.

Hierarchical Packet Scheduler for Supporting Fairness and QoS Over DVB-S2 ACM Systems

ManKyu Park, DongBae Kang, MinSu Shin, DeockGil Oh
Satellite Broadcasting & Telecommunications Convergence Research Team, ETRI, Korea

Abstract—**In this paper, we propose a hierarchical scheduler offering fairness and QoS for DVB-S2 ACM systems while minimizing the decrease of throughput. As a result of the simulation, we show that the proposed scheduler supports bandwidth fairness to the individual RCST and can provide some level of QoS differentiation for user traffics.**

Keywords-Satellite networks, ACM, fairness, scheduler, QoS

I. INTRODUCTION

Although satellite communication system with ACM technology has higher transmission efficiency, it does not offer fairness to each RCST on the ground. Because the data is transmitted with high MODCOD in regions of clear sky while it is transmitted with low MODCOD in regions with rain events. In fact, this situation is very irrational to RCSTs which have the same SLA contract. To overcome those problems, we propose a hierarchical scheduler offering fairness and QoS to RCSTs and minimizing the decrease of throughput. From the performance evaluation, we show that the proposed scheduler supports bandwidth fairness to the individual RCST and can provide some level of QoS differentiation between Expedited Forward (EF) , Assured Forward (AF) and Best Effort (BE) class traffics.

The paper is organized as follows. In section II we show the architecture of proposed hierarchical scheduler and describe algorithm conducted in each step. In section III we indicate the result of simulation and main conclusions and further research topics are outlined in the last section IV.

II. PROPOSED HIERACHICAL PACKET SCHEDULER

In this paper, we present the architecture and scheduling algorithm of a proposed hierarchical scheduler for supporting fairness and QoS. In Figure 1, the packets are classified as being of either the EF class, the AF class or the BE class of QoS in the packet classifier. Classified packets are processed by the first step scheduler and are transferred to MODCOD queues. To support fairness to the earth stations, we process the packets by providing low MODCOD queues with many different processing times[1][2][3].

A. *Scheduling Algorithm of the First Step*

We apply the priority scheduling algorithm in order to deal with packets which have other QoS levels, as seen in Figure 1. Packets from the high priority queue are always processed before packets from the medium priority queue are processed. Likewise, packets from the medium priority queue are always processed before packets from the low priority queue are processed.

B. *Scheduling Algorithm of the Second Step*

The packets processed by the first step scheduler are classified by each MODCOD. Each MODCOD queue is processed by a Weighted Round Robin Scheduling Algorithm in order to provide RCSTs with fairness. The weight value is defined as

$$W_i = \frac{C_{\max} \times timeslot}{C_i} \tag{1}$$

where C_{max} = *max* $\{C_i \mid i=0, \ldots, n\}$, *timeslot = packet size /* C_{max} and C_i is each MODCOD level. RCSTs in regions with bad weather conditions are allocated more timeslots than RCSTs in regions with clear sky conditions.

Due to bad weather conditions RCSTs in regions with bad channel conditions transmit packets with the low transmission rate. So they need more timeslots than RCSTs located in a region with clear sky conditions. Although they have a different transmission rates in a separate location, this algorithm can offer RCSTs having fairness with respect to different weather conditions.

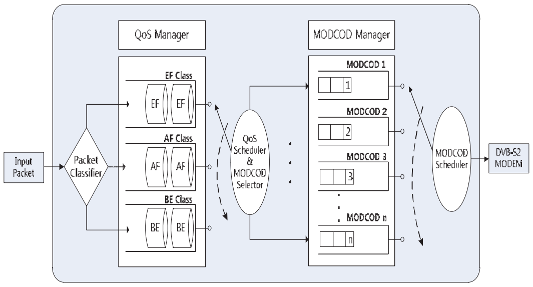


Figure 1. Architecture of the Hierarchical Scheduler

By applying this algorithm, we offer the same throughput to all RCSTs. However the overall throughput decreases because users belonging to a region of clear sky conditions will be affected by the bad weather conditions. In this paper, we apply the priority scheduling algorithm in the first step scheduler in order to minimize the loss of EF, AF class data. So most of loss data is filled with the packets of BE class.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. Performance Evaluation

we evaluate performance for our proposed scheme through computer simulations using sim++ [4], which is one of the queuing simulators. Simulation configurations are shown in Table 1, and the queuing network topology used for the simulation is shown in Figure 2.

TABLE I. Simulation configuration

Bandwidth	120 Mbps ~ 300 Mbps
Packet size	1500 bytes
Traffic ratio	1:1:2 (EF:AF:BE)
MODCODs	MODCOD 8 (QPSK, 4/5) MODCOD 13 (8PSK, 2/3) MODCOD 19 (16APSK, 3/4) MODCOD 25 (32APSK, 4/5)

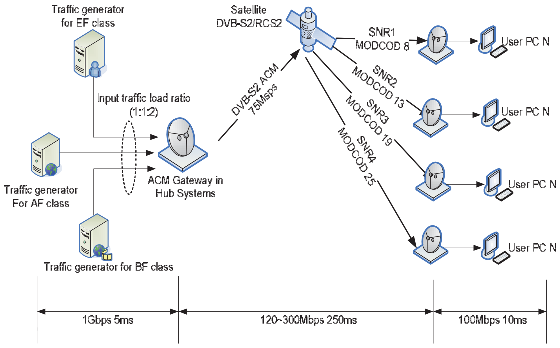
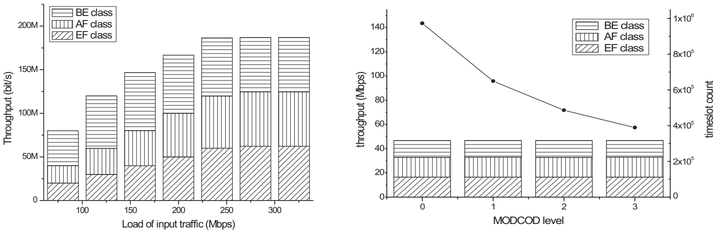


Figure 2. Simulation Network Topology

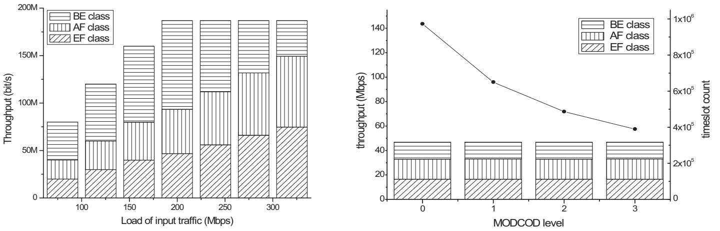
A. Packet Throughput and Fairness

a) Round-robin scheduling Algorithm: In the round-robin scheduling algorithm, the total thoughput is equal regardless of the priority of the QoS after the throughput reaches the maximum values. Because the weight value in each MODCOD is perfectly equal, the RR algorithm does not offer a guarantee of QoS in Figure 4(a). In addition, although the packets are transmitted to the different regions having different MODCODs which have different transmission rates, we should assure fairness in Figure 4 (b). As a result of the simulation in 4(b), more timeslots are allocated to the regions with bad conditions having a low MODCOD than to regions with clear skies having a high MODCOD. Therefore the amount of transmitted packets to regions having defferent weather conditions are completely equal regardless of the type of packets.

b) Priority Scheduling Algorithm: Figure 5(a) shows the throughput of the priority scheduling algorithm. The BE class traffic is gradually decreased when the traffic is increased in the networks. We make up a loss of throughput with BE class which has occurred by supporting QoS, because high priority packets are always transmitted earlier than low priority packets. As a result of the simulation in 5(a), more packets of the EF and AF classes are transmitted than packets of the BE class.Therefore most of loss data is filled with the packets of BE class. To supporting fairness for each RCST, more timeslots are allocated in the region of bad conditions having a low MODCOD than in the region of clear skies having a high MODCOD in Figure 5(b). Therefore, we offer the fairness of throughput to each RCST considering different weather conditions. However, the amounts of transmitted packets are different depending on the weather conditions.



(a)Throughput per load of input traffic (b)Throughput of each MODCOD level
(0: MODCOD8, 1: MODCOD 13, 2: MODCOD 19, 3: MODCOD 25)
Figure 4. Total throughput of Round-robin Scheduling Algorithm



(a)Throughput per load of input traffic (b)Throughput of each MODCOD level
(0: MODCOD8, 1: MODCOD 13, 2: MODCOD 19, 3: MODCOD 25)
Figure 5. Total throughput of Priority Scheduling Algorithm

IV. Conclusion

In this paper, we have proposed a hierarchical packet scheduler for supporting fairness and QoS in DVB-S2 ACM systems. We applied it to a weighted round-robin scheduling algorithm in order to offer earth station fairness. The weight value is calculated by the transmission rate of each MODCOD queue. We show that the priority scheduling algorithm offers the earth stations fairness and QoS, while the Round-robin scheduling algorithm does not provide QoS to the earth stations. In addition, the priority scheduling algorithm in the first step scheduler makes up the throughput which has been decreased due to the support of fairness.

References

[1] F. Vieira, M. A. Vazques Castro and G. Seco Granados, “A tunable-fairness cross-layer scheduler for DVB-S2,” International Journal of Sat. Com. and Networking, vol. 24, pp. 61-69, 2006, pp. 61-69.

[2] E. Rendon-Morales, J. Mata-Diaz, J. Alins, J. L. Munoz, and O. Esparza “Adaptive Packet Scheduling for the Support of QoS over DVB-S2 Satellite Systems,” 9th International Conference on Wired/Wireless Internet Communications, vol. 6649, 2011, pp.15-26.

[3] U. Park, H. W. Kim, D. S. Oh, and B. J. Ku, “A Dynamic Bandwidth Allocation Scheme for Multi-spot-beam Satellite System, ETRI Journal, Vol. 34, No. 4, Aug. 2012, pp. 613-616.

[4] sim++ Version 1.0, www.cis.ufl.edu/~fishwick/simpack/simpp.ps

Improvement of Connectivity Between Infrastructure and Mobile Terminals for Infotainment Services

Eun-Jeong Jang, Rinara Woo and Dong Seog Han, *Senior Member, IEEE*

***Abstract*—An algorithm for improving connectivity between an infrastructure and mobile terminals is proposed based on IEEE 802.11p/WAVE for infotainment services. We improve the connectivity with the cooperation of neighbor terminals.**

I. INTRODUCTION

The U.S. and other countries are going to commercialize technologies for the infotainment services to mobile terminals in fast-moving environments [1]. Recently, the wireless communication technologies, networking architectures and protocols have been specified in the IEEE 802.11p and IEEE 1609 protocol set to provide infotainment services [2], [3].

The IEEE 1609.4 standard has been proposed to support multi-channel operations in IEEE 802.11p/WAVE (wireless access in vehicular environments) [2], [3]. There are two types of channels. One of them is the common control channel (CCH) for conveying safety and control messages on a fixed frequency band. The other is the service channels (SCHs) for data exchange to provide infotainment services. An access point (AP) as an infrastructure within a basic service set (BSS) provides infotainment services to mobile terminals. The CCH and SCH are alternately used with a time division multiplexing scheme. The length of an interval for CCH or SCH is 50 ms [3].

In a CCH interval, networks are used to convey status messages periodically from APs and mobile terminals [3]. There are two types of status messages, beacons and WAVE service advertisement (WSA) messages. Beacons are one-hop broadcasted status messages between mobile terminals. Beacons contain status information about the mobile terminal's position, speed, etc. These messages are transmitted according to the enhanced distributed channel access (EDCA) protocol [2].

The WSA is a broadcasted message during the CCH interval by an AP to mobile terminals within a BSS [3]. The main purpose of WSA is an advertisement of AP's own presence. The WSA specifies the initialization information such as the SCH frequency used for the BSS set-up, the offered services, and other connection parameters. Those are needed to access an AP during the SCH interval [4], [5]. When the channel is switched from the CCH interval to the SCH interval, mobile terminals have to tune to the SCH specified in the WSA. Then mobile terminals start to exchange data with the AP. Therefore, connectivity between mobile terminals and the AP during the SCH interval is strongly related to the successful reception of WSAs from the AP during the former CCH interval [4]. Mobile terminals which did not receive WSAs during the CCH interval can't exchange data with the AP during the next SCH interval. To solve this, Claudia Campolo and Antonella Molinaro proposed an algorithm called WSAp [4]. The algorithm is that mobile terminals which have received WSAs transmit piggybacked beacons with the received WSA according to a given probability for the piggybacking. WSAp with a high probability increases the number of mobile terminals received WSAs. It, however, produces excessive overheads on the channel.

In this paper, we propose an algorithm that increases connectivity between mobile terminals and APs during SCH intervals by extending WSA frame with a flag bit marking reception of WSAs and piggybacking a WSA according to the flag bit.

II. PROPOSED ALGORITHM

In the proposed algorithm, the beacon frame from the legacy is extended with a flag bit. Fig. 1(a) shows the beacon frame in IEEE802.11p/WAVE. The flag bit represents success or failure of the reception of WSAs. The flag bit is initialized to '0' at the starting point of the CCH interval. When a mobile terminal receives a first WSA from an AP during this CCH interval, the flag bit in the beacon frame is set to '1', otherwise '0'.

There is a mobile terminal **A** that couldn't receive any WSAs. Therefore, **A** broadcasts beacons with the flag bit '0'. Neighbor mobile terminals, which received beacons from **A** and WSAs and are within communication range, realize that **A** couldn't receive any WSAs. One of the neighbor mobile terminals broadcasts a beacon piggybacked with a received WSA according to given a probability p. As these procedures, **A** can receive a WSA, tune to the SCH and exchange data with the AP during the next SCH interval.

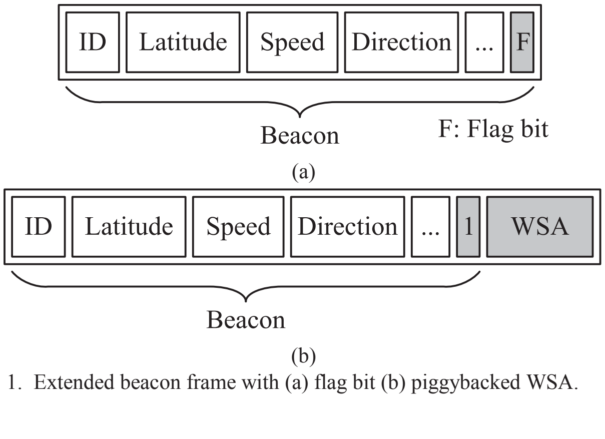


Fig. 1. Extended beacon frame with (a) flag bit (b) piggybacked WSA.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Additionally, the other neighbors that couldn't receive any WSAs also receive a WSA by receiving the beacon (piggybacked with a WSA). Every neighbor of **A** is no longer piggybacks the received WSA in its beacon. It is for avoiding overheads caused by unnecessarily piggybacking.

III. Computer Simulations

The physical and MAC layer parameters are set with IEEE 802.11p/WAVE standards [2]. The parameters are presented in Table. 1 [2].

TABLE I
Simulation Parameters

Parameter	Value
Frequency	5.9 GHz
Channel bandwidth	10 MHz
Transmission power	7 dBm
Carrier sense threshold	-95dBm
Transmission rate	3 Mbps
SINR threshold	10 dB
Slot time	13 µs
SIFS time	32 µs
Head length	40 µs

The number of terminals under the AP coverage is from 1 to 100. Terminals are randomly located. The beacon packet size is set to 100 bytes, the WSA packet size is same with the beacon packet size.

Fig. 2 shows the packet delivery ratio of WSA according to the number of mobile terminals. In Fig. 2, the packet delivery ratio of WSA of the proposed algorithm is higher than legacy IEEE 802.11p. That is, the proposed algorithm increases the connectivity between an AP and mobile terminals. The results of simulation show that a higher probability for the piggybacking used in the proposed algorithm bring the better performance than that with a lower probability.

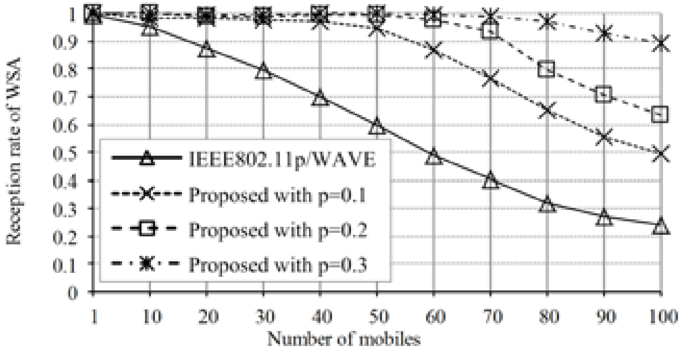


Fig. 2. Reception rate of WSA messages at each terminal.

IV. Conclusions

In this paper, we proposed an algorithm that improves connectivity between an infrastructure and mobile terminals to provide the infotainment services in fast-moving environments. The beacon frame is extended with a flag bit which represents success or failure of reception WSAs. Mobile terminals forward a WSA to other mobile terminals by broadcasting a beacon piggybacked with a WSA, according to a flag bit on a received beacon and a given probability of piggybacking. Through the simulation, the packet delivery ratio of WSA of the proposed algorithm is higher than that of the legacy IEEE 802.11p.

Acknowledgment

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the CITRC(Convergence Information Technology Research Center) support program (NIPA-2012-H0401-12-1006) supervised by the NIPA(National IT Industry Promotion Agency).

Reference

[1] K. Dar et al., "Wireless Communication Technologies for ITS Applications," *IEEE Commun. Mag*., vol. 48, no. 5, May 2010, pp. 156–162.

[2] *IEEE Standard for Information technology-- Local and metropolitan area networks-- Specific requirements-- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments*, IEEE std. 802.11p, July. 2010.

[3] *IEEE Standard for Wireless Access in Vehicular Environments (WAVE)—Multi-Channel Operation*, IEEE Std. 1609.4, Sep. 2010.

[4] C. Campolo, H.A. Cozzetti, A. Molinaro and R. Scopigno, "*Augmenting Vehicle-to-Roadside connectivity in multi-channel vehicular Ad Hoc Networks*," *J. Network and Comput. Applicat.*, Elsevier, 2012.

[5] C. Campolo and A. Molinaro, "*On Vehicle-to-Roadside Communications in 802.11p/WAVE VANETs*," in Wireless Communications and Networking Conference, Cancun, Mexico, 2011, pp.1010-1015.

An Audio-Haptic Feedbacks for enhancing User Experience in Mobile Devices

Jeong-Mook Lim[†], Jong-Uk Lee[†], Ki-Uk Kyung[†] and Jae-Cheol Ryou[‡]
[†]Electronics and Telecommunications Research Institute, Daejeon, Korea
[‡]Chungnam National University, Daejeon, Korea

Abstract*-- We introduce a haptic library that creates tactile feedback by analyzing audio data. Because the proposed haptic library uses audio signal of application to make tactile feedback, it doesn't need to modify application. Also, user can select a particular audio frequency band from multiple audio sources with haptic profiles; then only selected audio can be converted to the tactile feedback. We designed 4 haptic profiles for specific tactile effects. Finally, application examples of applying the haptic library were introduced.

I. INTRODUCTION

Recently, most of mobile devices provide tactile feedback. Tactile feedbacks were mainly used for alerting to users as regards incoming calls or messages in the early days. Nowadays tactile feedbacks are used for confirming user selection in GUI environment. They are also used to represent the surface texture of graphical objects, to improve user experience. In addition, applications such as video games make use of tactile feedbacks to concentrate on the games. Game applications provide more realistic feedbacks combining tactile feedbacks and audio effects so that users can be immersed in the game. Tactile feedbacks also can play an important role in a music player. Music is an art representing human feeling with sound and silence. Its common elements are melody, rhythm, dynamics, and timbre. Users can enjoy more dynamically representing these elements with tactile feedback.

There are several studies to provide a richer user experience using tactile feedback and audio effect. Menelas[3] showed that the combination of audio and haptic cues is very helpful to reach the target in a virtual environment. This indicates that the combination feedback could reduce cognitive load on users. Breamish[4] developed "D'Groove", an intelligent Disc Jockey (DJ) system that uses a haptic turntable for controlling the playback of digital audio. D'Groove provides tactile feedback for the tempo of a song so that DJ's beat-matching skill could be improved. Baillie[5] introduced a mobile music player combined with tactile feedback, and he showed that the user experience could be enhanced when the user hear the music but also feel it. Ichiyanagi[6] proposed a model that links emotional states evoked by music content to specific haptic stimuli. He stated this model would be helpful when visually impaired person select a musical piece in collections.

Unlike the earlier studies, we focused on a manner about converting audio to haptic effect in this paper. In general, sound effect in game or general music is consists of various sounds. For example, various sound effects are provided in racing car games, including sound of engine boosting, sound of the wheels depending on road condition, horn sound, or background music etc. However, to convert all the sound effects to the haptic effect, the user could be interfered to concentrate on the game. It is more effective to enjoy the game that the only selected sound effect would be converted to haptic effect.

The purpose of this paper is to introduce the haptic library that converts the selected audio to the haptic effect, and to present its utilization.

II. HAPTIC LIBRARY

We developed the haptic library that converts an audio source to tactile output. The haptic library was designed running on Android (v.2.3). The conversion process is described in Fig. 1.

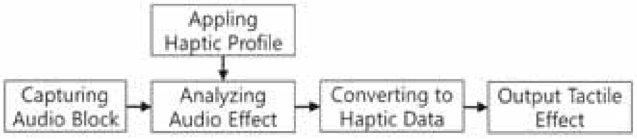


Figure 1. Audio-Haptic Conversion Process

To analyze the audio signal, the library captures the signal as a block. The block capturing rate is adjustable between 1~100 blocks/s. As the block capturing rate decreases, the time interval between the blocks increases and then causes a time delay. As the block capturing rate increases, there will be less of a time delay, but the conversion requires a high computation load in a smartphone. The allowable delay may depend on the application and the system performance. Therefore, the number of capturing blocks in 1 second was designed to be adjustable in the haptic library. In other words, for music with fast rhythm, we set the number of capturing block as 50. Therefore, time interval between capturing blocks should be 20msec. The user can hardly notice the delay in 20msec interval.

Table 1. Haptic Profiles (Frequency/Intensity Threshold)

Haptic Effect/ Frequency(KHz)	***0.04~ 0.08***	***0.2~ 0.3***	***0.64~ 0.86***	***1.7~ 1.9***	***2.5~ 2.8***	***3.8~ 4.3***
Engine Boosting Sound		0.1	0.02	0.02		
Background Music	2		0.02			0.02
Vocals					0.02	0.02
Drum Beats	2	2				

We designed the haptic library that the user can choose the frequency range of audio converted to tactile feedback. To easily select a particular frequency band, the haptic library provides four kinds of tactile profile. It contains frequencies and threshold intensity of the each frequency. Therefore, the selected tactile profile will be applied the FFT(Fast Fourier Transform) filtering step. Each capturing audio block is analyzed by the FFT filter. We evenly divided the audio frequency into 512 sub ranges. We investigated the intensity of each frequency range that was defined in the tactile profile. The haptic library will convert the frequency band that has more intensity than the threshold to the tactile feedback.

In addition the delay time, another property determining the tactile quality is the intensity of the vibration stimuli. Fig. 2 (R) shows the frequency response. We measured the acceleration range according to the input voltage and the frequencies. To ensure the maximum intensity range (Figure 2, 0 ~ 3.2G), the vibration frequency was fixed at 100Hz. The intensity of the vibration was adjusted by the intensity of the audio frequency band.

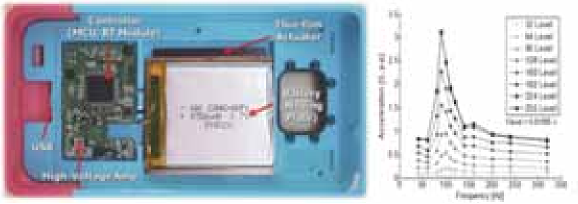


Figure 2. SHIFT(L) and Frequency response(R)

III. Haptic Interface

We developed a haptic interface to test our haptic library. The haptic interface was designed as a bumper case for a smartphone[1]. The electro-active polymer(EAP) actuator was used in the bumper case. The EAPs are polymers that exhibit a change in size or shape when stimulated by an electric field[2]. We attached a mass on the EAP, then the mass vibrated by the contraction and extension of the EAP. The tactile feedback was created by the motion of the mass. The range of input voltage is 0 to 3.021V. Fig 2 (L) shows inside structure of the bumper case.

Figure 3. Applications (L) Music Player, (R) Racing Game

IV. Application

We tested the haptic library with a music player and a game on a smartphone.

Interactive Game: Conversion of sound effects to haptic effects can be efficiently applied to a game. We applied the haptic library to a racing game that contains a variety of sound effects, such as background music, engine boosting effect, road friction noise and others. There are two types of profiles for a racing game. One is for the engine boosting sound, and the other is for the background music. A player can select a profile to make tactile feedback and a haptic feedback corresponding to each sound effect gives the player a more exciting feeling. Even though the game was not designed with haptic feedback, if the game contains sound effects, the player can feel the corresponding tactile feedback without any modification.

Music player: The proposed haptic library can be applied to a music player. The haptic library provides two haptic profiles for the music player. With these profiles, the proposed haptic library distinguishes drum beats and vocals sound in same music piece, and it converts the selected sound to the tactile feedback. With this feature, the user is able to enjoy tactile feedback that represents rhythm, tempo and melody line. Figure. 3(L) shows a user who listen music with the proposed haptic library.

V. Conclusion

In this paper, we introduced a haptic library that creates tactile feedback using audio signal. The user can determine the audio frequency and intensity threshold for converting to the tactile feedback, so the only selected audio would be converted to the corresponding tactile feedback. We designed 4 haptic profiles for game and music player. The audio frequency and the intensity threshold are defined, in each profile.

There are 2 key properties that determining the quality of the tactile feedback. One is time delay and the other is the intensity of the tactile stimuli. The converting process is designed to be handled in a very short time so that the user could feel hardly the delay. The converting process works every 20msec in the proposed haptic library. In addition we can adjust the intensity of tactile stimuli, the audio volume changes can be presented realistically with the haptic library.

For the future research, we will add new haptic profiles reflecting the sound characteristic of more instruments, like a piano, a guitar, etc. Also we will design new UI that shows the frequency distribution of currently playing audio effect, and the user will be able to set the frequency and the intensity threshold dynamically in the same UI.

Acknowledgment

This work has been supported by the S/W Computing R&D program of MKE/KEIT (10035360, TAXEL: Visio-haptic Display and Rendering Engine).

VI. References

[1] J. W. Lee, J. M. Lim, H. S. Shin, K. U. Kyoung, “SHIFT: Interactive Smartphone Bumper Case”, in Proc. of EuroHaptics 2012.

[2] Carpi, F., Rossi, D., Kornbluh, R., Pelrine, P. and Sommer-Larsen, P. Dielectric Elastomers as Electromechanical Transducers: Fundamentals, Materials, Devices, Models and Applications of an Emerging Electroactive Polymer Tech nology. Elsevier, Jan. 2008.

[3] Ménélas, B., Picinalli, L., Katz, B. F. G. and Bourdot, P. “Audio Haptic Feedbacks for an Acquisition Task in a Multi-Target context” CNRSLIMSI, BP. 133 91403 Orsay, France

[4] T. Beamish, K. v. d. Doel, K. Maclean, and S. Fels, “D’Groove: A Haptic Turntable for Digital Audio Control,” in Proc. of ICAD, Boston, MA, 2003.

[5] L. Baillie, D. Beattie, L. Morton, “Feel what you hear: haptic feedback as an accompaniment to mobile music playback”, in Proc. of IwS’11, Stockholm, Sweden,

[6] Y. Ichiyanagi, E. Cooper, V. Kryssanov and H. Ogawa, “A Haptic Emotional Model for Audio System Interface”, In proc. of: Human-Computer Interaction, 14th International Conference, HCI International 2011, Orlando, FL, USA, July 9-14, 2011,

Parallel Implementation of Aggressive PNN Method for Devices with GPUs

Akiyoshi Wakatani (Konan university, JAPAN), *Member, IEEE*
Akio Murakami (Konan university, JAPAN), *non Member, IEEE*

***Abstract*— We implement the Aggressive PNN method on GPUs by using CUDA to generate codebooks for VQ compression and the speedup of up to 4.01 is achieved compared with the tau PNN method on the CPU. Our second method enhances the parallelism by using indirect vectors to reduce idle threads. We also improved the algorithm by about 20% by using indirect vectors.**

I. INTRODUCTION

Most of recent CPUs consist of multiple processing cores and some of them are heterogeneous multicore architecture such as Intel Ivy Bridge and AMD fusion architecture. In general, the heterogeneous multicore processor consists of conventional CPUs and GPU, so the GPU should be utilized for the general purpose computation in order to achieve a high performance on it. Therefore it has been more important that real applications, including image processing and image coding, should be implemented on the device having the heterogeneous multicore processor or the GPU. Thus, GPGPU (General Purpose computing on Graphic Processing Unit) attracts a great deal of attention [1].

The VQ (Vector Quantization) compression is one of the most important methods for compressing multimedia data, including images and audio, at a high compression rate. A key to achieving a high compression rate is to build an efficient codebook that represents the source data with the least quantity of the bit stream, but it requires lots of computational resources [2]. In this paper, Aggressive PNN method [3], which has been developed for the codebook generation on multicomputers with distributed memory, is applied to GPU by using indirect vectors and its effectiveness is evaluated.

NVIDIA's GPU consists of several MPs (Multi Processors), and several kinds of memory units including the global memory and the shared memory. Each MP contains 8 SPs (Stream Processors).[1)] Each SP executes a thread and plural threads share the shared memory for the data transfer with each other at a high speed. Accesses to the global memory are very slower than the shared memory. In order to avoid such a difficulty, we use the coalesced communication, which can coalesce several access requests from SPs to adjacent addresses of the global memory into one memory request. Meanwhile, each MP executes the computation collectively with 32 threads. This is called "warp."

II. CODEBOOK GENERATION FOR VQ COMPRESSION

PNN (Pairwise Nearest Neighbor) method [4] consists of the following four steps: 1) select initial training vectors with the size of T, 2) calculate all the distances between vectors, 3) find the minimal distance and merge the vector pair with the minimum, and 4) iterate the second and third steps until the size of vectors reaches K. The second step is called "distance calculation step," and the third step is called "merge step." In order to reduce the computational complexity, tau PNN method has been proposed [5], but this method is inherently sequential, so the effectiveness of parallelization for it is not so large. Aggressive PNN method merges plural vector pairs with the first L minimums, where L is given in advance. This L is called Aggressive parameter. When the Aggressive parameter is large, the quality of generated codebook might be worse. We will consider it later.

We apply the Aggressive PNN method to the GPGPU in two ways. The first method (aPNN1) is that each SP is in charge of one vector in the distance calculation step to calculate all the distances of the other vectors and find the nearest neighbor vector. It should be noted that all the SP are not necessarily active because the distance calculation is needed for only vectors whose nearest neighbor vectors are merged. The merge step is done by one thread, except for the data transfer between the global memory and the shared memory. The data transfer is a coalesced communication done by plural threads.

The second method (aPNN2) utilizes indirect vectors. For example, in order to access array elements A[0], A[2], A[9] and A[10] consecutively, the accesses to A[K[0]], A[K[1]], A[K[2]] and A[K[3]] can be assigned to consecutive threads if K[0]=0, K[1]=2, K[2]=9 and K[3]=10. Thus, if the vectors which the distance calculation is needed for are assigned to consecutive threads by using the indirect vector, there are no idle threads. It should be noted that the usage of indirect vectors increases the memory access cost to the global memory.

III. EXPERIMENT AND DISCUSSION

We implement our methods on GPGPU1 system with GeForce 9500 GT (compute capability 1.1) and GPGPU2 system with Tesla C1060 (compute capability 1.3), and evaluate the elapsed time of generating codebooks with the size (K) of 512 from training vectors with the size (T) of 2048.The GPGPU1 consists of NVIDIA GeForce 9500 GT (32 cores), Intel Core 2 Duo E8400 (3 GHz) and 4 GB memory under OS Windows 7 Professional and CUDA toolkit 3.2. The GPGPU2 consists of NVIDIA Tesla C1060 (240 cores), AMD Phenom IIx4 945 (3 GHz) and 4 GB memory

[1)] We are currently implementing our methods on Fermi GPU (GeForce GTX 590).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

under OS Windows 7 Ultimate and CUDA toolkit 3.2. Figures 1 and 2 indicate the elapsed times on both the GPGPU1 and the GPGPU2.

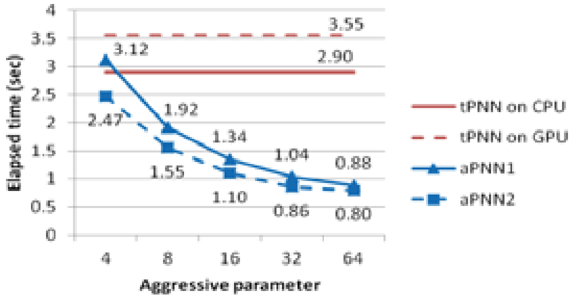


Figure 1: Experimental results (GPGPU1)

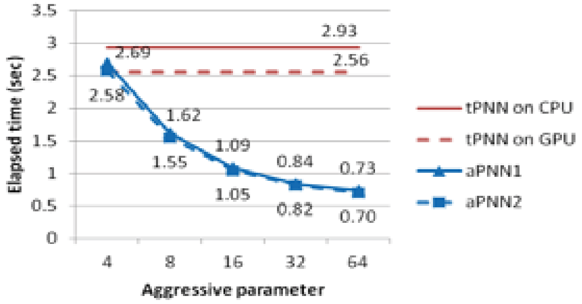


Figure 2: Experimental results (GPGPU2)

As shown in Figure 1, the GPU implementation of the tau PNN method on the GPGPU1 is slower than the CPU implementation because of both the overhead of GPGPU and the low parallelism of the algorithm. However, the elapsed time of the Aggressive PNN dramatically decreases as the Aggressive parameter increases. The speedup of over 3.0 is achieved when the Aggressive parameter is 64.The reason is that the parallelism is enhanced due to the increase of the Aggressive parameter and the elapsed time of the merge step is alleviated because of the reduction of the number of the merge steps. As shown in Figure 2, the CPU implementation of the tau PNN method is slower than the GPU implementation on the GPGPU2, but the speedup of the GPU is slight. However, the speedup of up to 4.01 (=2.93/0.73) is achieved when the Aggressive parameter is 64, due to the same reason as the GPGPU1.

Figure 3 shows the ratio of the elapsed times of the aPNN1 and the aPNN2. On the GPGPU1, the aPNN2 is superior to the aPNN1 without regard to the Aggressive parameter, but the superiority is getting small as the Aggressive parameter increases. Each MP executes 16 warps on the aPNN1, while it executes one warp on the aPNN2 when the number of threads is up to 128. As the Aggressive parameter increases, the parallelism also increases, and thus the number of idle threads decreases on the aPNN1. Therefore, the superiority of the aPNN2 decreases when the Aggressive parameter is large

On the other hand, the superiority of the aPNN2 on the GPGPU2 is smaller than that on the GPGPU1. Each MP executes 2 or 3 warps on the aPNN1, while it executes one warp on the aPNN2 when the number of threads is up to 960. The difference of the number of warps to be executed on both algorithms is smaller than on the GPGPU1 case, and the aPNN2 requires the overhead of memory accesses to the global memory because it utilizes indirect vectors.

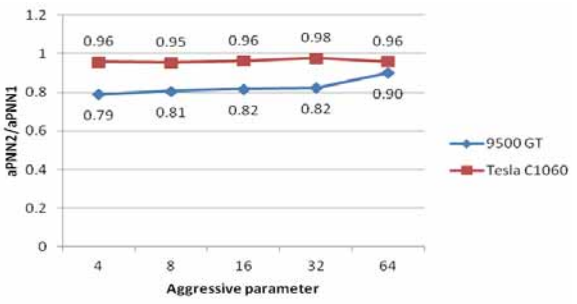


Figure 3: Parallel programs 1, 2 and 3 for Lena image

Finally, the quality of the compressed images should be mentioned. The value of PSNR (Peak Signal Noise Ratio) of compressed images by codebooks produced by the tau PNN method is higher than that by the Aggressive PNN method by 0.05 dB when the Aggressive parameter is up to 32. However, the difference of PSNRs increases when the Aggressive parameter is over 64. Thus, it is important to select the efficient Aggressive parameter in consideration of the quality of the codebook and the parallelism of the algorithm

IV. CONCLUSIONS

A key to achieving a high compression rate of VQ compression is to build an efficient codebook, but it requires lots of computational resources. We proposed a parallel algorithm suitable for GPGPU systems, called Aggressive PNN method, which exploits the parallelism of GPUs with keeping the quality of generated codebooks excellent. The speedup of up to 4.01 is achieved by the Aggressive PNN method compared with the tau PNN method on the CPU. We also improved the algorithm by about 20% by using indirect vectors.

ACKNOWLEDGEMENT

This work was supported in part by MEXT, Japan.

REFERENCE

[1] "Parallel Programming and Computing Platform | CUDA | NVIDIA," http://www.nvidia.com/object/cuda_home.html.

[2] Akiyoshi Wakatani, "Preliminary Implementation of V Q Image Coding using GPGPU," in *Proc. 2010 Int'l Conf. on Consumer Electronics*, P1-12, 2 pages, 2010.

[3] Akiyoshi Wakatani,, "Parallelization of VQ codebook generation using lazy PNN algorithm," Parallel Computing: Software Technology, Algorithms, Architectures & Applications, pp. 415-422, 2004.

[4] Timo Kaukoranta, Pasi Franti and Olli Nevalainen, "Fast and space efficient PNN algorithm with delayed distance calculations," in *Proc. 8th Int'l Conf. on Computer Graphics and Visualization*, pp. 219-224, 1988.

[5] William Equitz, "A new vector quantization clustering algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 10, pp. 1568-1575, 1980.

An IEEE 802.15.4g SUN OFDM-Based RF CMOS Transceiver for Smart Grid and CEs

Seung Sik Lee, Byounghak Kim, Jae Young Kim, Sangsung Choi, and Changwan Kim

***Abstract*—The proposed SUN OFDM-based RF CMOS transceiver, with replacement of ZigBee, can be used as not only an energy saving intelligent green home, which is related to Smart Grid, but also an universal remote controller, a Building Automation, and so on. With our proposed SUN OFDM RF transceiver, wireless connectivity among CEs (Consumer Electronics) devices and electric meters can save electronic energy and make lives more comfortable. The proposed RF transceiver consists of a RF front-end, a TX BBA (Base Band Analog), a RX BBA, and a PLL. Re-using of a *LC* resonator in the RF front-end can reduce the chip size of RF transceiver, DC power consumption, and the cost. The proposed RF transceiver is implemented in a 0.18-μm CMOS technology and consumes 37-mA in TX and 38-mA in RX mode from a 1.8-V supply voltage. In addition, with a fabricated RF transceiver chip, we have succeeded a public demonstration.**

I. INTRODUCTION

In these days, increasing number of digital household CE devices, HDTV (high definition television), entertainment system, personal computers and other Internet consumer devices and uprising electronics charges demand real time checking of power consumption for each devices. Add to that, advanced metering infrastructures (AMIs) for water meter and gas meter are needed for eco-friendly life. The smart utility network (SUN) will be a good solution [1].

The SUN system is a telemetry system closely related to the smart grid framework, which targets designing a modernized electricity network as a way of addressing energy independence, global warming, and energy response.

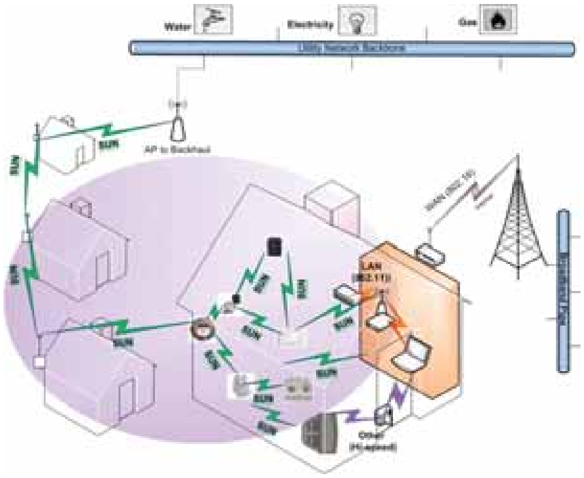


Fig. 1. SUN conception diagram.

This work was supported by the IT R&D program of MKE/KEIT [10035236, development of wireless system for smart utility network] and by National Research Foundation of Korea Grant funded by the Korean Government (2010-0011679).

The SUN system is a ubiquitous network that facilitates efficient management of utility services. It is expected to encourage energy conservation and reduces resources wastage. The conception of SUN is shown in Fig. 1 and its standardization has been processed in IEEE 802.15.4g working group. The first of all, AMIs for gas, water, and power metering will be a good application of SUN. It can also be used for PC peripherals and personal healthcares. Considering replacement of ZigBee, a URC (Universal Remote Control) and home control are also good applications of SUN [2]. Fig.2 shows SUN applications.

The SUN system needs two kinds of system. One is a very low power and low data-rate system by adopting a frequency shift keying (FSK) and the other is a high data-rate system using an orthogonal frequency division multiple access (OFDM). In this paper, a high data-rate OFDM based system will be described.

AMI for Water Meter
AMI for Power Meter
AMI for Gas Meter
URC
SUN Application
PC Peripheral
Building Automation
Home Control
Personal Healthcare

Fig. 2. SUN applications.

II. SUN OFDM RF TRANSCEIVER

Fig. 3 shows the block diagram of the proposed SUN OFDM RF transceiver, which consists of a RF front-end, up/down conversion mixer, base band analog block, and PLL-VCO block.

In Fig. 3, the proposed RF front-end combines a driver amplifier, a low noise amplifier (LNA), and a RF switch. In this configuration, the driver amplifier (DA) and LNA share a common LC resonant circuit, which is simultaneously used as an input matching circuit for the LNA and as an output load for the DA. Therefore, smaller chip-size and lower dc power consumption in the RF front-end have been achieved.

To suppress common-mode noise signals and LO leakages, the proposed transceiver adopts double balanced topology up-conversion mixer. The proposed mixer adopts a V-I converter circuit as its transconductance stage to satisfy the requirement of peak-to-average power ratio (PAPR) of about +10 dB.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

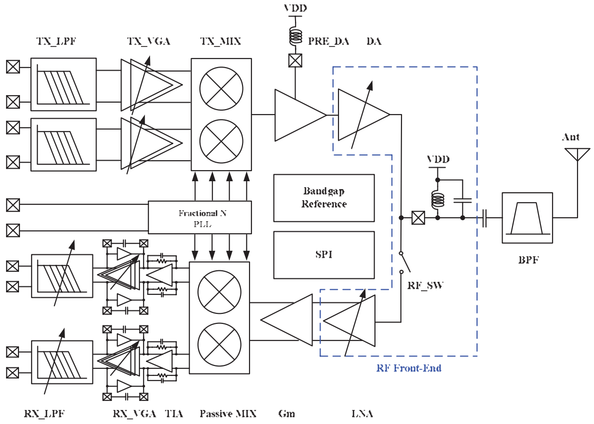


Fig. 3. The block diagram of proposed SUN RF transceiver.

III. Measurement Results

The transmitting channel power can be increased up to +10-dBm but its normal channel power is about 0-dBm due to PAPR of 10 dB. The occupied maximum channel bandwidth is 1.2-MHz and also can be changed into 800/400/200 KHz, which is satisfied with the IEEE 802.15.4g standard. The measured transmitting and receiving spectrum are shown in Fig.4. Fig. 5 shows a PCB board for testing and photograph of the fabricated chip. The fabricated chip is implemented in a 0.18-um CMOS technology and its size is 2.8-mm X 3.0-mm. The fabricated RF transceiver consumes 37mA in TX and 38mA in RX mode for a 1.8V supply voltage.

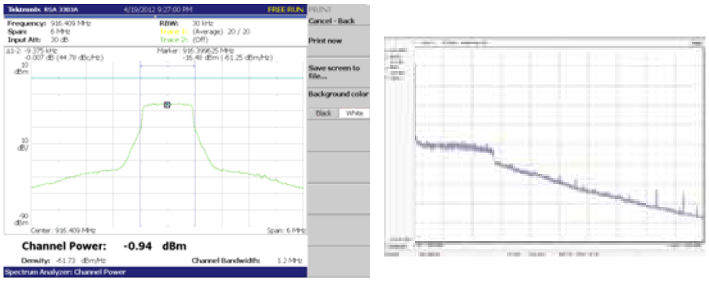

Tektronix RSA 3303A 4/19/2012 9:27:00 PM FREE RUN
Frequency: 916.409 MHz
Span: 6 MHz
Input Att: 30 dB
RBW: 30 kHz
Trace 1: (Average) 20 / 20
Trace 2: (Off)
Δ1-2: -9.375 kHz -0.007 dB (44.70 dBc/Hz)
Marker: 916.399625 MHz -16.48 dBm (-61.25 dBm/Hz)
Center: 916.409 MHz
Span: 6 MHz
Channel Power: -0.94 dBm
Density: -61.73 dBm/Hz
Channel Bandwidth: 1.2 MHz
Spectrum Analyzer: Channel Power

Fig. 4. The transmitting (left) and receiving (right) spectrum of proposed SUN RF Transceiver.

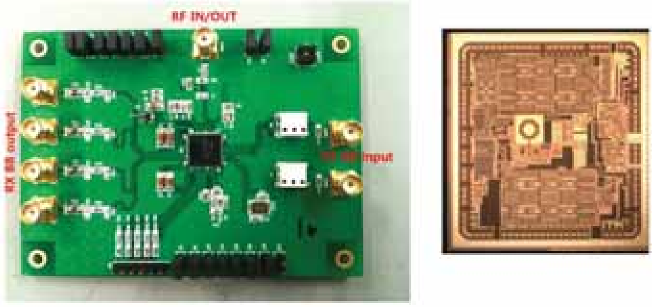


Fig. 5. The test board (left) and photograph (right) of proposed RF transceiver.

IV. Public Demonstration

There are a lot of SUN applications in many wireless connection fields. One of them can be dockyard. During welding, lives and cost can be saved by distinguishing between real fire and sparkle with our SUN system. But it is very difficult to distinguish at control head, because the distance from construction sites to the control head is very long. The public demonstration shows one feasible solution. When a spark is made in sites, a camera takes a picture and sends it to the control head with our SUN wireless system which has adopted the proposed SUN OFDM RF transceiver. Finally, the control head can easily judge fire or not, without checking-fire-processing. The concept and public demonstration are shown in Fig. 6 and 7, respectively.

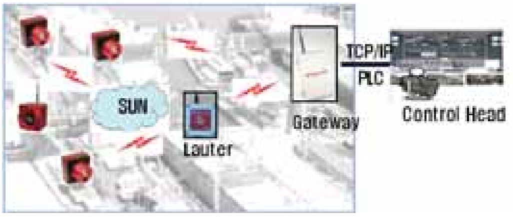


Fig. 6. The conception of public demonstration.

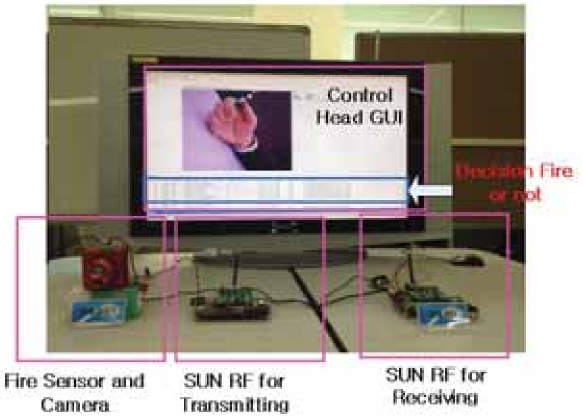


Fig. 7. The public demonstration.

V. Conclusion

A proposed SUN OFDM CMOS RF transceiver is presented. The measured results show a maximum transmitting output power of +10 dBm and dc power consumption of 37mA in TX and 38mA in RX mode from a 1.8V supply voltage. With an implemented SUN OFDM RF transceiver, we have succeeded a public demonstration, which is one of good feasible applications.

Reference

[1] Chin-Sean Sum, H. Harada, Zhou Lan, and Funafa, R, "Smart Utility Networks in TV White Space," *IEEE Comm. Magn*, vol. 49, pp. 132-139, July. 2011.

[2] Wan-Ki Park, Intark Han, and Kwang-Roh PArk, "ZigBee based Dynamic Control Scheme for Multiple Legacy IR Controllable Digital Consumer Devices," *IEEE CE. Trans.*, vol. 53, pp. 172-177, Feb. 2007.

Distributed Multicast Protocol Based on Beaconless Routing for Wireless Sensor Networks

Hosung Park, Jeongcheol Lee, Seungmin Oh, Yongbin Yim, and Sang-Ha Kim, *Member, IEEE*

***Abstract*—Distributed geographic multicast protocols are efficient and scalable for wireless sensor networks but could not be applied to beaconless routing. We propose a novel distributed multicast protocol based on beaconless routing without exchanging beacon messages.**

I. INTRODUCTION

Distributed geographic multicasting (DGM) [1, 2] has been considered as an efficient and scalable protocol for wireless sensor networks since each forwarding node determines whether to divide the path instead of using the preconstructed multicast tree. DGM is also tolerable to the network dynamic such as node and link failure since it is stateless approach. On the other hand, beaconless routing [3, 4] is recently proposed to solve the beacon problem of geographic routing. Conventional geographic routing selects next-hop node for packet forwarding based on the position information of their 1-hop neighbors. This information can be gathered by a periodic exchange of beacon messages. To avoid this message exchange, beaconless routing provides a completely reactive routing. In other words, the forwarding node broadcasts the packet without next-hop selection and one neighbor is elected as next-hop node by competition among the neighbors.

This paper focuses on the distributed multicast protocol based on beaconless routing to take advantages of both DGM and beaconless routing. However, existing DGM protocols could not be applied to beaconless routing. DGM protocols uses geographic routing as underlying routing protocol and each forwarding node carries out branching decision based on the two location information: destinations and neighbors. If DGM protocols use beaconless routing as underlying routing protocol, the forwarding node could not perform branching decision since it could not obtain the locations of neighbors from beacon messages.

In this paper, we propose a novel distributed multicast protocol based on beaconless routing (DMPB) for wireless sensor networks. In DMPB, each forwarding node carries out branching decision based on the only locations of destinations without neighbors' locations.

The locations of destinations can be obtained from the header of the packet. Simulation results show DMPB has batter performance in terms of signaling overhead since DMPB could work based on beaconless routing without exchanging beacon messages.

II. THE PROPOSED PROTOCOL

The problem we are facing can be described as follows. A data packet generated by a source node is delivered to multiple destinations by distributed multicasting using beaconless routing as underlying routing protocol. Each forwarding node should determine whether to divide the path or not without exchanging beacon messages, i.e. without location information of neighbors. Since this paper is focused on multicast protocol, we assume that the source node is aware of locations of destinations. To know that information, the source node can employ destination location service schemes such as [5]

The source node virtually divides the network into four sectors with itself as the center and sends copied packets separately. Each packet contains locations of destinations located in the sector and a temporary guiding point (TGP). In a sector, a packet is delivered toward TGP which is average position of destinations located in the sector.

Each forwarding node uses locations of destinations and TGP for branching decision. When the forwarding node receives a packet, it checks the number of destinations. If the packet contains only one destination, the forwarding node directly sends the packet to the destination. If the number of destinations is more than one, the forwarding node carries out branching decision. The forwarding node also divides the network into four sectors with itself as the center but turn the quadrant to make TGP become central angle of a sector. This rule could prevent unnecessary branching. If all remaining destinations locate in a sector, the forwarding node does not branch off the path and sends packet toward TGP. Otherwise, the forwarding node separates destinations into two groups based on the line between TGP and itself and sends packet separately. Destination list of each packet is replaced to locations of destinations belonging to each group. If a group has more than one destination, the forwarding node calculates new TGP for the group and replaces packet's TGP as new TGP. This branching decision process is repeated until all destinations receive the packet. After all destinations receive the packet, a trajectory the packet pass becomes multicast tree.

Figure 1 shows the example of branching decision process. N5 is no need to carry out branching decision since the packet contains only one destination. The packet to D1 directly delivered by beaconless routing. N2 and N3 do not branch off the path and forward the packet toward TGP1 since all destinations D2-D8 located in a sector. N4 branches off the path since destinations locate more than one sector. N4 separates destinations into two group, D2-D4 and D5-D8,

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

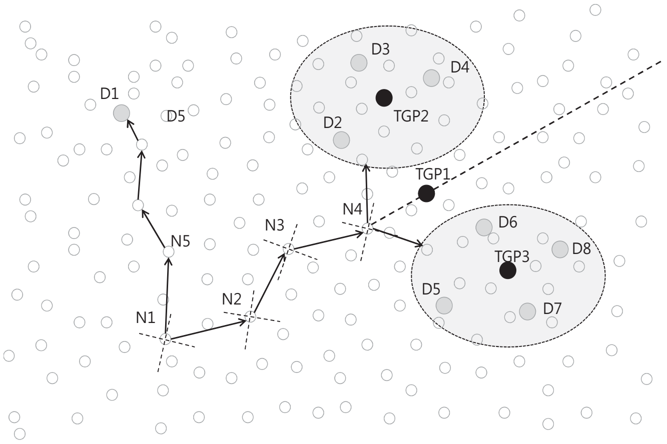


Fig. 1. Example of branching decision process

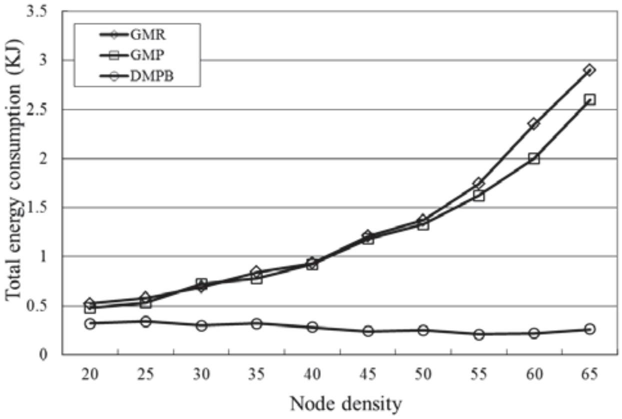


Fig. 2. Average energy consumption in respect of node density

based on the extended line passing N4 and TGP1. N4 calculates TGP2 for D2-D4 group and TGP3 for D5-D8 group and then forward packet to TGP2 and TGP3 separately.

III. Performance Evaluation

In this section, we present simulation results to evaluate performance of DMPB. The purpose of simulations is verification that DMPB has less energy consumption than distributed geographic multicast protocols, GMR [1] and GMP [2]. We simulate DMPB on QualNet simulator [6], The simulation network space is 1000m X 1000m. The number of destinations is 20. Node and destination placement follows random deployment. The transmission range of each node is 30m. A sensor node's transmitting, receiving, and idling power consumption rates are 21, 15, and 0.03mW respectively. The device parameters are chosen in reference the MICA specification [7].

Figure 2 shows total energy consumption in respect of the node density. The node density is the average number of sensor node in 100m X 100m space. Energy consumption of GMR and GMP is rapidly increases as the node density increases. In GMR and GMP, every node periodically exchanges beacon messages with its neighbors regardless of data forwarding. Therefore, as the node density increases, energy consumption for exchanging beacon messages also increases. DMPB consumes less energy than GMR and GMP and is not influenced by the node density since DMPB do not use beacon messages. Energy consumption of DMPB slightly decreases as the node density increases since the probability of electing batter next-hop node increases.

IV. Conclusion

Beaconless routing is the most energy efficient routing protocol in resource-constrained wireless sensor networks. However, existing distributed multicast protocols could not be applied to beaconless routing since they use location information of neighbors for branch decision. This paper proposes a novel distributed multicast protocol based on beaconless routing. In the proposed protocol, each forwarding node carries out branching decision based on the only locations of destinations without neighbors' locations. The proposed protocol has batter performance in terms of energy consumption since it eliminates signaling overhead for exchanging beacon messages.

References

[1] J. Sanchez, P. Ruiz, J. Liu, and I. Stojmenovic, "Bandwidth-Efficient Geographic Multicast Routing for Wireless Sensor Networks," IEEE Sensors Journal, vol.7, no. 5, pp. 627-636, May 2007.

[2] S. Wu and K. S. Candan, "Demand-scalable geographic multicasting in wireless sensor networks," Computer Communications, vol. 30, pp. 2931-2953, Oct. 2007.

[3] H. Fußler, J. Widmer, M. Kasemann, M. Mauve, and H. Hartenstein, "Contention-Based Forwarding for Mobile Ad Hoc Networks," Ad Hoc Networks, 1(4):351 – 369, 2003.

[4] M. Heissenbuttel, T. Braun, T. Bernoulli, and M. Wachli, "BLR: Beacon-Less Routing Algorithm for Mobile Ad-Hoc Networks," Elsevier's Computer Communications Journal (ECC), 27(11):1076–1086, July 2004.

[5] D. Liu, I. Stojmenovic, and X. Jia, "A scalable quorum based location service in ad hoc and sensor networks," in Proc. IEEE Int. Conf. Mobile Ad-Hoc and Sensor System (MASS), Oct. 2006.

[6] Scalable Network Technologies, Qualnet, [online]. Available: http://www.scalable-networks.com

[7] J.Hill and D. Culler, "Mica: a wireless platform for deeply embedded networks," IEEE Micro, vol. 22, no. 6, pp. 12-24, Nov./Dec. 2002.

Embedded Software Development Kit for Sustainable Distributed Mobile Geocast

Hiroyuki Kasai, The University of Electro-Communications, Tokyo, Japan

Abstract — **Our new distributed mobile cache system, a sustainable distributed Geocast technology, enables data caching temporarily in a designated local area. We released the open software development kit (SDK) for embedded systems. This paper explains details of the developed open SDK, an implementation guide, and three applications using this SDK[1].**

I. INTRODUCTION

In order to provide asynchronous communication, an area-based caching capability in a designated local area is expected to develop to foster a sustainable social network. One such technology is Abiding Geocast [1]. As one practical system of Geocast, we earlier proposed a new area-based distributed mobile cache system [2] formed by distributed multiple terminals. It is maintained by collaboratively sharing and relaying cache data among these terminals that pass in or near the local area. The cache data are not propagated over other areas in a borderless fashion, which enables applications to leverage this cache system anywhere such as in crowded city areas or in large exhibition centers. As described in this paper, we deployed and released this open software development kit (SDK) for embedded systems. This paper explains the developed open SDK, an implementation guide, and three applications.

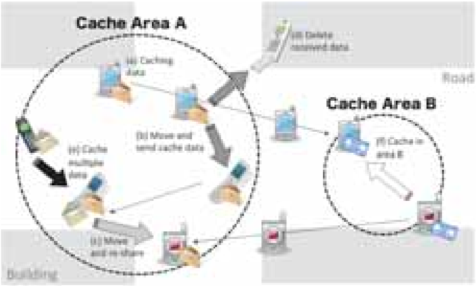


Fig. 1. Area-based Distributed Mobile Cache System [2].

II. AREA-BASED DISTRIBUTED CACHE [2]

This cache system relays and shares data across multiple devices via short-range wireless communication, thereby producing an area-based 'virtual cache' through collaborative access of devices while obviating a network infrastructure. The basic concept is presented in Fig. 1. Fundamentally, every terminal moves freely around in a real physical environment. Once an end terminal decides to cache data in a current area, it attempts to store target data attached to its target cache area information (Fig. 1(a)). Next, these cache data are transferred to neighboring terminals (Figs. 1(b), 1(c)). Receiving terminals judge whether the received data should be cached inside the current area or not, and attempt to store it temporarily and re-share it with others. Alternatively, the cached data are deleted outside the area (d). Each end terminal can be a relay terminal candidate for any data in any area. Therefore, it might carry multiple cache data in multiple cache areas (e).

III. OPEN SOFTWARE DEVELOPMENT KIT (SDK)

We deployed and released the open software development kit (SDK) for embedded systems. This SDK, which includes middleware for the sustainable distributed Geocast, is provided to the public as library files. Third-party developers implement this Geocasting functionality by including the special library file into their original application program. The developer can access and control the middleware via an open API defined in the special header file that the SDK contains.

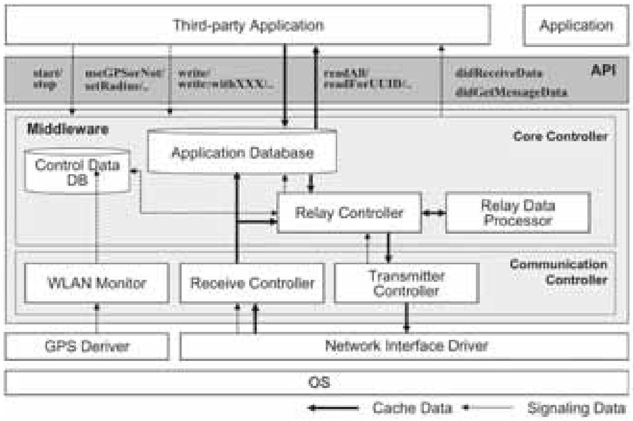


Fig. 2. System architecture diagram and APIs [2].

A. *Middleware Layer Architecture and Implementation*

The middleware, which is implemented by an Objective-C language native module, is separated into two layers: the Core Controller and the Communication Controller. The Communication Controller supports the functionality to access a WLAN or Bluetooth module and to receive/transmit relay data. The Core Controller manages when and to which terminal the internally cached multiple data should be sent via the Communication Controller. Regarding data reception, once the receiving event is detected by the Communication Controller, the Core Controller stores it in the database after data de-multiplexing. Moreover, this manager is executed periodically on a timed basis. This controller deletes the cached data outside its designated cache area. The application data handled by

[1] This work is supported in part by the National Institute of Information and Communications Technology (NICT), Japan, under the new generation network R&D program for innovative network virtualization platform and its application.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

applications are stored in a local SQL database, which is implemented using SQLite. The various control data are also managed in SQLite, of which examples are the current location, a list of devices in proximity. The location is measured through a local GPS module or detectable access points (APs). Such device modes as a Bluetooth module or GPS are useful with a framework equipped in the device OS. One special feature is that the application data to be shared across devices can be encrypted before being written into the middleware from applications. The encryption policy is that the application data can be decrypted only by the same application. This encryption is performed based on a key provided by an application programmer or provider.

B. *Application Programming Interface (API)*

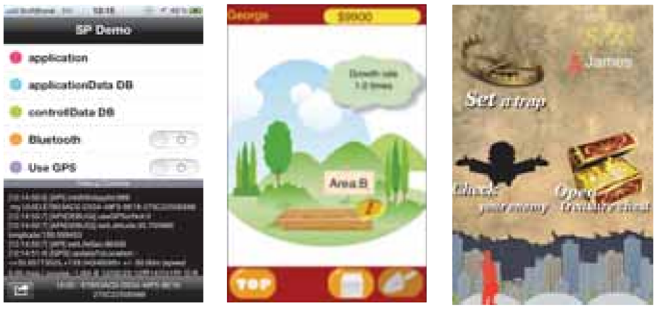
The middleware has various application programming interfaces (APIs) to produce various third-party applications using this distributed mobile storage platform. The first basic application interface is the “**write**” method to write data to be shared on this Geocasting platform (Fig. 3: L13). Here, the location, radius and expiry time for data must be configured using “**setLati:long**” (L9) or “**setRadius**” method before practical writing. Alternatively, it can be done at a time using overloaded methods like “**write:withLati:long:radius**” when writing. The second basic interface includes data-reading methods to retrieve cached data from the local middleware, which are “**readAll**” method for reading all of stored data, or “**readForUUID**” method to retrieve the data by creating a device with the indicated UUID. Next, regarding how to detect data arrival, a special delegate method “**didReceiveData**” that is registered with the middleware informs the application (L17). The common data format used in these methods includes a UUID of its device of origin, a timestamp when the data was generated, the central location data to be cached, the radius size, and the application data itself. As a notable useful method, a message notification method from the middleware is “**didGetMessage**”, which includes “level” and “number.” The “level” has a notification, a warning and an error, whereas “number” indicates a detailed message such as “Writing data is too big” or “A critical error occurred in Bluetooth communication.” Fig. 3 shows sample application code.

```
 1  #import "SPSpotPocket.h"
 2  @interface AppDelegate : UIResponder< SPSpotPocketDelegate>
 3
 4  // initialization, middleware starting, and data sending
 5  - (BOOL)applicationDidFinishLaunching:(UIApplication*)application{
 6     // make an instance
 7     SPSpotPocket* sp= [[SPSpotPocket alloc] initWithInterval:60];
 8     [sp useGPSorNot:YES];                                   // use GPS
 9     sp.delegate = self;                                // use delegate
10     [sp start];                                   // start middleware
11     [sp setRadius:1000];                     // set default cache range
12     [sp setLifeSec:60*60*6];                 // set default expire time
13     [sp write:@”test”];              // write a message into middleware
14  }
15
16  - (void)sp:(SPSpotPocket*)sp didReceiveData:(NSDictionary*)dic {
17     NSArray *datas = [spotPocket readAll];;              // data receiving
18  }
```

Fig. 3. Sample pseudo code for application development using SDK.

IV. Implemented Applications

Three applications were used to demonstrate how SDK is useful. The first is a verification test application that helps application developers to interpret the SDK behavior, and to ascertain whether it operates properly. The application provides developers various user functions to write/read application data to/from the middleware, to see cached data, to simulate device location change, and to check detectable devices in proximity. Screenshots are portrayed in Fig. 4. The other two applications are geo-location social games depicted in Fig. 4. They enable users to enjoy games with passengers who walk through designated areas when commuting to school or work. The first is for users to grow vegetables with other people in an asynchronous manner. The vegetable fields are shared with people. They collaborate to water the vegetables, fertilize them, and destroy insects. In the other one, users compete with other people by virtually setting traps in the physical area. When others are trapped when walking through a trap-area, the trap-setter gains a reward. The user sets a couple traps on the way to school in the morning. Then the user enjoys checking whether other people were trapped or not. Of course, the user might also be caught by the traps by the other people when entering to check the results.



(a) Verification Test (b) Vegetable Grow (c) Trap Game

Fig. 4. Applications using Deployed SDK.

V. Conclusion

The deployed SDK, which is flexible and easy to use, has great potential to drive new applications such as a geo-local message exchange and targeted advertisement. Real-time social games are also an important target application because they exemplify rapidly expanding applications [3], and have achieved explosive popularity and a large business market.

References

[1] C. Maihöfer, T. Leinmüller, and E. Schoch, (2005). Abiding Geocast: Time-Stable Geocast for Ad Hoc Networks. The Second ACM International Workshop on Vehicular Ad Hoc Networks (VANET 2005).

[2] H. Narimatsu, H. Kasai, and R. Shinkuma, “Area-based Collaborative Distributed Cache System using Consumer Electronics Mobile Device,” *IEEE Trans. on Consumer Electronics*, vol. 57, no. 2, pp. 564-573, 2011.

[3] A. S. Y. Lai and A. J. Beaumont, “Mobile Bluetooth-Based Multi-player Game Development,” *Ubiquitous Computing. Journal of Computational Information Systems*, vol. 6, no.14, pp. 4617-4625, 2010.

An Non-uniform Sampling Strategy for Physiological Signals Component Analysis

Molin Jia, *Student Member*, *IEEE*, Chaoyang Wang, Kui-Ting Chen, *Student Member*, *IEEE*,
and Takaaki Baba, *Member*, *IEEE*

***Abstract*--The conventional approach cannot meet the requirement of physiological signal analysis to extract the main component of the acquired signal. This paper proposes a non-uniform sampling strategy with corresponding fast Fourier transform (FFT) for signal component analysis. The simulation results of different approaches are analyzed and compared. The proposed strategy has the superior anti-aliasing performance than conventional approach with lower sampling rate.**

I. INTRODUCTION

Several companies and researchers are developing healthcare devices and investigating how to detect vital physiological signals, such as respiration, heartbeat intervals, heart sound, pulse oxygen saturation, blood pressure, body temperature, etc. These vital signals need to be extracted by the bio-signal acquisition systems like electrocardiogram (ECG) devices [1]. With the rapid progress of wireless technology, the biomedical devices generate a new model for providing healthcare [2]-[4]. The acquired signal is mixed by various physiological signals carrying different characteristics and the noise from wireless environment. Biomedical electronics for healthcare and clinical applications expect not only the accurate physiological signal acquisition of behavioral modifications, but also the capability to extract the information out from the acquired signals [5]. The component analysis of physiological data obtained from patient has created new challenges for healthcare devices to separate the different physiological signals and extract useful information from the raw data.

Conventional methods of signal component analysis including numerous matrix operations are too complex to implement with low overhead, and cannot meet the requirement of signal analysis. Since the different components of acquired signal have own frequency characteristics, the analysis of band limited signal in time-frequency domain can constitute an efficient technique instead of conventional way. The key of signal component analysis is solving the aliasing problem during the sampling and frequency analysis.

Shannon sampling theorem emphasizes that the sampling frequency f_s must be at least twice the highest frequency component f_{max} in the signal or aliasing error will be introduced. Hence, the maximum frequency component determines the maximum sampling frequency. When the signal frequency increases, the sampling rate has to be increased in double speed [6]. High sampling frequency calls for high cost of signal analysis systems. Shannon theorem applies only to the case of uniform samples [7]. Besides, uniform sampling will cause the aliasing problem which is not acceptable in the system of signal component analysis, since the components of acquired signal cannot be forecasted. Several non-uniform sampling methods have been proposed for signal reconstruction or data compression [7]-[10]. However, none of them can be employed to signal analysis, because the non-uniform discrete Fourier transform is too complex.

With proposed non-uniform sampling and fast Fourier transform (FFT), a novel component analysis strategy for physiological signals is presented in this paper. In this approach, the highest frequency of non-uniform sampling can be lower than Shannon sampling frequency. More significantly, the aliasing problem happened in the uniform sampling during the FFT operating can be effectively compressed in the proposed strategy, and much purer analysis results can be achieved with higher reliability.

The rest of the paper is organized as follows. In Section II, the whole architecture of the system is displayed at the beginning. Then, the component analysis strategy with proposed non-uniform sampling and FFT is demonstrated from three sub-sections. First, non-uniform sampling approach is presented. Second, the corresponding FFT is explained. Moreover, the aliasing compression is discussed, and a further method for reducing the aliasing is introduced. Section III shows the simulation results and the comparisons between different methods. The conclusion of this study is shown in Section IV.

II. SIGNAL COMPONENT ANALYSIS BASED ON NONUNIFORM SAMPLING AND FFT

The system concept of proposed strategy is shown in the Fig. 1. The input is the acquired physiological signal. The output is the analysis result by the proposed strategy. It mainly consists of three parts (A, B and C). In the method of proposed non-uniform sampling, the signal are sampled by non-uniform impulse response and the sampled composite signal is going to be resolved to several uniform sampled discrete signals according to the Fourier series expansion. In the second part, after the uniform FFT operation for the uniform discrete sampled signals, the uniform FFT results are synthesized and the major component of the signal will be extracted. In addition, we are going to further compress the aliasing caused by multi-component of sampling signal.

A. Proposed non-uniform sampling

Any acquired physiological signal can be described as the following equation.

This research was supported by Waseda University Grants for Special Research Projects, and also by funds from MEXT via Kitakyushu innovative cluster projects, Japan.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

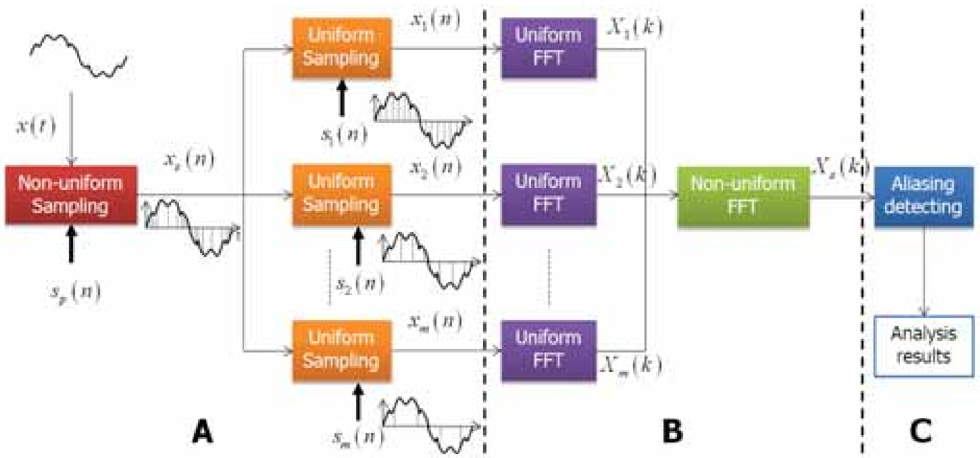


Fig. 1. The system concept of proposed strategy for non-uniform sampling and FFT

$$x(t)=\sum_{i}x_i(t)+\sigma \tag{1}$$

$x_i(t)$ are the different components of the acquired signal. σ is the interfering noise.

Sampling provides the bridge between the continuous and the discrete-time worlds. The impulse response of non-uniform sampling is defined as

$$s_p(n)=\delta(t-nT(t)) \tag{2}$$

where, *$T(t)$*, the function with the variable t , is the sampling interval. n presents the number of non-uniform sampling point. As a result, the sampling signal is not uniform. Then, the sequence of non-uniform samples can be expressed as

$$x_s(t)=\sum_{n_p}x(nT(t))\bullet s_p(n) \tag{3}$$

The aiming is to develop a fast algorithm to analyze the samples components in frequency domain as shown in the following.

$$X_s(k)=\sum_{n_p=0}^{N-1}x_s(n)e^{-jw_k^p n_p} \tag{4}$$

Conventional discrete Fourier transform cannot meet this non-uniform sampling approach within simple calculation.

On the basis of Fourier series expansion, the equation (2) (sampling response) can be resolved to

$$s_p(n)=\sum_{m}c_m e^{j2\pi mt/T} \tag{5}$$

We can define the sampling response and make the c_m equal to a constant. Thus, any non-uniform sampling response can be presented as

$$s_p(n)=\sum_{m}\lambda_m s_m(n) \tag{6}$$

$S_m(n)$ is uniform sampling response with multi-level harmonics. m is the harmonics number of sampling signal. λ_m can be unitized by the sampling signal definition. The equation (3) (non-uniform samples) can be shaped as

$$x_s(t)=\sum_{n_p}\left[x(nT(t))\bullet\sum_{m}s_m(n)\right] \tag{7}$$

which can be further transformed to

$$x_s(n)=\sum_{m}x_m(n)=\sum_{n,m}\left[x(nT_m)\bullet s_m(n)\right] \tag{8}$$

wherein *$x_m(n)$* is the samples under uniform sampling predefined before sampling with different sampling rate. T_m is the sampling period in the mth sampling components.

So far, the work of proposed strategy in time domain is completed. The non-uniform sampling is resolved into the combination of several uniform samplings with different rates. We can predefine the number and rates of different sampling signal components to predefine the m and n and shape the non-uniform sampling response.

B. Corresponding FFT and synthesis

For non-uniform samples, conventional FFT is unable to easily analyze the signal components in frequency domain, because w^p and n_p in equation (4) are not uniform transforming.

In accordance with the equation (8), in the proposed non-uniform sampling strategy, the equation (4) can be transformed to

$$X_s(k)=\sum_{n=0}^{N-1}\left\{\left[\sum_{m}x_m(n)\right]e^{-j2\pi kn/N}\right\} \tag{9}$$

Then, according to the linear characteristic of FFT, this equation can be further inferred to

$$X_s(k)=\sum_{m}\sum_{n=0}^{N-1}x_m(n)\cdot e^{-j2\pi kn/N} \tag{10}$$

wherein, as well known

$$\sum_{n=0}^{N-1}x_m(n)\cdot e^{-j2\pi kn/N}=X_m(k) \tag{11}$$

Therefore, the simple transform of equation (9) can be achieved as the core strategy of signal analysis in frequency domain, as shown in the following.

$$X_s(k)=\sum_{m}X_m(k) \tag{12}$$

in which, *$X_m(k)$* is the uniform FFT for the samples from one uniform sampling components in proposed non-uniform sampling. The non-uniform FFT corresponding to the non-uniform sampling can be resolved to several uniform FFTs. The synthesis of several uniform FFTs can return to the original result of non-uniform FFT.

In short, we achieved an efficient method for signal analysis by predefining the non-uniform sampling signal and utilizing the linear characteristic of FFT.

C. Aliasing compression processing

In the uniform sampling, aliasing problem will happen in the frequency response of the acquired signal shown in equation (1). The frequency response includes two parts, original response part and aliasing part, as shown in the follows.

$$\left|X(k)\right| = \sum_i \left|X_i(k)\right| + \sum_i \left[\left|X_a(rk^s \pm k_i^x)\right|\right] + \sigma(k), \quad r \in N^+ \tag{13}$$

X_a(k) is defined as the aliasing response. k^s is the frequency of uniform sampling signal. k^x presents the different frequencies component of the acquired signal. $\sigma(k)$ is defined as the frequency response of white Gaussian noise. It appears that the power of aliasing part is same with the original part. It is too difficult for signal analysis system to identify the components.

In the proposed strategy, the original frequency response is enhanced and the aliasing part is compressed. The frequency response is described by

$$\left|X_s(k)\right| = \sum_i m\left|X_i(k)\right| + \sum_m \sum_i \left[\left|X_a(rk_m^s \pm k_i^x)\right|\right] + \sigma(k) \tag{14}$$

k^s presents the different frequencies component of sampling signal. Because the proposed strategy meets to band limited signal, according to FFT feature, Equation (14) is changed to

$$\left|X_s(k)\right| = \sum_i m\left|X_i(k)\right| + \sum_m \sum_i \left[\left|X_a(k_m^s - k_i^x)\right|\right] + \sigma(k) \tag{15}$$

Although more aliasing frequencies are introduced, the different components of original frequency response have been enhanced by *m* times in comparison with the aliasing part. Therefore, the system has better anti-aliasing ability and can easily distinguish the original signals with different physiological information.

However, there is a situation that can introduce the aliasing into the analysis results. If $m \leq i$ and

$$k_m^s - k_l^x = k_{m-1}^s - k_{l-1}^x = \ldots = k_1^s - k_{l-m+1}^x = \Delta k \tag{16}$$

Equation (15) transforms to

$$\left|X_s(k)\right| = \sum_i m\left|X_i(k)\right| + mX_a(\Delta k) + X_a\left(k_m, k_i\right) + \sigma(k) \tag{17}$$

The power of Δk is also be enhanced by m times and mixed to the component analysis results of real physiological information.

Here, we propose a method to remove this aliasing by detecting the following matrix.

$$\begin{bmatrix} k_m^s - k_1^x & \cdots & k_m^s - k_i^x \\ \vdots & \ddots & \vdots \\ k_1^s - k_1^x & \cdots & k_1^s - k_i^x \end{bmatrix} \tag{18}$$

If the value appearing with most times in equation (18) is one frequency of *X_i(k)*, this component will be abandoned from *X_i(k)*. Then, the correct analysis results of acquired physiological signal can be figured out.

III. SIMULATION RESULTS AND COMPARISONS

A. Simulation parameters

To investigate the signal analysis ability of the proposed strategy, the simulation of the proposed and conventional approaches are implemented and compared in MATLAB R2006b. Since the frequency of heart sound signal or the ECG signal is within the range from 0Hz to 500Hz as a band limited signal, the scope of component analysis is defined in the same range. Several simulations are performed with different methods, sampling rates, frequency components of acquired signal. The parameters are defined in Table I. In order to facilitate the observation and comparison, the unit is Hz.

TABLE I
SIMULATION PARAMETERS FOR DIFFERENT METHODS

	Simulation (a)	Simulation (b)	Simulation (c)
Frequency components	f_1 =10, f_2 =120 f_3=250, f_4 =455	f_1=230, f_2 =250 f_3=280, f_4 =300	f_1 =120, f_2 =270 f_3 =314, f_4 =338
Uniform sampling rates	f_s=1100 $f_s > 2f_{max}$	f_s=750 $f_s > 2f_{max}$	f_s=750 $f_s > 2f_{max}$
Uniform sampling rates	f_s=565 $f_s < 2f_{max}$	f_s=565 $f_s < 2f_{max}$	f_s=565 $f_s < 2f_{max}$
Proposed non-uniform strategy	f_{s1}=659, f_{s2}=709 f_{s3}=673 $f_{smax} < 2f_{max}$	f_{s1}=547, f_{s2}=571 f_{s3}=523 $f_{smax} < 2f_{max}$	f_{s1}=503, f_{s2}=547 f_{s3}=571 $f_{smax} < 2f_{max}$
Proposed non-uniform strategy	with aliasing removing	with aliasing removing	with aliasing removing

B. Results comparison and analysis

Fig. 2 shows the signal analysis results based on conventional uniform sampling. Although the sampling rate is higher than twice the highest frequency component, the main signal components cannot be easily recognized. Fig. 3 describes the signal analysis results based on uniform sampling whose sampling rate is lower than twice the highest frequency component. The results are totally incorrect owing to the aliasing problem, because the aliasing form noise has inundated the main signal component. The signal analysis results based on proposed non-uniform sampling and FFT strategy is described in Fig. 4. In Fig. 4(a) and Fig. 4(b), the propose approach presents the superiors performance in extracting the main signal components. The main signal components are effectively enhanced. Moreover, the highest sampling rate is lower than twice the highest frequency component. However, the aliasing introduced by different sampling signal happens in Fig. 4(c), as described in equation (16). The result shown in Fig. 5(c) removes this aliasing component by employing the proposed further method for aliasing compression as explained at the end of Section II. These simulation results prove that the proposed strategy, of which the sampling rate is lower than Shannon sampling

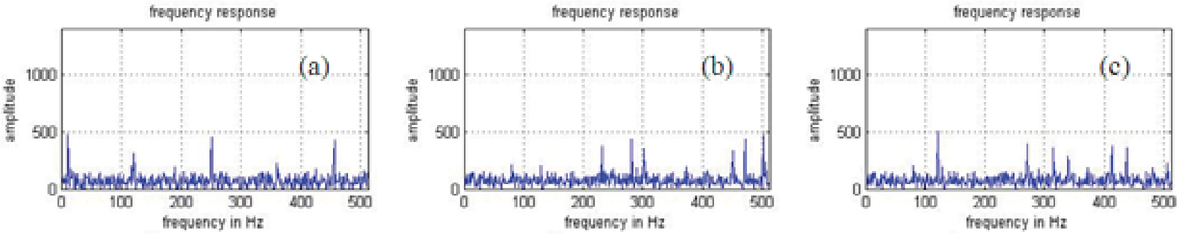


Fig. 2. The signal analysis results based on uniform sampling rates which is higher than twice the highest frequency component

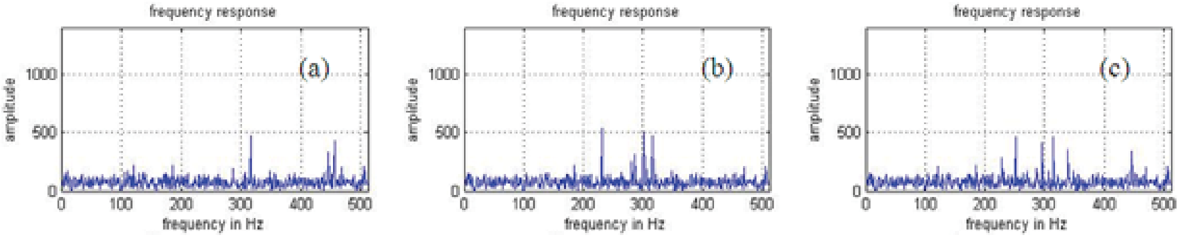


Fig. 3. The signal analysis results based on uniform sampling rates which is lower than twice the highest frequency component

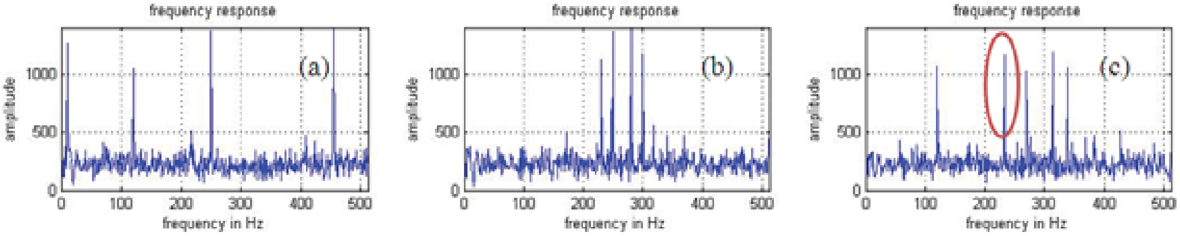


Fig. 4. The signal analysis results based on proposed non-uniform strategy without further aliasing removing

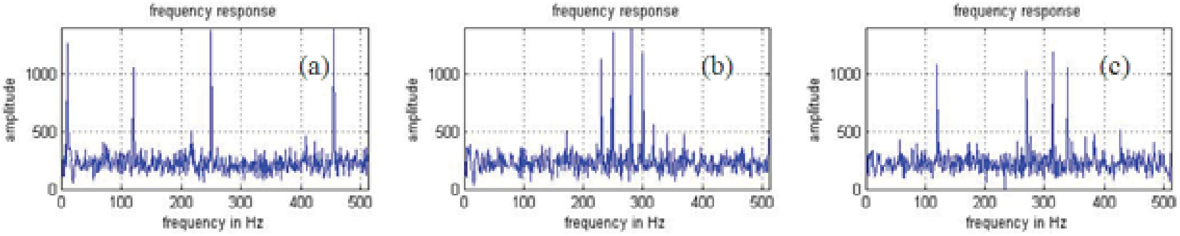


Fig. 5. The signal analysis results based on proposed non-uniform strategy with further aliasing removing

frequency, has the superior performance than conventional approach in signal component analysis, has a better anti-aliasing performance.

IV. CONCLUSION

To meet the requirement of physiological signals analysis, we have proposed a strategy based on non-uniform sampling and FFT for purer extraction of main components and lower sampling rate. The proposed concept of non-uniform sampling was elaborated and the corresponding FFT method was presented. The simulation results prove that the proposed strategy has the superior anti-aliasing performance on signal component analysis than conventional approach and the lower sampling rate than Shannon sampling frequency. In addition, the proposed strategy will be suitable for any other kinds of signal component analysis system.

REFERENCES

[1] J. Lee, and S. Kang, "Heartbeat detection based on filter banks and fuzzy inference for u-healthcare," *IEICE Electronics Express*, vol. 6, No. 13, pp. 936-942, July 2009.

[2] S. -L. Chen, H.-Y. Lee, C.-A. Chen, H.-Y. Huang, and C.-H. Luo, "Wireless body sensor network with adaptive low-power design for biometrics and healthcare applications," *IEEE System Journal*, vol. 3. No. 4, pp. 398-409, Dec. 2009.

[3] A. Boukerche, and Y. Ren, "A secure mobile healthcare system using trust-based multicast shceme," *IEEE Journal on Selected Areas in Communication*, vol.27. No. 4, pp. 387-399, May 2009.

[4] E. Nemati, M. J. Deen, and T. Mondal "A wireless wearable ECG sensor for long-term application," *IEEE Communications Magazine*, pp. 36-43, Jan. 2012.

[5] C. I. Ieong, M.-I. Vai, P.-U. Mak, and P. -I. Mak, "ECG heart beat detection via mathematical morphology and quadratic spline wavelet transform," *IEEE International Conference on Consumer Electronics*, pp. 609-610, Jan. 2011.

[6] Y. Xiong, Y. Huang, P. Sun, M. Evans, and T. Cronk, "A non-uniform sampling tangent type FM demodulation," *IEEE Transactions on Consumer Electronics*, vol. 50, No. 3, pp. 844-848, Aug. 2004.

[7] E. Margolis, and Y. C. Eldar, "Nonuniform sampling of periodic bandlimited signals," *IEEE Transactions on Signal Processing*, vol. 56, No. 7, pp. 2728-2745, July 2008.

[8] M. Ben-Romdhane, C. Rebai, A. Ghazel, P. Desgeys, and P. Loumeau, "Non-uniform sampling schemes for IF sampling radio receiver," *International Conference on Design and Test of Integrated Systems in Nanoscale Technology*, pp. 1-9, Sep. 2006.

[9] V. Singh, and N. Rajpal, "Data compression using non-uniform sampling," *IEEE International Conference on Signal Processing, Communications and Networking*, pp. 603-607, Feb. 2011.

[10] S. M. S. Zobly, and Y. M. Kakah, "Compressed sensing: Doppler ultrasound signal recovery by using non-uniform sampling & random sampling," *Radio Science Conference (NRSC)*, pp. 1-9, Apr. 2011.

Seamless and Non-Contact Health Monitoring System in Cloud Computing

Ee-May Fong, Tae-Ha Kwon, and Wan-Young Chung, *Member, IEEE*
Department of Electronic Engineering, Pukyong National University, Busan 608-737, South Korea

***Abstract*-- Promising developments in healthcare technology have fostered an interest in noninvasive and indirect contact capacitive coupled ECG measurement technique. In this paper, we propose an indirect contact ECG measurement system to measure biomedical data invasively in cloud computing environment. Bio-signal data can be measured and collected at any time without the presence of any personal assistant or physician. However, patient identity is not recognized in other researches and this may causes patient identity confusion. Medical data may be wrongly recorded and results in a poor medical record management. Thus, ECG identification is important to make sure that the data is coming from the right person. Bio-signal data and patient biological information are stored in a web server monitoring system to enable real-time collection and dissemination of personal health data by patients and health-care professionals at anytime and from anywhere.**

I. INTRODUCTION

Electrocardiography (ECG) is the electrical signal of our heart and it is one of the major vital signals monitored in ubiquitous healthcare. Generally, a nonintrusive biomedical signal monitoring in daily life aims at continuously monitor health-related information without interrupting the subject's ordinary daily activities, without requiring additional operations and cooperation to make measurements, and without the subject's awareness. Thus, capacitive coupled method is applied to measure biomedical signal for a long term monitoring with minimal inconvenience [1]. However, it has its drawback where the patient identity is not recognized. Patient identity confusion may causes serious consequences such as medical data is wrongly recorded, medical history is wrongly retrieved, wrong medical or financial decisions are made based on the data. Thus, ECG identification is important to ensure the biomedical data is coming from the right person. The physiological and geometrical differences of the heart in different individuals display certain uniqueness in their ECG signals [2]-[3]. In this paper, we propose an ECG-based patient identification and remote health monitoring system. Indirect contact ECG measuring approach is adapted to measure ECG signal through chair and human identification is done using non-fidual method where ECG template feature extraction can truly represent the distinctive characteristics of a person. After ECG authentication is done, bio-signal data and patient biological information are uploaded and stored in a web server monitoring system to enable real-time collection and dissemination of personal health data by patients and health-care professionals anytime and from anywhere.

II. SYSTEM DESCRIPTION

Fig. 1 shows the overall system architecture of the indirect contact ECG measurement approach and human identification together with web server health monitoring system. The proposed system consists of three parts: Capacitive coupled ECG measuring sensor from the user from multiple locations, human identification to recognize the user whose ECG signal is being measured, and web server health monitoring system.

A. *Hardware Capacitive Coupled ECG Sensor*

Indirect contact capacitive electrocardiogram (ECG) approach is employed to measure biomedical data. This approach does not require any direct contact between the sensor and the human skin and thus it enables long term healthcare monitoring without disturbing the subject's daily life.

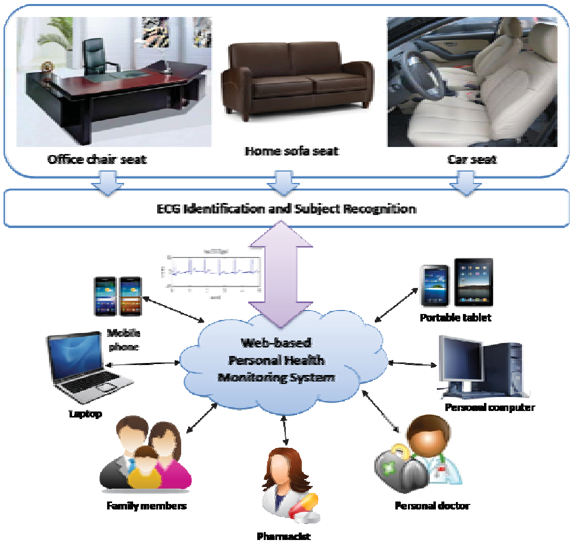


Fig. 1. System architecture of indirect contact ECG measuring and subject identification with web server health monitoring system.

By installing a pair of capacitively coupled sensor electrodes on the chair back and conductive textile on the chair seat, biomedical signal can be measured. These capacitive electrodes are connected to an electronic circuit for analog signal processing. Along with the electronic circuit, a sensor node transmits the bio signal data through IEEE 802.15.4 Zigbee based radio protocol at 250Kbps to pc or mobile phone for human identification processing. This approach can be employed at office chair seat, home sofa seat, or automobile chair seat for long term ECG monitoring.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

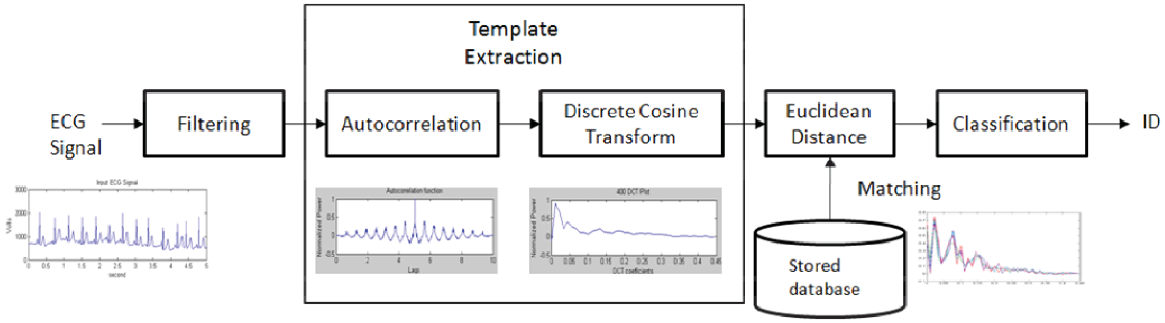


Fig. 3. Subject identification and feature extraction algorithm.

B. *Patient Identification Methodology*

ECG signals can be measured from different users from multiple locations. Without proper biomedical identification, the ECG data may be recorded wrongly and wrong medical decision may occur. This will leads to poor medical record management. Fig. 3 shows the subject identification and recognition algorithm. ECG signal acquired is being digitally filtered again whereas preserving the unique characteristics. Then a feature template is extracted from the user using AC/DCT method [4]–[5]. Feature extraction operates directly on the autocorrelation of a few seconds of ECG to form distinctive personalized signatures for every subject. The, by comparing the template to the history database, subject is identified easily using Euclidean distance method.

III. EXPERIMENTAL RESULTS

To test the feasibility of our approach for ECG based human identification, ECG data are collected from 10 users sitting on a chair seat wearing cotton shirt. ECG signals are measured using noncontact measurement methods. Then, ECG signals are sent through wirelessly to a computer for subject identification purpose. Template extraction extracted from ECG data has very appealing unique characteristics for subject recognition. We applied this template extraction algorithm to 10 series of ECG data recorded from 10 subjects. Fig. 5 shows the comparison of template extraction in 2D and 3D plot from multiple ECG windows from 3 subjects, A, B and C among 10 subjects. Thus, we concluded that human electrocardiogram (ECG) exhibits unique patterns that can be used to discriminate individuals. After subject identification and recognition, subject personal details and biomedical data are uploaded into web server monitoring system. Thus, clinical data and applications are readily available to patients, care takers and physicians through via internet in a community of clouds. Relevant health data may be retrieved from web server as well, enabling intelligent decision making of diagnosis and prognosis from the healthcare service network.

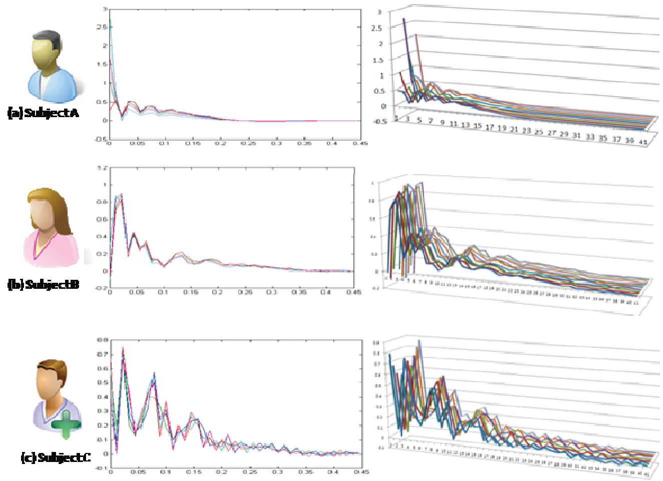


Fig. 5. Comparison of template extraction in 2D plot and 3D plot from multiple ECG windows from different subjects.

IV. CONCLUSIONS

Indirect contact biomedical signal measurement and ECG-based identification is presented for the continuous health monitoring in cloud computing environment. By integrating this technology on a chair seat in office, home, or automobile, biomedical data can be acquired unnoticeably. Biomedical data acquired from any location is processed digitally to identify and recognize the subject whose ECG is being measured. The results demonstrate the validity of long time healthcare monitoring by the technology.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012004140).

REFERENCES

[1] Lopez A, Richardson PC, Capacitive Electrocardiographic and Bioelectric Electrodes, IEEE Transactions on Biomedical Engineering, BME-16(1): 99, 1969.

[2] F. Agrafioti, D. Hatzinakos, ECG Biometric Analysis in Cardiac Irregularity Conditions, International Journal of Signal, Image and Video Processing. Springer London, Sep. 2008.

[3] P.Sasikala, Dr. R.S.D Wahidabanu, Identification of individuals using Electrocardiogram, IJCNS International Journal of Computer Science and Network Security, Vol.10 No.12, Dec. 2010.

[4] Rafik Matta, Johnny K. H. Lau, Foteini Agrafioti, Dimitrios Hatzinakos. Real-time continuous identification system using ECG signals. In Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE) May 8-11, 2011.

[5] Chiu, C.-C., Chuang, C.-M., Hsu, C.-Y.: A Novel Personal Identity Verification Approach.Using a Discrete Wavelet Transform of the ECG Signal. In: Proceedings International Conference on Multimedia and Ubiquitous Engineering, pg201-208, 2008.

Graph-structured Hierarchical Scheduling for Mobile Cloud Computing

Eun-Sung Jung
Samsung Advanced Institute of Technology, Samsung Electronics
Email: es73.jung@samsung.com

Abstract—**Mobile cloud computing is an emerging technology which brings cloud computing services to mobile devices. Recent studies investigated feasibility of mobile cloud computing from the perspectives of computation offloading to the cloud where cloud computing resource are available. However, those research work fails to address issues on multiple available resources including nearby servers and peers and online scheduling algorithms to quickly adapt to the changing environment. We propose a distributed scheduling algorithm, graph-structured hierarchical scheduling, which distributes not only task workload but scheduling workload so that responsive online computation offloading is actionable in mobile cloud computing. The theoretical performance analysis shows that the proposed algorithm outperforms a centralized single scheduler.**

I. Introduction

Cloud computing is getting more prevalent as large scale computing platforms as well as high speed networking infrastructure have become mature. Cloud computing can be summarized as computing as a utility, which can further can be further categorized as IaaS(Infrastructure as a Service), PaaS(Platform as a Service), and SaaS(Software as a Service).

The fast growing number of smart devices such as smart phones has triggered a new emerging research area, mobile cloud computing. Mobile cloud computing provides mobile devices with convenient computing services as cloud computing does desktop computers. For a simple example of mobile cloud computing, a man with a smart phone can run a compute-intensive job by offloading computation to nearby local servers, say, public servers at Starbucks. Formally, it was defined as "the availability of cloud computing services in a mobile ecosystem" in Open Gardens blog.

technologies/challenges As enabling technologies for mobile cloud computing, four key technologies are listed up in [1]. They are collaborative programming abstraction, dynamic workload profiling and scheduling, real-time elastic resource provisioning, and privacy protection. For instance, consider a scenario that one with a mobile phone enters a Starbucks coffee shop, where public access local servers are installed, and wants to run a high-quality graphical game, which can be played in low-quality graphics on the mobile phone itself. First, the workload of the application is profiled and which parts of the application should go to local servers are determined. To run the remote execution codes on local servers, real-time elastic resource provisioning on local servers and collaborative programming abstraction among mobile devices and local servers are needed. In addition, the data transferred to local servers should be protected against possible hostile attacks.

In this paper, we deal with dynamic workload scheduling issues. In the literature of grid/cloud computing[4], [3], a job is represented by a directed acyclic graph (DAG), called workflow, and is scheduled by a centralized scheduler upon arrival of a request to efficiently utilize distributed resources and minimize job completion time. Finally, the partitioned tasks are deployed on remote sites. The recent work in mobile cloud computing[2] takes a little different approach. The call graph of an application is analyzed offline and a scheduler then computes optimal schedules for several possible mobile environments, e.g., low network bandwidth/high compute resource and high network bandwidth/low compute resource. The limitations of [2] are as follows. Since the schedules for an applications are determined offline based on typical execution environments due to restricted computing resources of a mobile phone, the predetermined schedules may not fit into at present execution environment. Moreover, the simplified computation offloading model, which allows only to the cloud, can hardly be extended to situations where many nearby servers, called cloudlet, exist.

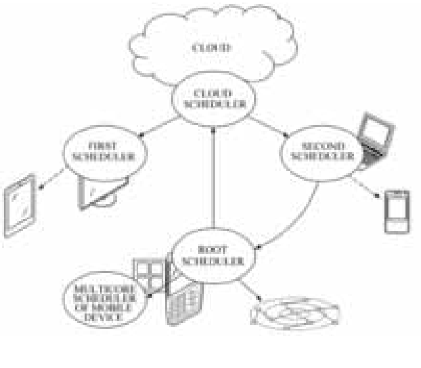


Fig. 1. Illustration of Graph-structured Hierarchical Scheduling

In this paper, we first propose a distributed scheduling algorithm, graph-structured hierarchical scheduling (GHS), which addresses online task scheduling and rescheduling issues relating to mobile cloud computing. We then analyze the performance of the algorithm based on analytical model. Finally we will present conclusions.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

II. Graph-structured Hierarchical Scheduling

The graph-structured hierarchical scheduling (GHS) a distributed scheduling algorithm which distribute not only workload of a job but workload of scheduling itself among multiple compute resources recursively until workload distribution performs better. The workload distribution can be represented by a graph $G = (V, E)$, where a node denotes a compute resource and an edge denotes a master-worker relation between start and end node. In a graph G, only one root node exists, which has no incoming edge, and it matches with a mobile device launching a job.

Figure 1 illustrates one example of GHS. The root node represents a mobile device running a certain application/job. A mobile device user have several available resources including sensor networks, laptops, cloud, and so on. The GHS scheduler on the mobile device first finds out available resources and partitions tasks of a job into three groups. The first task group is assigned on the internal multicore scheduler, and the second is assigned on the sensor network scheduler. The last group is assigned on the remote cloud scheduler. The last group may have the largest set of tasks, therefore the cloud GHS scheduler decides to further distribute workload among nearby compute resources such as a laptop. This mechanism can be formally described as in Algorithm 1.

Algorithm 1 GHS Algorithm

```
Input: a task graph $G_t = (V_t, E_t)$ // node: task, edge: precedence relation
1: GHS scheduler on a mobile device builds a resource graph $G_r = (V_r, E_r)$ after detecting available resources in the neighborhood. // node: available resource, edge: network connectivity between two resources
2: GHS scheduler partitions resources and tasks into same number of groups.
3: GHS scheduler sends a list of resources and tasks to a group leader for each group. // Group leader is elected based on computing power and network connectivity.
4: GHS schedulers on group leaders do the above steps recursively until partitioning is expected to perform worse.
```

If some resource fail, rescheduling happens at the lowest node which covers those failed nodes.

III. Analysis

In this section, we analyze the performance of GHS compared to a centralized algorithm. The main goal here is not to develop a detailed and exact model of the system and analyze this, but to provide a coarse understanding of the performance of GHS. The random variables used for performance analysis is summarized in Table 2.

The basic assumptions about GHS algorithm for analysis are: 1) GHS algorithm continues to distribute tasks whenever possible, 2) the number of tasks in a job is always greater than the number of available resources, and 3) the time complexities of scheduling and partitioning are all polynomial time, and given by αn^{β} and γn , respectively, where n is the number of tasks.

Random Var./Dist.	Description	Mean
N_r/exp.	Number of available resources	R
N_t/exp.	Number of tasks in a job	T
T_d/exp.	Data transfer time	D

Fig. 2. A performance comparison between a centralized algorithm and GHS algorithm

Regarding scheduling time, it is straightforward that the time complexity is αn^{β} since a mobile phone should scheduling all tasks alone. In case of GHS, the time complexity can be formulated as a recursive form, $T(n) = 2T(\frac{n}{2}) + (\gamma n + 2D)$. If the number of available resources are greater than or equal to the number of tasks, the time complexity becomes $O(n \log n)$ by applying master's theorem. However, since R is assumed to be less than T and the depth of recursive tree is $\log_2 R$, the time complexity is mathematically expressed as $\sum_{i=1}^{\log_2(R+2)-1}(\gamma n + 2^i D) + \alpha(\frac{n}{2^{log_2(R+2)-1}})^{\beta}$. As for rescheduling time, if one node fails at one time and each node has same failure probability, the time complexity of GHS rescheduling is expressed as $\frac{1}{R}(g(n) + 2g(\frac{n}{2}) + \cdots + \log_2(R+2)g(\frac{n}{log_2(R+2)}))$ where $g(n)$ is the time complexity of GHS scheduling.

	Centralized	GHS
Scheduling time	αn^{β}	$\approx (\log_2 R)rn + 2RD + \alpha(\frac{n}{R})^{\beta}$
Re-scheduling time	αn^{β}	$\approx \alpha \cdot \frac{n^{\beta}}{R^{\beta+1}}$

Fig. 3. A performance comparison between a centralized algorithm and GHS algorithm

IV. Conclusions

We propose a distributed scheduling algorithm, graph-structured hierarchical scheduling, which distributes not only task workload but scheduling workload so that responsive online computation offloading is actionable in mobile cloud computing. The theoretical performance analysis shows that the proposed algorithm outperforms a centralized single scheduler. As future work, we will conduct research on more sophisticated partitioning algorithms for graph-structured hierarchical scheduling.

References

[1] Kyung-Ah Chang and IL-Pyung Park. Challenges and enabling technologies in mobile cloud computing. *Multimedia Communication Technical Committee (MMTC) E-Letter*, Vol. 6, No. 10, pp. 24-26, 2011.

[2] Byung-Gon Chun, Sunghwan Ihm, Petros Maniatis, Mayur Naik, and Ashwin Patti. Clonecloud: elastic execution between mobile device and cloud. In *Proceedings of the sixth conference on Computer systems*, EuroSys '11, pages 301–314, New York, NY, USA, 2011. ACM.

[3] Eun-Sung Jung, Sanjay Ranka, and Sartaj Sahni. Workflow scheduling in e-Science networks. In *2011 IEEE Symposium on Computers and Communications (ISCC)*, pages 432–437. IEEE, July 2011.

[4] Jia Yu and Rajkumar Buyya. A taxonomy of workflow management systems for grid computing. *Journal of Grid Computing*, 3(3):171–200, 2005.

An ICA-Based Automatic Eye Blink Artifact Eliminator for Real-Time Multi-Channel EEG Applications

Jui-Chieh Liao and Wai-Chi Fang, *Fellow*, IEEE
Department of Electronics Engineering and Institute of Electronics, National Chiao Tung University, R.O.C.

***Abstract*--This paper presents an ICA-based automatic eye blink artifact eliminator for real-time multi-channel EEG applications. Since EEG signals are very feeble, they are easy to be contaminated by artifacts. Among all artifacts, eye blink artifact dose the most significant harm. ICA has been shown to separate the artifacts from brain activity sources. After the processing of ICA, the results can be used for further applications such as brain computer interface (BCI). Because the behavior of eye blink artifact might mimic the brain activities which might mislead the operation of BCI, the independent component contained eye blink artifact needs to be removed before further applications. For the availability and feasibility of BCI, a real time ICA algorithm, online recursive ICA (ORICA), is developed. With ORICA, a set of ICA result of each EEG sample time can be accomplished after each EEG acquisition. That will reduce the reaction time of ICA and make the applications of BCI more feasible. In order to take advantage of ORICA completely, a real-time eye blink artifact eliminator which can detect the existence of eye blink artifact for every single set of ICA result is needed. The proposed eliminator is designed using TSMC 90nm CMOS technology. The validity of the proposed eliminator is also given in this paper.**

I. INTRODUCTION

In this paper, a VLSI design of automatic eye blink artifact eliminator for real-time ICA-based multi-channel EEG applications is presented. The eliminator is used to remove the data of independent component with eye blink artifact before further applications after the processing of ICA. Since EEG signals are very feeble, they are easy to be contaminated by artifacts. Among all artifacts, eye blink artifact dose the most significant harm to the EEG signals. ICA has been used to extract the eye blink artifact from the EEG signals caused by sources of brain activities. After processing of ICA, the results can be used for further applications such as BCIs [1]. Since the artifact cause by eye movements might mimic the behavior of brain activities and is in a frequency range of 0.5 to 3 Hz (within the delta waves range) [2], it must be removed to avoid misleading the operation of BCIs. Previously, well-trained observers were employed to identify the eye blink artifact visually after ICA processing. Although this method allows the component containing eye blink artifact to be precisely identified, it is not convenient and not feasible for BCI users need to be well-trained to remove the eye blink artifact by themselves. In [3], an ICA-based automatic eye blink artifact correction method is presented and 5000 sets of ICA results are needed to accomplish the eye blink artifact correction automatically. In [4], an online recursive ICA is presented. This type of ICA algorithm produces a result from each independent component at each sample time instead of a segment of EEG raw data. This reduces the reaction time of ICA and makes BCI applications based on ICA more feasible. But this is useless if the eye blink artifact eliminator cannot remove the eye blink artifact in real time. Therefore, an artifact eliminator which can detect the existence of the eye blink artifact for every single result of ICA is required. Due to the above reasons, we present a VLSI design of automatic eye blink artifact eliminator for real-time ICA-based multi-channel EEG applications. The remainder of the paper is organized as follows. In section II, the algorithm adopted for the eliminator is described. In section III, a series of MATLAB® simulations are described to verify the validity of algorithm. In section IV, the design of proposed hardware architecture using TSMC 90nm CMOS technology is provided. Finally, the result and conclusion are given in section V and section VI.

II. ALGORITHM

In this work, a statistic algorithm, sample entropy, is adopted. Sample entropy is an algorithm used to measure the regularity of data sets in a time interval and is proposed by Richman [5]. After the computation of ICA, the algorithm is applied to evaluate whether each independent component contains the eye blink artifact. Since the eye blink artifact appears suddenly and dramatically changes results of ICA, the data sets containing the eye blink artifact should have lower sample entropy values for lower data regularity than other brain activity sources. By taking the advantage of this characteristic, the independent component with eye blink artifact can be distinguished. The computation flow of the algorithm is described as below.

Define Data Vectors:

The results of each independent component are defined as data vectors $d(1)$, $d(2)$, …, and $d(N)$.

Define Data-Set Vector:

The data-set vector is shown as (1) with $i = 1\sim N\text{-}m+1$, and m is the embedded dimension.

$$D(i)= [d(i),\ d(i+1),\ \ldots,\ d(i+m\text{-}1)] \qquad (1)$$

Define the Distance:

The distance between two data-set vectors is defined using (2).

$$d[D(i),D(j)] = \max[|\,d(i+k)-d(j+k)\,|], k = 0\sim m-1 \quad (2)$$

Set Threshold r:

Threshold $r = 0.2\times \text{SD}$, where SD is the standard deviation of the data segment in the time interval. $B^{m}_{i}(r)$ is then calculated, with i = 1~N-m and $i \neq j$, as shown in (3).

$$B_i^m = \{number\ of\ \ d[D(i),D(j)] < r\}/(N-m-1) \qquad (3)$$

Define $B^m(r)$:

$$B^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} B_i^m(r) \qquad (4)$$

Find Sample Entropy Value, SamEN(m,r):

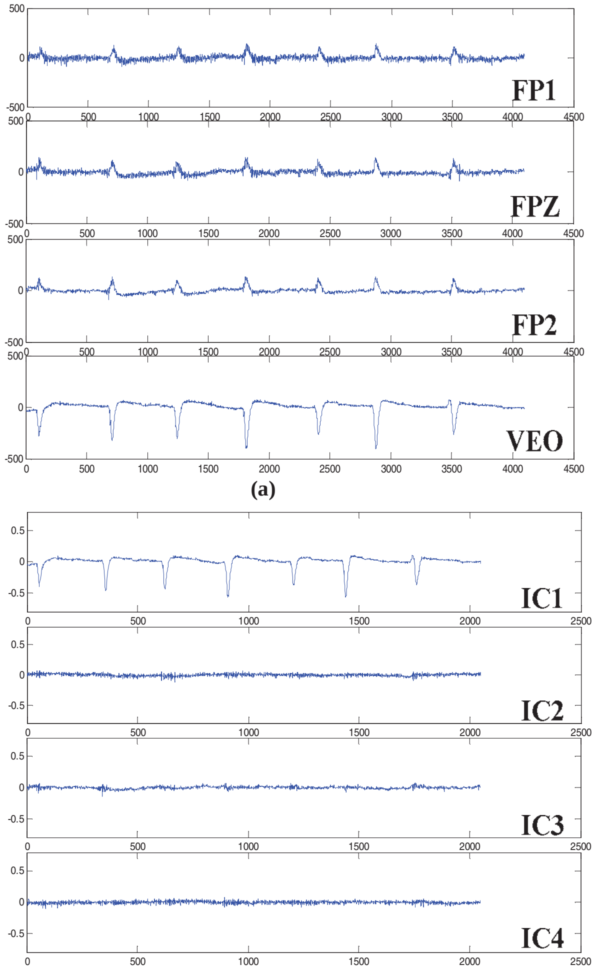
$$SampEn(m,r) = \ln[B^m(r)] - \ln[B^{m+1}(r)] \qquad (5)$$

Before applying the algorithm, there are some parameters and value need to be defined. The first is data amount, N. The second is embedded dimension, m. Finally, a threshold of sample entropy value used to evaluate whether the independent component contains eye blink artifact is required. All of the parameters and value will be defined in III, and a series of MATLAB® simulations will be given to find the threshold of sample entropy value and to verify the validity of the algorithm.

III. Algorithm Verification

In this section, a series of MATLAB® simulation results are given to verify the validity of the algorithm mentioned in section II. The simulations can be separated into two steps: the step to find threshold of sample entropy values and the step to verify supposes in the first step. To execute simulations of the first step, m is set to 2 as [3]. In [3], the data amount, N, is set to 5000. However, a huge amount of hardware resources will be consumed since large size of memory is required to save the 5000 sets of ICA results of each independent component. Here N is set to 128.

The ICA results used for simulation are from the toolbox, EEGLab, and the process data rate is 128 per second. The EEG raw data is acquired from the commercial system, NeuroScan, with the band pass frequency ranged from 0.15 to 100 Hz and the sample rate being 256. The electrodes are located at FP1, FPZ, FP2, and VEO. In order to fit the data process rate of EEGLab, the raw data fed into EEGLab are obtained by averaging each two sample times. The EEG raw data acquired form NeuroScan are shown in Fig. 1a, and the ICA results of EEGLab are shown in Fig. 1b. After EEGLab completes the process of ICA, 2048 sets of ICA results can be acquired. Then the data sets are divided into 16 segments for performing the simulations with N=128. The simulation results are provided in table I. Through observation in Fig. 1b, it can be found that IC1 (independent component) is the component which contains eye blink artifact. From table I, sample entropy values of IC 2, IC 3, and IC 4 are always higher than 1.3. However, the sample entropy values of IC 1 are ups and downs irregularly. This makes the definition of threshold of sample entropy value hard. After we inspect the data segments with the sample entropy value higher than 0.8, we find that they are either not contained eye blink peaks or contained less than half of an eye blink peak like segment 6 and segment 7 shown in Fig 2.



(a)

(b)

Fig. 1. (a) EEG raw data from NeuroScan and (b) ICA results of EEGLab.

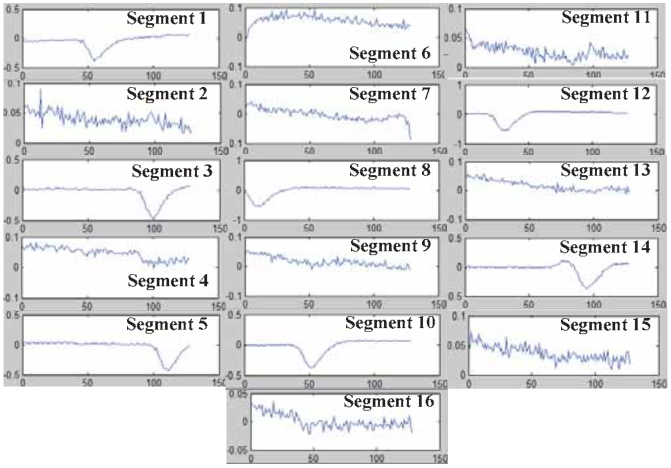


Fig. 2. Data segments of IC 1.

TABLE I
Sample Entropy Value of IC1, IC2, IC3, and IC4

	IC 1	IC 2	IC 3	IC 4
segment 1	0.2343	1.8971	1.9062	1.7750
segment 2	1.9459	2.4849	1.9966	2.4277
segment 3	0.0848	2.1302	1.8028	2.3812

	IC 1	IC 2	IC 3	IC 4
segment 4	1.3823	2.1893	2.0949	2.1879
segment 5	0.1750	1.8795	2.6692	2.0431
segment 6	1.9459	1.9828	1.8718	2.1172
segment 7	1.4083	2.4423	2.3461	2.3593
segment 8	0.0826	2.0971	2.5649	2.2561
Segment 9	1.8357	2.4248	2.2513	2.2687
segment 10	0.1214	1.7918	1.6843	2.3609
segment 11	1.7011	3.0082	2.2914	2.0794
segment 12	0.1111	2.5974	1.8412	1.8608
segment 13	1.3912	2.9042	1.8383	2.8234
segment 14	0.1267	2.0244	1.3587	2.0075
segment 15	1.3573	2.3026	2.4108	2.4681
segment 16	1.7463	2.3026	2.0883	2.6391

Here we define the threshold of sample entropy value as 0.8 and perform another simulation with different ICA results of different raw data. The data segments of the component which contains the eye blink artifact are evaluated as the data segments without eye blink artifact when the data segments do not contain eye blink peaks or contain less than half of an eye blink peak. The results show that the misjudgment occurs when the data segments contain less than half of an eye blink peak. After the MATLAB® simulations, the defect of the algorithm is found. In the next section, the architecture we proposed to overcome this defect is presented.

IV. Design of The Proposed Eliminator

In this section, the design of proposed eliminator for real-time ICA-based multi-channel EEG applications is presented. To achieve real-time elimination, the eye blink artifact detection of each single ICA result is required. In addition, the algorithm defect must be revised. The operation is transformed as follows. As shown in Fig. 3, the data segment collector is used to collect 128 sets of ICA outputs of each sample time. Once the first 128 sets of ICA outs are collected as Segment 1st in Fig. 3, an operation is performed to acquire the first sample entropy value. After this operation, the collector discards the first data set and puts the latest ICA output into the collector like Segment 2nd in Fig. 3. Then the second sample entropy value is generated. The same operations are repeated after each ICA out is generated like Segment 3rd, Segment 4th, Segment 5th, and Segment 6th. This way, a sample entropy value can be acquired after each ICA out. To reduce the reaction time, it is the best situation if the data sets in the segment collector can be used to evaluate the state of the latest ICA out. However, this will lead to misjudgment of the eliminator like the segment 7 in Fig. 2. Although the sample entropy value is higher than 0.8, the last data set which should be indicated as eye blink artifact component is located at the beginning of the eye blink peak. Since the average speed of a human eye blink is 300-400 milliseconds, 30 data sets are sufficient to describe half of an eye blink. In order to avoid the situation like Segment 7 in Fig. 2, each sample entropy value is used to define the state of the 96th data set in the segment collector. Although the reaction time of the eliminator will be longer, the defect of the algorithm can be fixed.

The proposed hardware architecture of the eliminator is provided in Fig. 4. The ICA out cache is acted as the real-time updated data segment collector. In order to be compatible with eight-channel elimination, the ICA out cache is expanded to 1024 words. The blue area is used to compute the threshold r. The five red registers are used to execute the formula (3). The two gray registers are given to execute the formula (4). Then, the sample entropy value is obtained after the computation of nature log as (5). Finally, the output from 96th data set of each independent component in the data segment collector is set to zero to remove the eye blink artifact if the sample entropy value is lower than 0.8. Otherwise, it is delivered to the output directly without elimination. The same steps are repeated until the operations of each component are all finished. Here we suppose that 1/(N-m-1), 1/(N-m-2), and 1/(N-m) are all identical, therefore no additional divisors are required. This might lead to a little error, but the accuracy is good enough for eye blink artifact detection. The proposed hardware architecture is implemented using Verilog hardware description language and the silicon layout is also implemented using SOC encounter with TSMC 90nm CMOS technology as shown in Fig. 5. The specification of the eliminator is summarized in Table II. The operation results will be discussed in the next section.

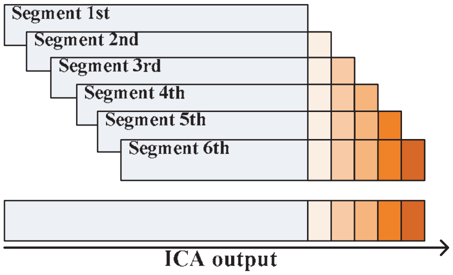


Fig. 3. Real-time updated data segment collector.

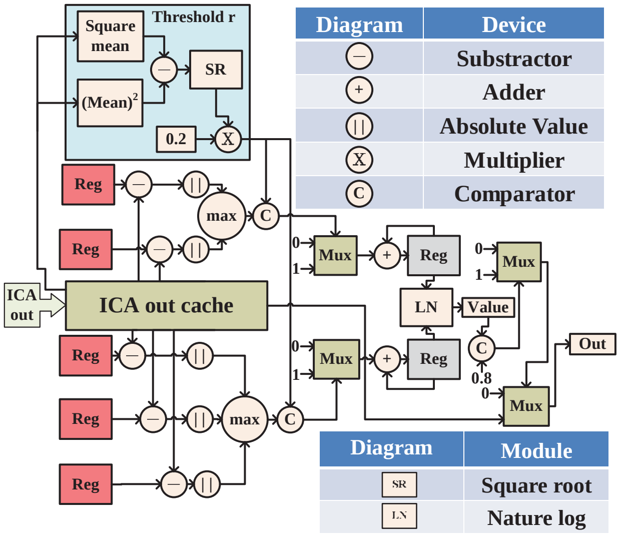


Fig. 4. The proposed hardware architecture.

V. Result

In this section, the operation results of the proposed eliminator are presented. The results of IC2, IC3, and IC4 are shown in table III. The start-up preserved data sets are preserved since the sample entropy values are not defined from the first set to the ninety-fifth set of ICA out. In addition, all sample entropy values of the data sets in these independent components are higher than 0.8. It can be demonstrated that the proposed eliminator will not affect the independent component with brain activity in Table III. Fig. 6 shows the operation results of IC 1 in Fig. 1. As shown in Fig. 6b, the data sets located at the eye blink peaks are all set to zero after the detection. Fig. 6c shows the FFT power spectrum of IC 1 from 0-10 Hz which is almost distributed in 0.5 to 3 Hz (within the delta waves range) before the operation of elimination. In Fig. 6d, the FFT power spectrum after eye blink artifact elimination is shown. The power distribution from 0.5 to 3 Hz is suppressed successfully. After verifications, the proposed eliminator is proven to exactly remove the eye blink artifact of ICA results automatically in real-time.

VI. Conclusion

In this paper, a VLSI design of automatic eye blink artifact eliminator for real-time ICA-based multi-channel EEG applications has been presented. The proposed architecture has also been shown to remove the eye blink artifact exactly. With the proposed architecture, the feasibility and convenience of BCI applications can be improved greatly.

Acknowledgment

This work was supported in part by the National Science Council of Taiwan, R.O.C., under grant NSC101-2220-E-009-049 and NSC101-2221-E-009-169-MY2. The authors would also like to express their sincere appreciation to the National Chip Implementation Center for chip fabrication and testing service.

Reference

[1] Palumbo, A.; Calabrese, B.; Cocorullo, G.; Lanuzza, M.; Veltri, P.; Vizza, P.; Gambardella, A.; Sturniolo, M.; , "A novel ICA-based hardware system for reconfigurable and portable BCI," Medical Measurements and Applications, 2009. MeMeA 2009. IEEE International Workshop on , vol., no., pp.95-98, 29-30 May 2009.

[2] Mammone, N.; La Foresta, F.; Morabito, F.C.; , "EEG Eye-Blinking Artefacts Power Spectrum Analysis," International Conference on Computer Systems and Technologies - *CompSysTech'06*, IIIA.3-1 to IIIA.3-5.

[3] Dan-hua Zhu; Ji-jun Tong; Yu-quan Chen; , "An ICA-based method for automatic eye blink artifact correction in multi-channel EEG," Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on , vol., no., pp.338-341, 30-31 May 2008.

[4] Muhammad Tahir AKHTAR, Tzyy-Ping Jung, Scott Makeigy, and Gert Cauwenberghs, "Recursive Independent Component Analysis for online Blind Source Separation" 2012 IEEE International Symposium on Circuits and Systems (ISCAS2012), COEX, Seoul, Korea, May 20-23, 2012.

[5] J.S. Richman, J. R. Moorman, "Physiological time-series analysis using approximate and sample entropy, " Am J Physiol Heart Circ Physiol, vol 278, pp. 2039-2049, 2000.

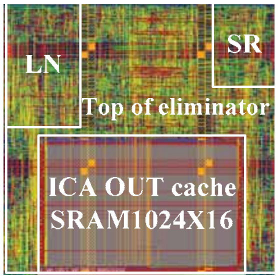
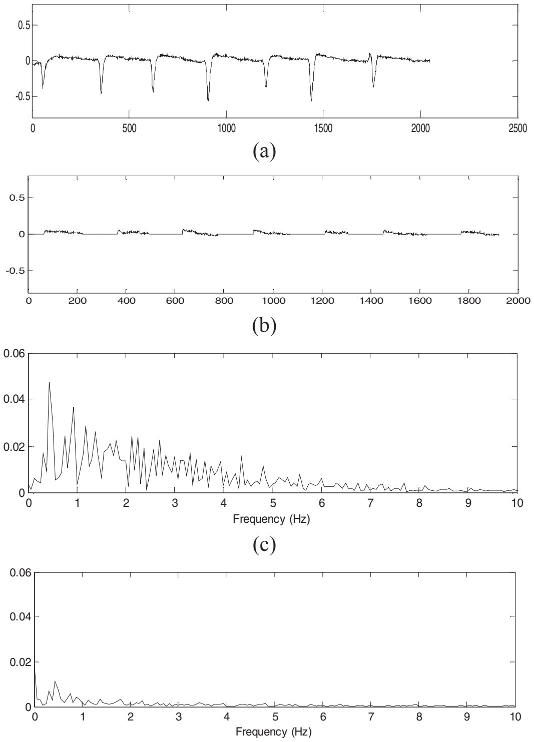


Fig. 5. Silicon layout of the proposed eliminator.



(a)

(b)

(c)

(d)

Fig. 4. (a)ICA results with eye blink artifact, (b) Operation results after elimination, (c) FFT power spectrum before operation, and (d) FFT power spectrum after operation.

TABLE II
Specification of The Proposed Eliminator

Parameter	Value
Technology	TSMC 90 nm
Core Voltage	1.0 V
Core Size	320x320 um2
Operation Frequency	10 MHz
Power Consumption	0.138 mW

TABLE III
Specification of The Proposed Eliminator

IC	Data Amount	> 0.8	<0.8	Start-up Preserved data sets
IC 2	2048	1953	0	95
IC 3	2048	1953	0	95
IC 4	2048	1953	0	95

A Novel Wearable Vibro-tactile Haptic Device

Andrei Ninu[a], Strahinja Dosen[c], Dario Farina[c], Frank Rattay[b], Hans Dietl[a]

[a] Otto Bock Healthcare Products GmbH, Vienna - Austria
[b] Institut for Analysis and Scientific Computing, TU Vienna - Austria
[c] Department of Neurorehabilitation Engineering, University Medical Center Göttingen - Germany

***Abstract* – This paper describes a wearable haptic technology able to produce complex haptic sensations through vibro-mechanical stimulation of the human skin. The technology can be implemented in cost effective devices, can be embedded in small geometries, and is able to provide elaborated stimulation patterns by independently modulating the amplitude and frequency of the mechanical vibro-stimulation. In order to evaluate its capacity to encode information haptically, we tested the device under stress conditions demonstrating the proof-of-principle, showing its strength and limitation. To better describe its capacity to generate vibro-stimuli, a direct comparison with a regular vibration motor has been done.**

I. Introduction

Haptic is a field which increasingly captures the attention of the scientific community and industry. Over time, a wide variety of wearable haptic devices have been proposed: electro-cutaneous stimulators, linear tactors, vibration motors with eccentric mass, kinesthetic displays (e.g., rolling, sliding), piezoelectric devices [1–3] etc. However, the only stimulation technology widely adopted by the industry is the vibration motor with eccentric mass. Its success is due to its low cost, small geometry and high efficiency. The most important disadvantage of this technology is that the stimulation parameters (vibration amplitude and frequency) cannot be independently modulated, and the range of available stimulation patterns is therefore quite limited. In this paper, we propose a novel methodology to generate more elaborated vibro-tactile stimulation by allowing independent modulation of the stimulation amplitude and frequency. We present the first experimental evaluation of the proposed methodology, demonstrating the proof-of-principle.

II. Methods

A. Operating principle

The principle of functioning is based on the Newton's third law of motion which states that for every action there is always an equal and opposite reaction. In our case the "action" is generated by the rotor of a rotational electro-mechanical machine. The rotor is accelerated by the magnetic field of the motor coils, and the resulting inertial force ("reaction") drives the stator. The stator is the stimulating part of the device, preloaded against the skin and embedded into a socket or case by using bearings with a small coefficient of friction. The inertial force transmitted by the stator is perceived as mechanical stimulation. If a square wave voltage is applied to the motor, an alternating, rotational electromagnetic torque is generated by the stator coils. The torque brings the rotor into back and forth motion which consequently coerces the stator to move, generating mechanical vibrations.

B. Implementation

This approach can be implemented by using any rotational electrical drive, but for reliability reasons we have chosen a brushless DC motor (BLDC). The BLDC motors use electronic instead of mechanical commutation. This property guarantees long life of the motor in dynamic applications, which is the case when the motor has to continuously accelerate and decelerate to generate vibrations. In the current implementation, a MAXON BLDC motor with Hall sensors and flat construction (EC32) has been chosen. With a diameter of 32 mm, supplied with 12V and having 15W electrical power, the motor provides a good compromise between efficiency and geometry.

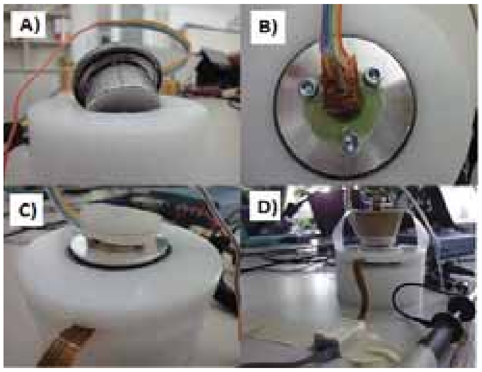


Fig. 1: Experimental setup: A) The stator of the motor was connected to the grounded plastic case via bearings; B)The sensor placement on the stator; C) Silicon layer simulating the skin; D)The stator is preloaded with 200 g.

The stator is constrained to circularly move around the rotational axis of the rotor and consequently stimulates the skin by stretching it tangentially, which has been demonstrated to be more efficient than normal indentation of the skin [4].

The stimulation range of the proposed device was tuned to generate vibrations in the range of 10-250 Hz; in this way the skin mechano-receptors sensitive to vibrations will be optimally activated: the Merkel disk (5-15Hz) sensitive to extremely low frequencies, Meissner's corpuscles (20-50 Hz) sensitive to midrange stimuli and Pacinian corpuscles (60-400Hz) with the lowest detection threshold at 250Hz, sensitive to high frequency vibrations [5].

C. Experimental evaluation

In order to systematically evaluate an implementation of the proposed method to generate vibrations, we conducted a set of tests using an "artificial" setup emulating to some extent realistic conditions. We preloaded the stator, the stimulation part of the device, with a 200g mass through a silicon layer

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

(see Fig. 1D). The stimulation device is a two-input, two-output system having as **I**nputs the reference for **V**ibration **A**mplitude (IVA) and **F**requency (IVF) and outputs the measured frequency and amplitude of the vibration. In the first group of measurements, the IVA was kept constant at 20, 40, 60 and 80% of its maximum, and for each of these levels the IVF was changed from 0 to 100% in 26 equidistant steps. In the second group, the IVF was kept constant at 20, 40, 60, and 80% while IVA was changed from 0 to 100% in 26 equidistant steps. The generated vibrations were measured by using a three axes accelerometer (MMA7368L). The sensor was placed on the stator with the two sensitive axes in the horizontal plane. The signals were acquired at 2 kHz. Ten measurements were performed for each condition. The vibration amplitude was assessed using a peak to peak value of the measured acceleration. In order to emphasize the differences between the novel and classical methods to generate vibrations, we have conducted the same test (preload, silicone base, accelerometer) using a conventional coin type 12mm vibration motor with an eccentric mass. We changed the motor voltage, which is the only degree of freedom, in 26 equidistant steps and we recorded the vibration frequency and amplitude.

III. Results

Fig. 2 depicts the summary results (average values, standard deviation less than 5%). The colored circles are recordings from the “constant amplitude” and the black crosses from the “constant frequency” measurements. The profiles described by the colored circles show the nonlinear correspondence between the IVA, which was kept constant, and the actually measured vibration amplitude over the entire frequency range (10-250Hz). This behavior is due to a resonant effect caused by the elasticity of the silicon layer simulating the natural skin. As visible from these plots, the natural frequency of the silicon layer is approximately 50Hz: the amplitude increases with frequency up to this point and afterward it decreases. The measured vibration was not linearly related to the IVA: the shift between values at lower vibration amplitudes was higher than the shift at higher amplitudes. Moreover, the shift at higher frequencies was smaller than that at lower frequencies. In the second group of measurements, the measured frequency well followed the IVF when the IVA was changed in 26 equidistant steps.

Finally, in the case of a coin type motor, the vibration frequency varied linearly with the applied voltage, while the vibration amplitude was a quadratic function of it. It is well known that the centripetal force which generates vibrations is proportional to the square of the motor’s angular velocity which is proportional to the applied motor voltage.

IV. Discussion and Conclusions

We have presented a novel methodology to generate more elaborated vibro-tactile stimulation. The stimulation achieved using the proposed methodology, plotted in amplitude-frequency domain, covers a 2D space (circles and crosses); the frequency and amplitude are independently adjustable. The stimulation generated by the coin type vibration motors in the same representation is only a quadratic curve (black bullets); the frequency is tightly correlated with the amplitude.

Nevertheless, there is a compromise which has to be met with regard to the device volume and energy consumption. The vibration motor with the eccentric mass is the most efficient, requiring also very small volume. The novel methodology can be also implemented using very small motors, but this would limit the stimulation range. In actual applications, there is tradeoff between the geometry and the stimulation range; the larger the motor, the larger the stimulation range. However, the potential of this methodology is not entirely exploited by the current implementation. The reduction of the stimulation range observed when increasing IVA and IVF is caused by the current approach of motor voltage control. Controlling the motor in “current mode”, the stimulation range can be extended, although this would also increase the system complexity.

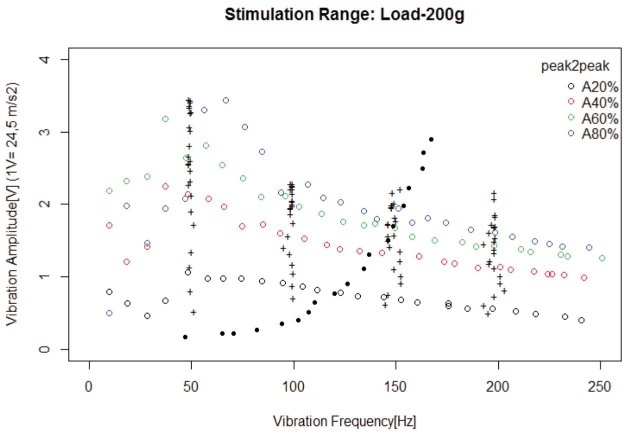


Fig. 2: Summary results: a) colored circles – recordings by varying the IVF in 26 steps while IVA was kept constant at 20,40,60 80%; b) black crosses - recordings by varying the IVA in 26 steps while IVF was kept constant at 20,40,60 80%; c) black bullets – recordings by varying the voltage of coin type vibration motor between 0-100% (0-3V).

V. References

[1] K. A. Kaczmarek, J. G. Webster, P. Bach-y-Rita, and W. J. Tompkins, “Electrotactile and vibrotactile displays for sensory substitution systems,” IEEE Transactions on Bio-Medical Engineering, vol. 38, no. 1, pp. 1-16, Jan. 1991.

[2] D.-S. K. Seung-Chan Kim, “Haptic Interfaces For Mobile Devices: A Survey Of The State Of The Art,” Recent Patents on Computer Science, vol. 1, pp. 84-92.

[3] L. A. Jones and N. B. Sarter, “Tactile displays: guidance for their design and application,” Human Factors, vol. 50, no. 1, pp. 90-111, Feb. 2008.

[4] J. Biggs and M. A. Srinivasan, “Tangential versus Normal Displacements of Skin: Relative Effectiveness for Producing Tactile Sensations.,” in Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002, pp. 121-128.

[5] E.Kandel et altri, "Principles of Neural Science", pp. 437, Mcgraw-Hill Publ.Comp, 2000

Decoding Scheme of Error Correction using Fake Error Addition to Compress Data Transmission

Agi Prasetiadi, Dong-Sung Kim, *Member, IEEE*, Soo-Young Shin, *Member, IEEE*

***Abstract*—This paper proposes a novel method to enhance channel efficiency by exploiting the error feature of a transmission channel using low-density parity-check (LDPC) code. The code is reduced by moving several parity to message data as fake errors. Simulation results show that the trimmed code can be recovered until seven percent trimming for 1/2 rate LDPC code.**

***Index Terms*—Decoding Scheme; Error Correction; Reduction Method; Networked Consumer Electronics**

I. Introduction

Error or noise in a signal is usually thought of as undesirable factor that needs to be eliminated. Therefore error correction code is used to guarantee data realibility. In this paper, we propose to increase the channel performance by using a fake error method in low-density parity-check (LDPC) code [1][2].

The proposed method is based on trimming the parity code and redistributing the trimmed parity code over certain sections in the message code. The message code which is replaced by trimmed parity code experiences intentional error. This error is neccessary to optimize the massage space and at the same time also to exploit the correcting ability of the code. This approach results in a shorter code length which can be used to boost up data transmission such file transfer or video streaming, especially on superphone technology.

The proposed method is raised because the unequal error protection feature is applicable on the transmission [3] or the error correction ability sometimes is too big compared to the actual error that may be occurred. In other words, the channel bandwidth is reduced in order to accommodate the correction capability; thus, the code is inefficient in low-error channels. Although this condition provides another opportunity for further exploitation of the features of the LDPC code by cutting the code directly, the proposed method is also suitable for another error correction scheme.

II. System Modeling

If the data transmission is overprotected in a certain channel, then the rest of the channel resources can be used to exploit the ability of the data to be compressed. The channel in which the error rate is greater than the recovery ability of the code is not suitable for this method.

In this paper, we examine the reduction possibility of the output of the encoded 1/2 rates LDPC data, particularly the parity allocation, during transmission such that the overall bandwidth can be reduced. If some sections of the parity code are trimmed and distributed homogeneously over the actual data, some parts of the data are replaced by this parity. This action can be called as intentional error. Consequently, the code length is smaller than before. The proposed method is similar with watermarking approach [4][5]. The following equation shows how the particular parity code distributed on message code.

$$\bar{c} = \{\bar{m}, \bar{p}\} \rightarrow \hat{c} = \{\bar{m} - \bar{m}_{(a_1:a_l)} + \bar{p}_{(1:l)}, \bar{p}_{(l+1:s)}\}, \qquad (1)$$

where $\bar{m}$ is message, $\bar{p}$ is parity, l is the length of trimmed parity code, s is the length of all parity code, and a_i is the position of replaced message code. Based on this, after $\hat{c}$ is being transmitted to the receiver, the code should be rearrange its position as $\bar{c}$ format.

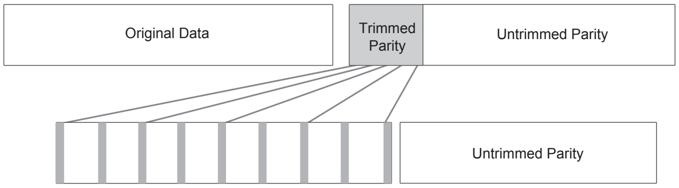


Figure 1. Schematic of Distributed Parity on Original Data as Fake Error

Figure 1 ilustrates the idea of certain sections of the message code being replaced by sections of the parity code. The top panel shows the original code that consists of two parts, the message data and the parity data. The distribution is carried out by moving and replacing the message sections of the data with the trimmed parity code.

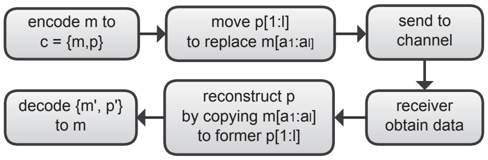

encode m to c = {m,p}
move p[1:l] to replace m[a_1:a_l]
send to channel
receiver obtain data
reconstruct p by copying m[a_1:a_l] to former p[1:l]
decode {m', p'} to m

Figure 2. Process Flow of Trimmed Code Transmission

Figure 2 shows process flow of how trimmed code involved on LDPC encode-decode process. The message data is encoded as LDPC code. Then the trimming process as described in Figure 1 is applied on the encoded code. At this state, the trimmed code is sent to the channel. After the data is received by the receiver, the trimmed parity bits are restored to their original positions in the parity block.

A. Prasetiadi, D.S. Kim, S.Y. Shin are with the Department of IT Convergence, Kumoh National Institute of Technology, South Korea (e-mail: {agiprasetiadi, dskim, wdragon}@kumoh.ac.kr)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. Simulation

The simulation is performed by using MATLAB. A 1152-bit LDPC code with a 1/2 parity rate is generated randomly. The capability of the code in terms of error correction is observed for several additive white Gaussian noise (AWGN) levels.

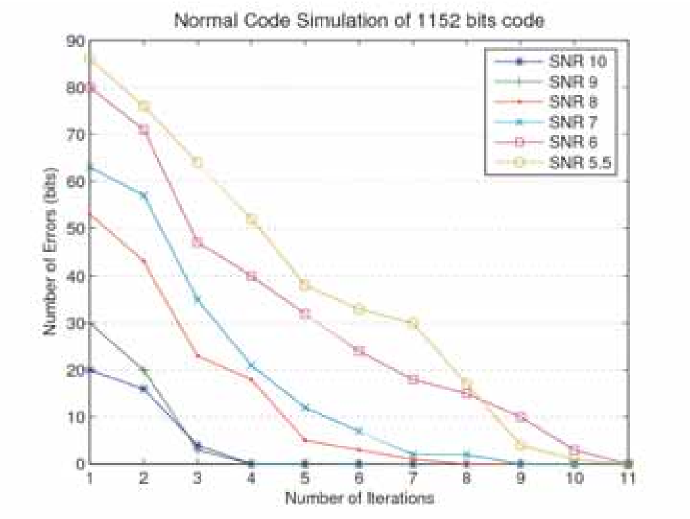


Figure 3. Normal Code Simulation

Figure 3 shows the normal LDPC code before the trimmed process is applied on the code. The code is applied with various degree of errors, where in this figure the error level is denoted as SNR. High SNR means the error occurred on the code is low, and vice versa. The x-absis of the figure shows the number of iterations experienced by the LDPC code during decoding process. The y-absis shows the remaining bit errors on the code after several iterations. It can be seen that if the code experienced high SNR, then the total number of iterations needed to recover the data is low. But when the code experienced low SNR, the total number of iterations needed to recover the data becomes high.

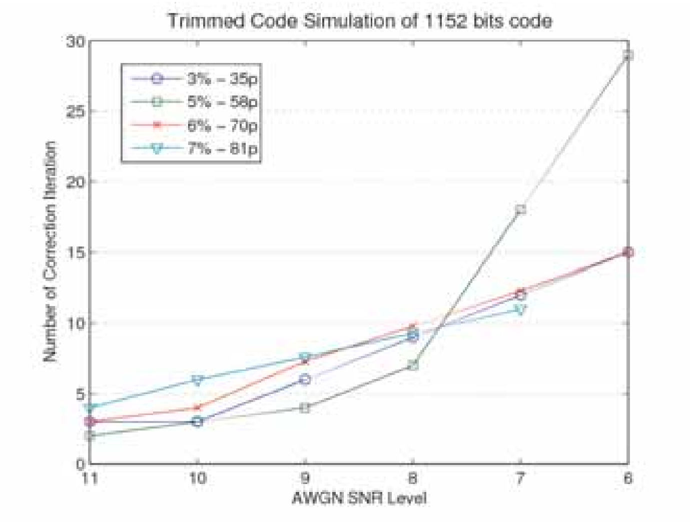


Figure 4. Trimmed Code Simulation

Figure 4 shows the decoding performance for several trimmed code variations, where various lengths of the trimmed parity were applied to the LDPC code. For example, when the code is subjected to 3% of the trimmed parity, then 35 parity bits are moved to replace 35 message bits, with the result that the code length is 35 bits shorter than the normal LDPC code.

The simulation shows that at 3% trimmed code, the total number of iterations needed to recover the code at an SNR level of 10 is three. The trimmed LDPC code with an SNR level of 9 requires six iterations as compared to three iterations for the normal LDPC code. The trimmed LDPC code with SNR level 6 requires 15 iterations as compared to 10 iterations for the normal LDPC code.

The simulation shows that the code reaches the limit of its ability for the 7% trimmed code condition. This is because when the SNR level is 6, the code cannot be corrected, while in the normal case, the code can still be corrected over 10 or 11 iterations. Moreover, based on the simulation results, if the code is subjected to 10% trimmed parity, the code simply cannot be recovered at all.

Intuitively, although the code length reduces, if the errors further increase, the decoding process becomes even more complex than before. Although this method is suitable for uplink process where energy consumption is not a major issue for decoding, but the possibilities to apply this method to downlink channel in superphone device is also promising.

IV. Conclusion

This paper proposed a method to enhance channel capacity by using a trimming code on the 1/2 rates LDPC code. The idea is to distribute some of the parity code within the data code. The proposed method increases the errors in the code by means of the addition of pseudo errors. The trimmed LDPC code can reduce the length of the transmitted data up to 6% reduction. However, the complexity of decoding increases at the receiving side. This method is suitable for superphone device to overcome increasing iteration process. Thus, the proposed method has promising usage to compress data transmission by reducing file transfer or video streaming directly at binary level.

Acknowledgments

This work was supported by National Research Foundation of Korea (NRF) under grant No. 2012-0008130.

References

[1] Robert G. Gallager, *Low-Density Parity-Check Codes*, Cambridge, MIT, June 1963.

[2] Davey, M.C. and MacKay, D.J.C., ”Low density parity check codes over GF(q),” *Information Theory Workshop*, Jun 1998, pp. 70-71.

[3] Sandberg, S. and Von Deetzen, N., ”Design of bandwidth-efficient unequal error protection LDPC codes,” *IEEE Transactions on Communications*, March 2010, Vol. 58, pp. 802-811.

[4] Nezhadarya, E., Wang, Z.J., and Ward, R.K., ”Robust Image Watermarking Based on Multiscale Gradient Direction Quantization,” *Information Forensics and Security*, Dec 2011, Vol. 6, pp. 1200-1213.

[5] Chuntao Wang, Jiangqun Ni, and Jiwu Huang, ”An Informed Watermarking Scheme Using Hidden Markov Model in the Wavelet Domain,” *Information Forensics and Security*, June 2012, Vol. 7, pp. 853-867.

Fast Operating System Switcher for Mobile CE Devices

Chei-Yol Kim, Soo-Cheol Oh, KangHo Kim, Chang-Won Ahn, Young-Kyun Kim
Electronics and Telecommunications Research Institute, South Korea

***Abstract*—This paper proposes a fast Operating System (OS) Switcher for ARM-based Consumer Electronics (CE) such as Smart Phones, Tablets and IPTVs. The OS switcher enables users to utilize multiple OSes on single H/W device without security problem from interference among OSes. The OS switcher is notably faster than the multi-boot approach and can preserve the previous working states. The OS switcher does not burden any run time overhead compared to virtualization and can be easily applied to the new product. We implemented the OS switcher in the bootloader layer without any modification of OS.**

I. INTRODUCTION

Most easy and traditional way to using multiple OSes on single device is multi-boot approach. The multi-booting chooses which OS will be booted when a system starts. The multi-boot approach does not generate any run-time overhead, but the system has to be rebooted to switch another OS. The other shortage of the multi-booting is that it cannot preserve the previous OS working states.

The other technology for using multiple OSes is virtualization. In the server and desktop area, the virtualization becomes most widely used technology. By rapid enhancement of mobile devices, the virtualization can be adopted to mobile CE devices. ViMo[1], VLX[2] and XenArm[3] are the results of the virtualization for mobile and embedded systems. The virtualization basically needs high performance hardware resource and makes performance degradation compared to native single OS. In addition to these points, the virtualization is usually difficult to implement and hard to port new devices.

Another approach of using multiple OSes is an OS switch technology [4]. We present the OS switch approach and implementation on ARM Cortex-A8 based mobile CE device. Android is the most popular and widely used mobile OS, so we implemented the two Android OS switching. By providing two Android OSes on single device, we can prohibit that each OS is affected by another. One can be used as personal with freely installing applications and modification and the other can be used as official purposed device without installing and modification.

In this paper, we first tell about OS switch features and then show our OS switcher design. Next, the implementation issues will be discussed and finally we will conclude the proposed OS switch.

II. FEATURES OF OS SWITCH

Compared with the multi-boot, time consumed to switch OSes is very shorter. Usually more than one minute is needed to halt the first OS and start the second OS in the multi-boot. The proposed OS switch just needs several seconds to switch from one to another OS. In addition to the switching time, the OS switch could preserve the previous OS system states, but the multi-boot mechanism loses the previous system states when it is rebooted.

The OS switch does not burden any run time overhead and has extremely less complexity against the virtualization. This means that the OS switch has high portability. Additionally, the OS switch could be implemented without OS modification. This point is very important in the market.

Table 1 shows the features of each approach.

Table 1. Comparison of each approach

	Multi boot	OS Switch	Virtualization
Switching time	> 1 minute	< 5 seconds	< 0.001 sec
Run time overhead	X	X	O
Preserving working states	X	O	O

III. OS SWITCHER DESIGN

The OS switcher consists of STR(Suspend To Ram), resume and switch mechanism for the decision whether boot or OS switching. In this section we will explain each procedure in detail.

A. Suspend-To-RAM

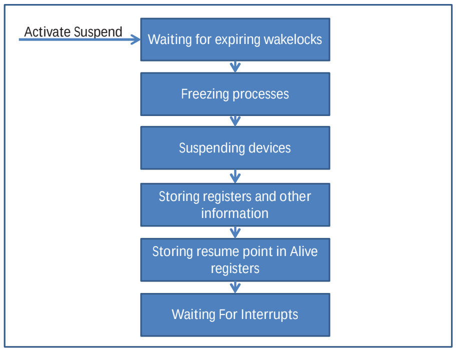

Activate Suspend
Waiting for expiring wakelocks
Freezing processes
Suspending devices
Storing registers and other information
Storing resume point in Alive registers
Waiting For Interrupts

Figure 1. STR procedure of Android

Fig. 1 illustrates the process of STR of Android. When a suspend signal is activated, the system states including CPU, memory and IO devices are frozen and stored in RAM. The stored point is written in the non-volatile CPU register, named *Alive register*. And then the system waits for the interrupt for wake up. The first block of fig. 1 about *wakelocks* will be specifically discussed at implementation detail section.

B. Resume

Resume procedure is waking the suspended OS to activate. Fig. 2 shows the Android resume process. When a reset signal is triggered, bootloader checks weather it is the boot or resume process. The bootloader checks the reset type and if it is set as resume, and then jumps to the resume pointer stored in *Alive registers*. Alive registers are located in the CPU. After that, resume process restores the CPU registers and activates the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

devices.

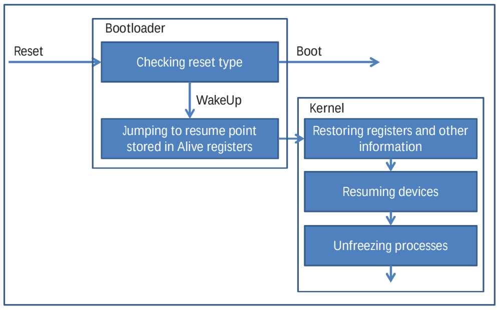

Bootloader
Reset
Checking reset type
Boot
WakeUp
Kernel
Jumping to resume point stored in Alive registers
Restoring registers and other information
Resuming devices
Unfreezing processes

Figure 2. Resume procedure of Android

C. OS Switching

OS switch is located in the bootloader layer. It has to know whether it is the first OS boot or second OS boot or switching. When it is switching, switcher has to decide which OS has to be resumed. Fig. 3 shows the sequence of switching when a reset signal is triggered.

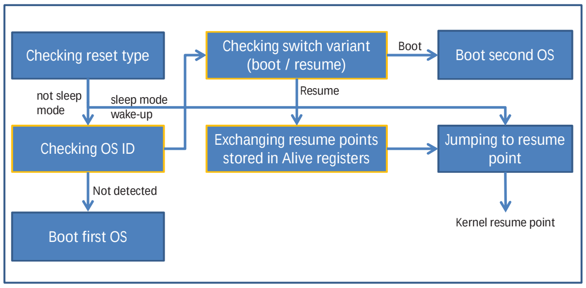


Figure 3. OS switch procedure

When system is activated, the bootloader checks the reset type. The reset type tells whether this activation is from general sleep mode or normal boot. When first booting, this is not the sleep mode wake-up state, then it checks the OS ID. If OS ID is not set, this means the first booting. After first boot, first OS will be STR and second reset signal will be triggered. In this second time, it will find the OS ID and determine whether it is switching or second boot. When switch variant is set as a boot, the second OS will be booted. The third case will have different switch variant value, and then the bootloader will run the switching process. The last case will be happen when the OS wants to sleep by itself. Then the system just wakes the OS not switching.

IV. IMPLEMENTATION DETAIL

The prototype platform board is ODROID [5] which has a Cortex-A8 based S5PC100 ARM processor, 512MB RAM, 10GB flash memory and 3.5-inch LCD with touch screen. Experimented OS is Android 1.5 which has 2.6.27 Linux kernel.

We used the non-volatile memory based register set which reside in ARM CPU core, named information register (INFORM0~INFORM7) for saving switching information. Also we divided 512MB RAM into two regions for each OS. Each OS is not able to know other OS's memory by configuring their physical memory. This makes each OS to be secure from other OS.

Android has a lock for waiting interrupt before entering the sleep mode for fast returning to the active state named *wakelocks*. *Wakelocks'* default waiting time is several seconds. This makes our OS switch to be slow than our expectation. We hoped not to make any modification of OS. But we had to change this value for fast switching.

V. CONCLUSION

In this paper, we introduced the OS switcher for two Android OSes on ARM processor. The OS switching technique has some advantages over the multi-boot and the virtualization. The OS switch is faster than the multi-boot and can preserve OS working status. The virtualization cannot avoid additional burden of system emulation for virtualizing CPU, memory and the other devices. The OS switch does not have any burden for the switching. In addition to this point, the OS switching is more portable and easier to adapting new hardware devices than the virtualization.

The disadvantage of the OS switching is to split memory for each OS. In some cases, this point cannot be ignored. But on enough memory environments, this will not be a problem.

Mobile OS like Android is going to be adapted to general CE devices beyond mobile devices. The OS switch can be an alternative technology of the multi-boot or the virtualization.

The OS switcher can be used for any special purpose devices such as e-book reader or smart TV for supporting general purpose OS without long switching time. These special purpose devices usually make use of OS only for their special jobs preventing interference from other application and operations. In this case, OS switcher can support general purpose OS with their special purpose OS without any effect.

ACKNOWLEDGEMENT

This work was supported by the IT R&D program of MKE/KEIT. [10038768, The Development of Supercomputing System for the Genome Analysis]

REFERENCES

[1] Soo-Cheol Oh, KangHo Kim, KwangWon Koh, and Chang-Won Ahn, " *ViMo (Virtualization for Mobile) : A Virtual Machine Monitor Supporting Full Virtualization For ARM Mobile Systems" Cloud Computing 2010.* Lisbon, pp. 48-53, Nov. 2010.

[2] VirtualLogix VLX. Avaiable http://www.virtuallogix.com

[3] Sang-Bum Suh, Sung-Kwan Heo, Chan-Ju Park, Jae-Min Ryu, Seong-Yeol Park, Chul-Ryun Kim, "*Xen on ARM: System Virtualization Using Xen Hypervisor for ARM-Based Secure Mobile Phones*" Consumer Communications and Networking Conference, 2008, pp. 257-261.

[4] Jun Sun, Dong Zhou, Steve Longerbeam, "Supporting Multiple OSes with OS Switching" in 2007 USENIX Annual Technical Conference.

[5] ODROID. Available http://hardkernel.com/renewal_2011/products/prdt_info.php?g_code=G129689092760

A Scalable Scheduling Algorithm for Coarse-Grained Reconfigurable Architecture

Hae-woo Park, Wonsub Kim, Donghoon Yoo, Soojung Ryu, Jeongwook Kim
Samsung Advanced Institute of Technology

***Abstract*—Coarse-grained reconfigurable architectures (CGRA's) are introduced as flexible architectures that can efficiently execute various types of applications in a single device. A CGRA often achieve high IPC by utilizing tens or hundreds of functional units (FU's). The key technique in exploiting a CGRA is to find an optimal mapping of operations over FU's. Modulo scheduling algorithm is known as the state-of-art technique to find fairly efficient solution; however it often takes too much time and occasionally fails as the number of FU is increasing. In this paper, we propose a novel two-stage scheduling algorithm which finds out a solution within a reasonable amount of time. The experimental result presents the proposed algorithm reduces the scheduling time by 92% and finds out schedules that are as efficient as the solutions given by the previous modulo scheduler.**

I. INTRODUCTION

Coarse-grained reconfigurable architectures (CGRA's) are introduced as flexible architectures that can efficiently execute various types of applications in a single device. These architectures often contain tens or hundreds of functional units (FU's), which can handle *word-level* operations. The major benefit of CGRA is that it can be configured every cycle at run-time, so that high-level programming languages such as C can be adopted to program it with aid of compiler techniques. This programmability greatly increases the productivity in developing. Figure 1 illustrates an example of a CGRA that contains 16 FU's, register file(s), and a control memory.

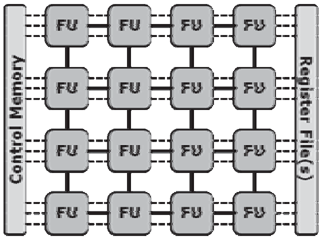


Figure 1. An example of a CGRA, containing 16 FU's

To maximize the performance, the compiler tries to find instruction level parallelism (ILP) in the given application and schedules the instructions exploiting the ILP. However, due to that ILP is constrained by data dependences between operations, it is impossible to get high degree of ILP through the entire application. Hence the previous researches [1][2] often focus on loop level parallelism (LLP) to find out large number of independent instructions through the loop iterations.

Modulo scheduling [3] is known as the state-of-art scheduling algorithm for CGRA. It improves the degree of ILP and LLP by overlapping the execution of consecutive iterations of a loop at intervals of *initiation interval* (II). While its various implementations [4][5] give quite efficient scheduling results, they have some problems: (1) they consume too much time and (2) they are not scalable, i.e. the scheduling time and the schedule quality get exponentially worse with increasing of the number of FU's.

This paper proposes a novel scheduling algorithm that notably reduces the scheduling time and finds out a schedule that is at least as efficient as the solutions given by previous solvers. The key idea is to assume the CGRA architecture as a composition of homogeneous FU clusters and schedule the instructions in two stages which are (1) *local scheduling* within each FU cluster and (2) *global scheduling* in between FU clusters. This algorithm enables fast and fairly efficient instruction scheduling.

The remainder of this paper is organized as follows. Section II presents the previous approaches. Section III describes the proposed architecture and algorithm. Section IV gives the experimental results. Finally, section V concludes this paper.

II. RELATED WORKS

DRESC framework [4] suggested a modulo scheduling algorithm that is based on simulated-annealing (SA). [6] pointed out that the reason why the searching space is so large is that it searches for valid routing paths after placing operation; hence proposed an algorithm, called *edge-centric* one, which first tries to find routing candidates before place operations. This hugely reduces the searching space; however, it causes performance degradation. The following work [7] enhanced the schedule quality by handling the recurrences in a special way. Another research [8] proposed an algorithm that considers the resource usage during operation placement so as to find out a solution within a reasonable amount of time. However, they all are not sufficiently fast and scalable.

Note that all the previous approaches attempted to find the solution with the whole loop code on the whole CGRA. The proposed approach is distinguished from them in that it firstly finds out a local solution on a portion of CGRA and then obtains a global solution by sewing the local ones.

III. PROPOSED SCHEDULING ALGORITHM

A. Analysis: why does it take so much time ?

Simulated-annealing (SA) –based approach [4] requires long time to find an efficient solution, since SA and similar meta-heuristics depend on repairing function which is not easy to design in this problem. Most of the other algorithms are based on branch-and-bound heuristics, which try to place instructions and their dependence edges over FU's and channels one by one. They often succeed to advance in early placement; however, easily fail in late placement, where there are few resources which are sparsely distributed. Note that the scheduler cannot predict what resources are needed for late instructions; hence it cannot reserve any rooms for them. To make matters worse, the remaining resources are often apart from each other; hence the routing between them becomes a serious issue, for the later instructions are deeply related among themselves as the early ones are.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Architectural Assumptions

We assume that we can logically group FU's into FU clusters as illustrated in Figure 2, where an FU cluster contains heterogeneous FU's but FU clusters are homogeneous. An FU cluster is expected to be capable of the full set of operations. The channels are categorized into two groups: local and global channels, which are intra- and inter-cluster ones, respectively. This architectural assumption is not a serious restriction, since such a modular design is popular and many CGRA's allocate each operation over FU's evenly for uniform use of FU's.

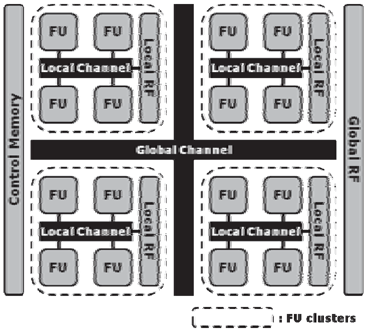


Figure 2. Abstract form of the proposed architecture

C. Two-stage Scheduling Algorithm

The proposed two-stage algorithm schedules the instructions in the given loop as follows:

① Loop unrolling/coalescing: this step is for raising the ILP in case that single loop iteration does not have enough ILP.

② Local scheduling for an FU cluster: single loop iteration is scheduled over FU's in an FU cluster. Figure 3 shows that a loop iteration denoted as A-B-C-D is executed on a single FU cluster. When there is a recurrence, the scheduler tries to schedule the producer as earlier as possible and the consumer as later as possible. This technique prevents the loop skew from being too large.

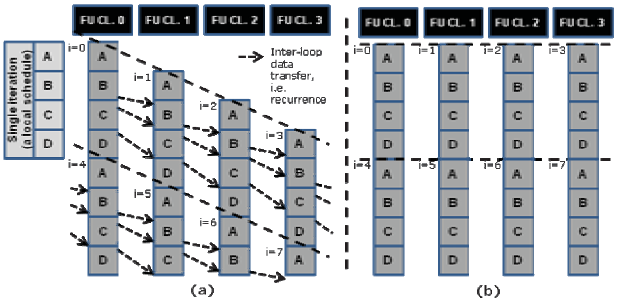

FU CL. 0
FU CL. 1
FU CL. 2
FU CL. 3
Single iteration (a local schedule)
A
B
C
D
i=0
i=1
i=2
i=3
i=4
i=5
i=6
i=7
Inter-loop data transfer, i.e. recurrence
(a)
(b)

Figure 3. The proposed global scheduling examples: (a) if there are recurrences, (b) if there is no recurrence.

③ Global scheduling of local schedules over FU clusters: making the loop skew between local schedules depending on inter-loop data transfer, i.e. recurrence. In case there are recurrences, the execution of a loop iteration has to be delayed for the data transfer as illustrated as Figure 3 (a). The loop skew is calculated as *max(L/c, max(d_k))*, where L is the local schedule length, c is the number of FU clusters, and d_k is the distance of producer-consumer pair of a recurrence r_k. In contrast, if there is no recurrence then no loop skew is needed, so more efficient scheduling is possible as shown in Figure 3 (b).

In this algorithm, the scheduling is never failed and we can achieve the IPC in the stable state as near c·P, where the IPC in the local schedule is P. This means the scheduling algorithm is scalable. Though this algorithm may give a worse solution if any of d_k is large, this problem should occur in previous algorithms.

IV. EXPERIMENTAL RESULTS

We examined the proposed algorithm with a CGRA simulator, which models 16 FU's, full crossbars in FU's within each FU cluster, 2-channel crossbars between FU cluster. For experiments, we tested 119 loops that are extracted from H.264/AVC video decoder. We compared the proposed scheduler with one of the previous modulo-scheduler [7]. For compilation, the scheduler takes 179 seconds total, while the modulo-scheduler takes 2195 seconds. Figure 4 presents the IPC ratio of the proposed scheduler's to the previous modulo-scheduler's, where the geometric mean value is 1.054.

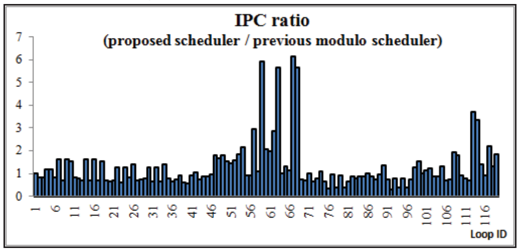


Figure 4. IPC ratio of the examined loops

The result presents that the compilation time is reduced by 92% and the IPC is slightly increased.

V. CONCLUSION

This paper proposes a two-stage scheduler for CGRA. By dividing the scheduling problem into local and global scheduling, it can make an efficient solution in reasonable time. The experimental result shows that the proposed algorithm greatly reduces the scheduling time and finds out schedules that are at least as efficient as the solutions given by the previous modulo scheduler. Since we use a simple list scheduler for the local scheduler at this time, we may increase the performance by developing CGRA-specific optimizers in near future.

REFERENCES

[1] G. Lu, H. Singh, *et al.*, The morphosys parallel reconfigurable system. Euro-Par'99, pp. 727-734, 1999.

[2] C. Ebeling, *et al.*, Mapping applications to the rapid configurable architecture. FCCM'97, 1997.

[3] B. R. Rau, Iterative modulo scheduling: an algorithm for software pipelining loops. MICRO 27, pp. 63-74, 1994.

[4] B. Mei, *et al.*, Exploiting loop-level parallelism on coarse-grained reconfigurable architectures using modulo scheduling. DATE'03, 2003.

[5] H. Park, *et al.*, Modulo graph embedding: mapping applications onto coarse-grained reconfigurable architectures. CASES'06, 2006.

[6] H. Park, *et al.*, Edge-centric modulo scheduling for coarse-grained reconfigurable architectures. PACT'08, pp. 166-176, 2008.

[7] T. Oh, *et al.*, Recurrence Cycle Aware Modulo Scheduling for Coarse-Grained Reconfigurable Architectures, LCTES'09, pp. 21-30, 2009.

[8] A. Hatanaka and N. Bagherzadeh, A modulo scheduling algorithm for a coarse-grain reconfigurable array template. IPDPS'07, pp. 1-8, 2007.

Device-Level Voltage Control Scheme of MLC NAND Flash Memory for Storage Power Failure Recovery

Sanghyuk Jung, *Student Member, IEEE*, and Yong Ho Song, *Member, IEEE*
Hanyang University, Seoul, Korea

***Abstract*—MLC NAND flash memory has been widely used as a storage device in mobile and desktop computing systems. However, MLC NAND flash memory may cause a data loss problem because the LSB-page programmed data can be lost when a power failure occurs in the middle of a MSB-page program operation. In this paper, we propose a device-level voltage control scheme in order to overcome this problem. With the theoretical feasibility of our proposed scheme, the storage controller could fully restore the LSB-page programmed data at device-level after a power failure.**

I. Introduction

The rapid development of NAND flash technology has enabled to provide light-weight, small size, and low power consumption. Due to these advantages, NAND flash has been widely used as a storage device in mobile embedded systems and desktop computers. Although NAND flash has several restrictions such as an *"erase-before-write"* characteristic and a limited lifespan resulting from its device architecture, many solutions including software supports (i.e., *flash translation layer* [1-4], *bad block management*, etc.) have been developed for hiding these drawbacks.

Recently, in order to produce NAND flash with lower price, the flash vendors are trying to squeeze more capacity into adopting *multi-level cell (MLC)* technology [4-5]. A 2-bit MLC NAND flash memory is able to store two bits in a cell, and two paired-pages share a word-line by dividing variable charged voltages in the cell. As shown in Figure 1, the *cell voltage distribution (CVD)* forms *dotted-blue line* from the ERASE state when a *least significant bit (LSB)*-page is programmed. After the LSB program, the CVDs form *solid-red lines* from the ERASE state or dotted-blue line when a *most significant bit (MSB)*-page is programmed. In a flash read operation, the cell voltage sensor in a flash device can detect the voltage level by using *RD1*, *RD2*, and *RD3* sensing points.

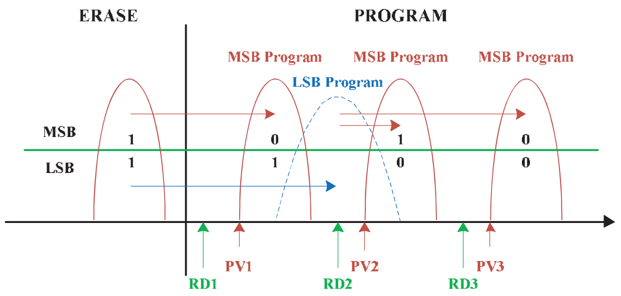


Fig. 1. MLC NAND operation method of generic ISPP technique.

However, MLC NAND flash may cause performance and reliability degradation because of its cell voltage management policy. In order to generate several CVDs, the *incremental step pulse program (ISPP)* [6] policy has been widely used. As shown in the Figure 2, we can make the CVDs much elaborate form if we use lower voltage pulses and larger number of program steps. To generate the CVDs of MSB-pages needs more number of elaborate operation steps than that of LSB-pages. And because the MSB-pages have smaller read sensing margins than LSB-pages, the program latency of MSB-pages is higher than that of LSB-pages. Particularly, the *power failure* problem on MLC NAND flash memory has been mainly issued these days because the LSB-page programmed data can be lost when a power failure occurs in the middle of a MSB-page program operation.

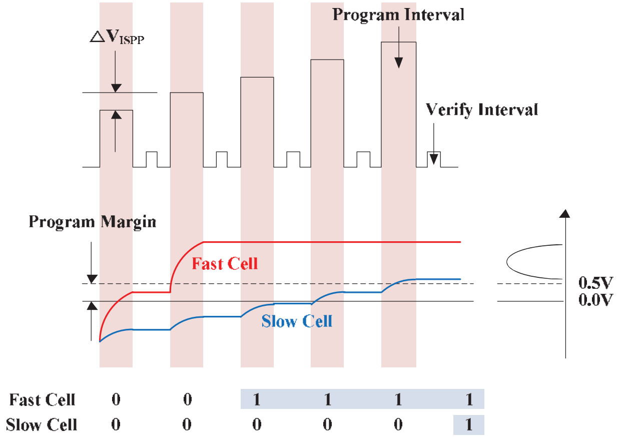


Fig. 2. Incremental step pulse program and voltage verification.

It is possible that the storage controller stores the LSB-page programmed data into a reserved storage area before the flash controller programs the MSB-page and restores the LSB-page in case of a power failure. However, this approach suffers from too much overhead because it generates additional page copy operations at every MSB-page program.

In this paper, we propose a device-level voltage control scheme which handles the ISPP voltage steps and shifts the voltage-level sensing points. With our proposed scheme, the storage controller can fully restore the LSB-page programmed data at device-level after a power failure.

II. Device-Level Voltage Control Scheme

Our approach, the device-level voltage control scheme, is divided into an *ISPP voltage control part (program operation)*

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0005325).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

and a *sensing point shifting part (recovery operation)*.

A. ISPP Voltage Control

As shown in Figure 3, it is possible that the CVD of (0) programmed LSB cell is moved to a higher voltage range by handling ISPP operations without overlapping the CVD of (0, 1) programmed MSB cell. In this operation, we must not decrease the voltage margins between four dotted-red lines. To decrease this voltage margin causes high bit-errors and a lifespan drop of MLC NAND flash memories. In the ISPP voltage control policy, compared with Figure 2, the CVD of the LSB-page has smaller voltage width and higher voltage level by using larger number of ISPP operations. Consequently, it is possible that the flash controller suffers from high LSB-page program delay, but the overhead is trivial because the LSB-page program is 2-3 times faster than the MSB-page program.

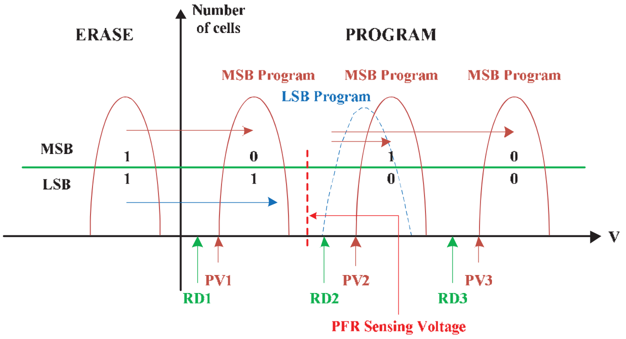


Fig. 3. MLC NAND operation method of ISPP voltage control scheme.

B. Voltage Sensing Point Shifting

The flash controller recovers the LSB-page data by shifting the read sensing point to the PFR sensing voltage after a power failure. As shown in Figure 3, the *dotted-red line* means the PFR sensing voltage in a page. This dotted-red line is located in the middle of voltage margin between CVDs of (0, 1) programmed MSB cell and (0) programmed LSB cell.

In this circumstance, although a power failure occurs in the middle of MSB-page program operations when the LSB cell has 1, the cell voltage sensor can recognize that the LSB cell stores 1 because the MSB-page programmed cell has lower voltage range than the PFR sensing voltage. Otherwise, if a power failure occurs in the middle of MSB-page program operations when the LSB cell has 0, the cell voltage sensor is able to recognize that the LSB cell stores 0 because the MSB-page programmed cell has higher voltage range than the PFR sensing voltage.

III. Feasibility

Park et al. [6] have designed a fast page program NAND flash by using an ISPP voltage control scheme. Since these researches have proven the probability of moving CVDs in NAND flash devices by handling ISPP operations, we can use an ISPP voltage control scheme considering the trade-off relation between the program time and durability ISPP. Moreover, the LSB-page program has smaller number of ISPP operations than that of MSB-page program, so it is possible for the LSB-page program operation to adopt a delayed ISPP voltage control policy without a significant overhead.

Dong et al. [7] have proven the probability that the flash devices are able to change read sensing voltages by shifting the sensing points. Their approaches are to shift the sensing points for adopting a *soft-decision LDPC algorithm* [8] to error correction of NAND flash memories. Unlike the variable sensing points of et al. moreover, since the proposed approach in this paper use a fixed read sensing point for providing MSB-page data loss, we can easily implement our design in NAND flash devices.

IV. Conclusion

In this paper, we discussed LSB-page data loss problems on MLC NAND flash memories when a power failure occurs in the middle of a MSB page program operation. The MSB-page programs were found to require ISPP operations on the LSB-page programmed CVDs, resulting in the flash controller suffering from the indistinguishable CVDs of LSB-page with a power loss. However, the proposed device-level voltage control scheme can eliminate this LSB-page data loss problem by ISPP voltage control and sensing voltage shifting policies. For the future work, we will develop a NAND flash device adopting the proposed ISPP voltage control and sensing voltage shifting policies.

References

[1] S. Jung, J. H. Kim, and Y. H. Song, “Hierarchical architecture of flash-based storage systems for high performance and durability,” *Proceedings of the IEEE/ACM DAC*, July 2009.

[2] A. Gupta, Y. Kim, and B. Urgaonkar, “DFTL: A flash translation layer employing demand-based selective caching of page-level address mappings,” *Proceedings of the ACM ASPLOS*, March 2009.

[3] Y. Lee, S. Jung, and Y. H. Song, “FRA: A flash-aware redundancy array of flash storage devices,” *Proceedings of the IEEE/ACM CODES+ISSS*, October 2009.

[4] S. Jung, S. Lee, H. Jung, and Y. H. Song, “In-page error correction code management for MLC flash storages,” *Proceedings of the IEEE MWSCAS,* August 2011.

[5] C. Lee, S.-K. Lee, S. Ahn, J. Lee, W. Park, Y. Cho, C. Jang, C. Yang, S. Chung, I.-S. Yun, B. Joo, B. Jeong, J. Kim, J. Kwon, H. Jin, Y. Noh, J. Ha, M. Sung, D. Choi, S. Kim, J. Choi, T. Jeon, H. Park, J.-S. Yang, and Y.-H. Koh, “A 32-Gb MLC NAND flash memory with Vth endurance enhancing schemes in 32 nm CMOS,” IEEE Journal of Solid-State Circuits, vol. 46, issue 1, January 2011.

[6] K.-T. Park, M. Kang, D. Kim, S.-W. Hwang, B. Y. Choi, Y.-T. Lee, C. Kim, and K. Kim, “A zeroing cell-to-cell interference page architecture with temporary LSB storing and parallel MSB program scheme for MLC NAND flash memories,” *IEEE Journal of Solid-State Circuits*, vol. 43, issue 4, April 2008.

[7] G. Dong, N. Xie, and T. Zhang, “On the use of soft-decision error-correction codes in NAND flash memory,” *IEEE Transactions on Circuits and Systems*, vol. 58, issue 2, February 2011.

[8] J. Wang, T. Courtade, H. Shankar, R. D Wesel, “Soft information for LDPC decoding in flash: mutual-information optimized quantization,” *Proceedings of the IEEE GLOBECOM*, December 2011.

Interactive Environment Management System Using ZigBee and Self-Configuration Algorithm in Modular Data Center for Effective Power Usage

Taehwan Shin, *Member*, *IEEE*, Insoo Lee, Jinsung Byun and Dukchul Kim

***Abstract*-- This paper proposes an interactive environment management system (IEMS) using low-power communication and self-configuration algorithm. We designed and implemented the IEMS in the test bed. The proposed system reduces the cooling power consumption of the test bed up to 6.6%.**

I. INTRODUCTION

As the digital information explodes, a data center has become important in recent years. At the data center, not only IT equipment such as servers, storages and network equipment but also dual power facilities, cooling devices and air conditioners for reliability are operated 24 hours, of which power consumption is great subsequently. In addition, as cloud services (IaaS, Paas, Saas) are vitalized, importance of the data center is growing bigger [1]. Global power consumption increases due to increase in the data center which has been doubled for every 5 years, and as of 2011, worldwide power consumption is expected to be 100 billion kWh. Particularly cooling cost of the data center accounts for 40% of total power cost [2]. Recently power usage effectiveness (PUE) [3] is applied as the indicator for power usage efficiency of the data center.

$$PUE = \frac{Total\ power\ consumption}{Power\ consumption\ of\ IT\ equipment} \qquad (1)$$

In order to increase PUE, power usage amount other than currently fixed amount of IT power included in total power consumption should be reduced. Ultimate goal of the cooling system can be said based on energy efficiency and availability. In the data center, there are number of cooling systems required per server rack. In other words, as server racks increase, the more number of cooling systems are required.

In this paper, an interactive environment management system (IEMS) was developed to minimize power consumption and operation cost and to enhance energy efficiency and availability for hot/cold aisle structure.

II. INTERACTIVE ENVIRONMENT MANAGEMENT SYSTEM

Fig. 1 shows overall system architecture of the IEMS. Based on low power sensor network, interior/exterior environmental data of the data center is collected, and the management server controls the cooling system, which enables the modular data center to improve the PUE.

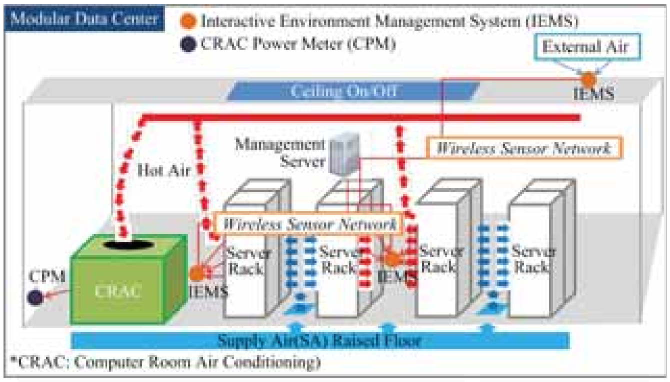


Fig. 1. Overview of the interactive environment management system.

A. Middleware Architecture

Fig. 2 shows middleware architecture of IEMS, which is composed of six components.

1) Information Management Component (IMC): IMC consist of five managers. The computer room air conditioning (CRAC) Power Manager manages power usage used in the CRAC power meter (CPM); the temperature manager (TM) monitors outdoor air temperature and indoor temperature generated in the server, and determines cooling temperature. As part of an effort to reduce greenhouse gas, the carbon manager (CM) measures and controls carbon dioxide in/out of the data center. The humidity manager (HM) detects humidity in/out of the data center and involves in selection of optimized cooling air by interoperation with TM. The sensing value manager (SVM) calculates data for concentrated cooling and outdoor air cooling, and intelligently controls environment by applying self-configuration algorithm. The environment

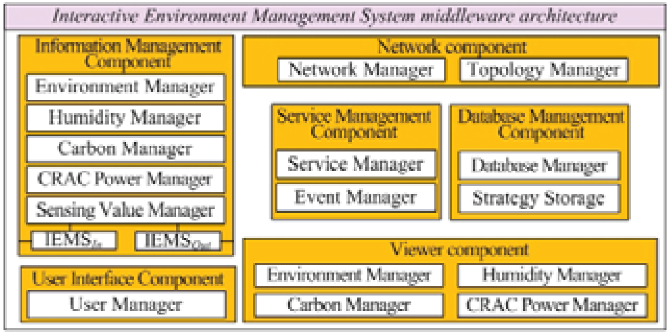


Fig. 2. Middleware architecture of the IEMS.

This work was supported by the Grant of the Korean Ministry of Education, Science and Technology (The Regional Core research Program/Zero Energy Green Village Technology Center) and by the Grant of Next-Generation Internet Data Center(IDC) Test bed based on Green Computing Infrastructures and Greenwares No.10035324

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

manager (EM) controls environment in real time by SVM.

2) Network Component (NC): The network manger (NM) is in charge of managing network elements such as network access, connection, Quality of Service (QoS), and security. It re-configures the sensor network for information measurement by the topology manager (TM).

3) DataBase Management Component (DMC): It collects and refines measured information in the data center, stores it in the database. It derives estimates according to statistical analysis on measured information and situation, and stores them in the strategy storage (SS).

4) Service Management Component (SMC): The event manager (EM) sets optimal service range and manages events. By current status control, the service manager(SM) generates alarm to alert the status of the new event.

5) User Interface Manager (UIM): It provides user interface and security function. If a user intends to change configuration of the IEMS, he/she may access the IEMS after authentication through UIM.

6) Viewer Component (VC): It provides information on power/temperature/humidity/CO2 in/out of the data center processed from IMC, and it allows users to easily access with smart devices.

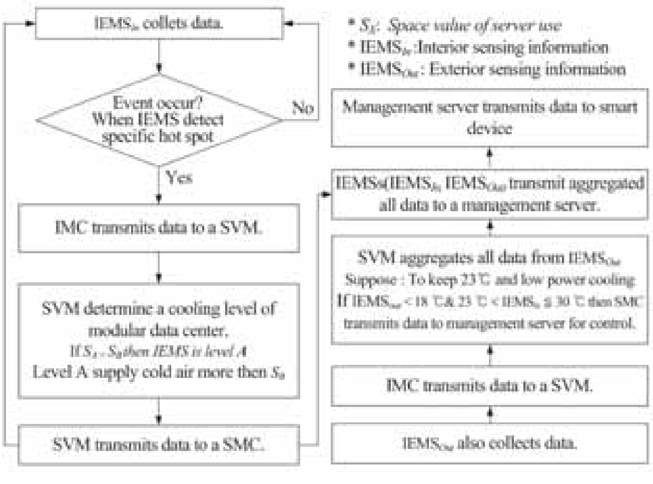
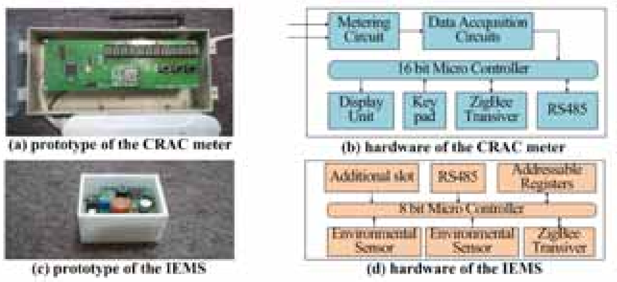

IEMS$_{In}$ collets data.
Event occur? When IEMS detect specific hot spot
No
Yes
IMC transmits data to a SVM.
SVM determine a cooling level of modular data center.
If $S_A > S_B$ then IEMS is level A
Level A supply cold air more then S_B
SVM transmits data to a SMC.
* S_X: Space value of server use
* IEMS$_{In}$:Interior sensing information
* IEMS$_{Out}$: Exterior sensing information
Management server transmits data to smart device
IEMSs(IEMS$_{In}$, IEMS$_{Out}$) transmit aggregated all data to a management server.
SVM aggregates all data from IEMS$_{Out}$
Suppose : To keep 23℃ and low power cooling
If IEMS$_{Out}$ < 18 ℃& 23 ℃ < IEMS$_{In}$ ≦ 30 ℃ then SMC transmits data to management server for control.
IMC transmits data to a SVM.
IEMS$_{Out}$ also collects data.

Fig. 3. Flow chart of the Self-Configuration Algorithm.

B. Self-Configuration Algorithm

Fig 3 illustrates the self-configuration algorithm flow chart. The built-in IEMS collects environment information on server racks of the data center, which is managed by the IMC. If the IEMS detects a certain spot which generates more heat than other spots, SVM generates an event signal. For instance, if server usage of a certain space A is higher than that of space B, it calculates data to supply cold air intensively to space A. In addition, the IEMS installed in external space also collects external data, which is managed by the IMC. The IMC transmits data to the SVM. Provided that the data center should maintain low power cooling at 23 °C, if external temperature is lower than 18 °C and internal temperature is higher than 23 °C and lower than 30 °C, the SVM calculates the value and transmits the data to the SMC. Events are managed by the SMC, and all the information is transmitted to the management server and is checked from smart devices.

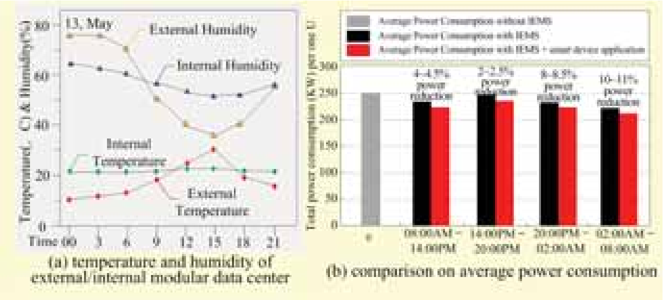


(a) prototype of the CRAC meter
(b) hardware of the CRAC meter
(c) prototype of the IEMS
(d) hardware of the IEMS

Fig. 4. Implementation of the CRAC meter and the IEMS

III. IMPLEMENTATION AND TEST

Fig. 4 shows the prototype and hardware block diagram of the IEMS. The main processor is based on 16-bit microprocessor. A ZigBee transceiver is used for communication with other networked devices because of ZigBee's low-cost and low-power characteristics. 250 kbps/2.4 GHz ZigBee module is used. There are various sensor modules (e.g. temp, humidity, carbon dioxide) on the IEMS. The IEMS operates by the battery. Fig. 5 shows evaluation results. For the test bed data center, dimensions were 42U standard rack. And the rated power per server rack was 10kW. For the test bed, Energy consumption was measured when the proposed system was applied and the system is not applied. As a result, the IEMS reduces energy consumption up to approximately 6.6%.



(a) temperature and humidity of external/internal modular data center
(b) comparison on average power consumption

Fig. 5. Experiment results.

IV. CONCLUSIONS

In this paper, we propose the IEMS using low-power communication and self-configuration algorithm. We designed and implemented the IEMS in the test bed. The proposed system reduces the cooling power consumption of the test bed up to 6.6%.

REFERENCE

[1] N. Leavitt, "Is Cloud Computing Really Ready for Prime Time?," *Computer*, vol. 42, no. 1, pp. 15-20, Jan. 2009.

[2] R. Schmidt, M. Iyengar, "Thermodynamics of information technology data centers," *IBM Journal of Research and Development* , vol. 53, no. 3, pp. 9:1-9:15, May 2009.

[3] E. Jaureguialzo, "PUE: The Green Grid metric for evaluating the energy efficiency in DC (Data Center). Measurement method using the power demand," *The IEEE 33rd International Telecommunications Energy Conference (INTELEC 2011)*, pp. 1-8, Oct. 2011.

In-Home Power Management System Based on WSN

Francisco J. Bellido Outeiriño, *CE Soc. Member, IEEE*, José Flores Arias *CE Soc. Member, IEEE*,
Matías Liñán-Reyes and Emilio Palacios-Garcia

Abstract-- **Smart Grid refers to the next generation power grid in which the power management is upgraded by incorporating advanced communications and capabilities for improved control and efficiency. Among all fields covered by the term Smart Grid energy management in home and building facilities is one of the aspects which is still not really developed. Our main goal is the creation of an application to manage the power consumption at home by controlling household appliances and other loads by means of setting the maximum power suitable and a priority algorithm based on timing schedule, temperature or ambient light.**

Index Terms— **Smart grids, wireless sensor networks, smart metering.**

I. INTRODUCTION

In the actual society, the electricity has become an essential actor in our lives. To plug in a device is so simple and usual that we have forgotten the cost that this energy causes, both from an economic and social point of view.

Nowadays systems are arising for home environment which aim to integrate renewable sources, manage HVACs, lighting control or just implementing simple timers for household appliances or advanced stand-by operation modes [1].

Public consumptions represents about 30% of the global consumption in western countries, and about 18% just for home consumption, a percentage big enough to try to apply the new technologies to control and manage this energy in a more efficient way [2].

New Concepts are arising now into home scenarios. Terms like “Smart Grid”, “Smart Energy” or “Smart Metering”. All of them pursue three main objectives: (i) Efficient distribution of the energy depending on demand, (ii) Real time power monitoring at home and (iii) Devices and methods that allow optimizing the energy demand from the user side.

The impulse given by the newest technologies for communication networks have also become a big step in this field, where the high cost of wired infrastructures plus the limits in the existing ones have been solved by means of the wireless networks like WiFi, Bluetooth or Zigbee[3].

The origin of the Domotics is in the 70's, when the first building automation devices such as X-10 technology, came on the market. From this time there arose a new branch of interest in automation aimed at finding a dream home. Arises then the conception of smart buildings and smart homes that have intelligent control systems that can regulate and adapt their actions more or less autonomously without the necessary intervention of humans.

The main standard for home control systems are BACnet, HBS, Batibus, CEBus, EHS, HES, EIB, Konnex / KNX, LonWorks or other proprietary systems like X10, Hometronic, SIMON VIS StarBox or EMerge. We cannot forget other specific standards for entertainment networks such as HAVi, UPnP, Jini or DLNA.On data networks that support these standards are Ethernet, USB, Firewire, CAN bus or carrier wave based like HomePlug or HomePNA. In wireless key we must mention WiFi, Bluetooth, Zigbee and other minor extended as Z-Wave, Hiperlan or IP500 [1].

This paper focuses on developing an energy & power management system for home enviroments by making use of wireless sensor networks [3] and under the Smart Energy concept. Materials used in the system are described and results about tests and measurements are presented.

II. BACKGROUND

Our main goal is the creation of an application to control the power consumption, which will be supported by Zigbee technology. This network will follow the philosophy of the WSN (Wireless Sensor Network), where a central coordinator is responsible for carrying out the task control algorithm, while the nodes simply provide parameters for the decision and are responsible for acting on the loads [5].

Currently, Zigbee is the leader in monitoring and control products to manage energy and water. The Zigbee Smart Energy profile is one of the main areas in development in recent years for energy efficiency. Furthermore, this profile is complemented with other profiles such as Building Automation, Home Automation or Light Link. As previously mentioned the term Smart Energy involves both home and business users of energy, and for suppliers, so is intended a constant communication by means of which is possible to reduce individual consumption and efficiently manage the connection and production of energy from facilities.

III. MATERIALS AND METHODS

A. Radio Modules

The basic element on our project is a Zigbee module from a well known manufacturer [4]. The architecture, called Z-Accel is based primarily in separating the application implementation itself from the network management. This encapsulates all RF functionality needed in a Zigbee network processor while the microcontroller is running the main application while communicating with the network manager through SPI link.

For programming and debugging the IDE environment IAR Embedded Workbench have been used. We have run the basic license so the code has been optimized and the final

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

application is less than 4KB. As alternative, Code Composer Studio[4] can be used, plus the tool called *Grace* which give a visual environment to manage all the peripherals of the system.

B. Developed system

1) Power sensing

The system is based on the calculus of the instantaneous power through the measurement of the current. Three main alternatives have been considered and the current transformer is the chosen because of its advantages for our purposes. We choose a 20A model, 0-5V output proportional to the I_{rms} flowing. We have designed a conditioning circuit based on the use of a CMOS OpAmp which performs the impedance and voltage adaptation for the ADC, obtaining 6.3MHz as minimum sampling frequency.

2) Temperature and light sensor

Both temperature and light sensors are mounted on board of the RF module, connected through resistors to the internal ADC channels. The datasheets give us the transfer equation between the voltage measured and the physical value of temperature and illuminance.

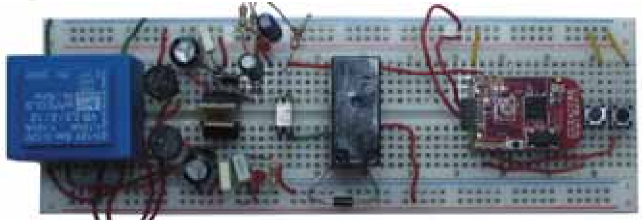


Fig. 1. AC Load drive module with mote and power source included.

3) AC Load drives

To drive AC loads we need a circuit that can manage $230V_{ac}$ and currents up to 12A. We have selected relays because of the simplicity and cost, but it is necessary to provide galvanic isolation among the relay coil and the microcontroller due to the electric noise which appears during commutation. We have chosen a Darlington optocoupler plus a free-wheel diode in the coil to avoid over voltage when current cuts off.

4) Power sources. AC/DC converters.

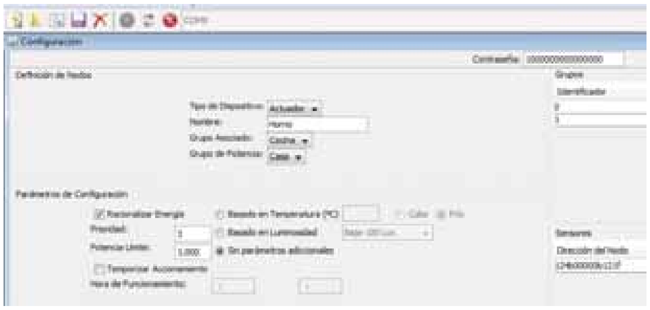
Due to the fact that all measurements are done near mains, from where we can take the energy, we consider unnecessary the use of batteries. Three different power sources have been designed because of the different needs in power and voltage levels for each kind of modules: power sensor modules, RF motes and AC Load drive ones.

IV. Graphic User Interface and Network Manager

The last element in this project is the software developed for the PC that will perform both the management and control system and that will serve as an interface with the user. It has been developed using Java and NetBeans as IDE.

The interface serves two key functions. At its highest layer it serves as a support to manage the network, observing the monitored data (graphically and log files) and system control. In the bottom layer it implements control algorithms for each type of management referred.

In the network management section, users can select the node type (sensor or actuator), assign a password to enable them into the network, assign a label or add into a group that the user can define, allowing for example to group all modules that are in the same room under a common nick.



Configuración

Contraseña: 0000000000000000

Definición de Nodos

Tipo de Dispositivo: Actuador ▾

Nombre: Horno

Grupo Asociado: Cocina ▾

Grupo de Potencia: Casa ▾

Grupos

Identificador

2

1

Parámetros de Configuración

☑ Racionalizar Energía

Prioridad: 1

Potencia Límite: 1.000

☐ Temporizar Accionamiento

Hora de Funcionamiento:

○ Basado en Temperatura (ºC) ○ Calor ● Frío

○ Basado en Luminosidad Bajo 100 Lux ▾

● Sin parámetros adicionales

Sensores

Dirección del Nodo

124b00000b1c11f

Fig. 2. Screenshot of the configuration wizard of the GUI

The control algorithm used for the management of the loads is based on a decision tree on various parameters, forming a structure basically divided into two layers.

In the first layer the load acting is determined on the basis of the current instantaneous power, the set limited value, the additional power that would result from activation of the new load element and its priority over other loads, which even could be disconnected to enable the connection of the above not to exceed the maximum power established threshold.

The second is where advanced control is implemented in addition to power limits. We can add variables such as timing, temperature or luminosity for making control decisions. Control conditions are additive (ORed), giving rise to complex control equations but very simple implementation by the user. There is also a manual control mode for unconditional activation and deactivation of loads.

V. Conclusions

We have presented an application to control the power consumption at home supported by Zigbee technology. The system is based on the calculus of the instantaneous power and it solves if new loads can be activated using priority criteria. A second level implements advanced control by considering variables such as timing, temperature or luminosity for making control decisions. The system has been mounted and tested.

References

[1] Bellido-Outeirino, F.J.; Flores-Arias, J.M. et al , "Building lighting automation through the integration of DALI with wireless sensor networks," *Consumer Electronics, IEEE Transactions on* , vol.58, no.1, pp.47-52, February 2012.

[2] *UIE.* Electricity for more efficiency: electric technologies and their energy savings potential. *EURELECTRIC. http://www.uie.org/*

[3] C. Buratti, A. Conti, D. Dardari, and R. Verdone, "An overview on wireless sensor networks technology and evolution", *Sensors*, vol. 9, pp. 6869-6896, Sep. 2009.

[4] Texas Instruments. http://www.ti.com. October 2012.

[5] M. Aliberti, "Green networking in home and building automation systems through power state switching," *IEEE Trans. Consumer Electron.*, vol. 57, no. 2, pp. 445-452 May 2011.

Proxy Mobile IP based Mobility Support in Heterogeneous Network Environments

Sunghyun Yoon, Noik Park, and Young Boo Kim
Converged Network Research Team, Electronics and Telecommunications Research Institute, Daejeon, Korea

Abstract— **Proxy mobile IP is a protocol that supports the network-based mobility. Since the network provides the mobility on behalf of mobile node, there is no need to mount a mobility protocol in the mobile node. So burden of the mobile node is reduced significantly, however, the proxy mobile IP does not consider mobility in heterogeneous networks. This paper proposes a scheme for proxy mobile IP based mobility in heterogeneous network environments.**

I. Introduction

Since the mobile Internet is positioned as a universal service, the IP mobility has become basic requirement. Thus, a variety of mobility technologies have been proposed. They are classified as host-based and network-based mobility depending on who provides the mobility. Whereas a heavy mobility protocol has to be mounted on a mobile node (MN) in the host-based mobility, the network-based mobility has no requirements except IP for the MN. Since portability is very important for MN, most MNs are implemented in the form of a small handset. Thus, the network-based mobility is currently preferred for this reason.

Proxy mobile IP (PMIP) is a leading network-based mobility protocol [1],[2]. The MN handles the mobility just at the lower layers of the IP and the mobility at the upper layers, including IP, is processed by network. The MN is identified by a mobile node identifier (MN-ID) such as network access identifier (NAI). The network authenticates the MN using a MN-ID of the MN and a pre-defined home network prefix (HNP) is assigned to the authenticated MN. After the acquisition of the HNP, the MN obtains an IP address (i.e. home address). Afterward the network makes the MN is unaware of the change at the network layer. To do this, the PMIP introduces novel network elements; mobile access gateway (MAG) and local mobility anchor (LMA).

The MAG is a kind of access router and the first system which a MN is connected at the network layer. When a MN is connected to the MAG, the MAG establishes a bidirectional tunnel with the LMA and intercepts all packets from the MN. The LMA works as a home agent (HA) of the MN. The LMA maintains binding information between the MN and MAG and forwards the all packets which destination is MN to the MAG that the MN is connected to.

In PMIP, the network supports the IP mobility on behalf of the MN. Thus, if the MN has just IP stack without any separate functions, the mobility can be provided. This reduces the burden of the MN in a number of ways such as performance, power consumption, and wireless resource utilization. However, the PMIP does not consider mobility in heterogeneous networks. Although the PMIP supports the multi-homing through access technology type option, there are limitations to provide the handover between different networks.

Even if the MN has the multiple network interfaces, the PMIP handles just one interface. In other words, a MN with multiple interfaces is recognized and treated as a separate device in each network. This means that, even if each network supports the PMIP, another technology is needed for heterogeneous mobility.

Meanwhile, several studies have been performed in order to support heterogeneous network with PMIP [3],[4]. These are focused on how to take advantage of the LMA such as using a LMA which is commonly used in heterogeneous networks or establishing bidirectional tunnels between the LMAs of each network. This means that the LMA have much of a burden, and cannot but lead to scalability issues eventually. Of course, the PMIP is constantly trying to mitigate the burden of LMA by optimizing the traffic path between LMA and MAG. However the network architecture must be open to each other in order to apply the route optimization in heterogeneous networks. This is not appropriate in terms of network security.

Although the scalability is definitely an issue to support PMIP in heterogeneous network environments, the most fundamental issue is the problem of address change. The PMIP allocates a HNP instead of a home addresses (HoA) to MN and the MN generates its own HoA through the auto-configuration. This means that the MN with multiple network interfaces can have multiple home addresses. Namely, reachability as well as mobility of the MN cannot be guaranteed.

To solve this different HoA problem, the network have to remember the address assigned to the first interface of the MN and re-assign to the second interface when the MN handovers. However, this increases the burden of address management in network and results in a waste of resource and damage of network accessibility of the MN since the only one interface is used at a time.

II. Proposed Architecture and Consideration

This paper proposes new network architecture to support PMIP-based mobility in heterogeneous network environments. Fig. 1 shows the proposed architecture. As shown in Fig. 1, the proposed architecture has a dedicated signalling node (DSN) which is a novel network element to set data path exclusively in control plane. The DSN has dedicated signalling channels with each conventional PMIP network elements as well as MN. The signalling with MN is performed via MAG.

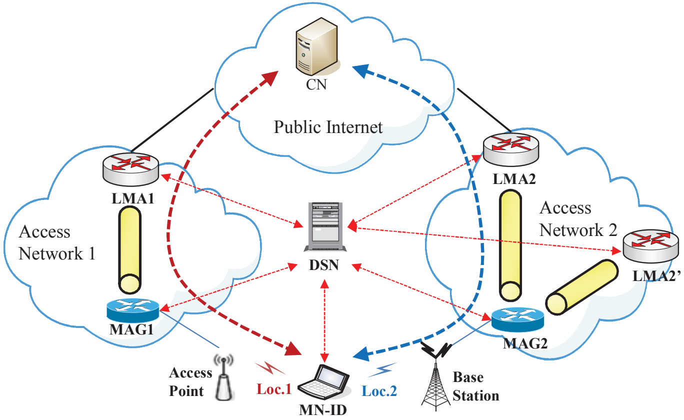


Fig. 1 Proposed architecture for PMIP-based mobility in heterogeneous network

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Whereas data packets are passed through the tunnel between the MAG and LMA, all signal messages are exchanged through the dedicated signalling channel. This means separation of the signal and data path.

The scalability issues mentioned above can be rather solved using the DSN. The LMA can reduce the burden on signalling for data path by focusing on data exchange with the MAG. The MAG is able to respond more flexibly MN's handover by new signalling process between previous MAG and new MAG.

The identifier (ID) and locator separation concept is applied with regard to the MN's address. There are two kinds of address which can be assigned to MN. One is the ID which identifies the MN and the other is the locator which is assigned to each interface indicating the location of the MN. Whereas the ID uniquely identifies the MN, the locator is different according to each interface and changed whenever the point of attachment (PoA) is changed. The ID can be obtained via the MN-ID, and each locator can be obtained by access network respectively.

The MN has a thin client for signalling with the DSN. The thin client manages consistently the network interfaces of the MN and informs the DSN of any change of network connection status in the MN. For this signalling process, the MN is always connected to the DSN through a secure channel. Persistent connection between the DSN and MN can be obtained by simultaneously exploiting multiple network interfaces of the MN. Each network interface of the MN is set to active or standby interface according to the current link quality. The thin client sets interface with the best connectivity to active interface, and the next best one, as standby interface. When the active interface loses the connectivity or the signal strength degrades, and it switches standby interface to the active. If the access network changes, the make-before-break handover is performed. Thus, the connection between the DSN and MN is always maintained unless the MN is out of the scope of its all interfaces.

The MN's ID is used for end-to-end communication and identifying the MN by the DSN. The DSN is in charge of ID/locator mapping and sets optimal data path using the current network connection status of the MN.

III. Performance Evaluation

To test the performance and practical aspects of the proposed scheme, the DSN and thin client software as well as PMIP network elements are implemented and a test bed is constructed.

Fig. 2 shows binding cache entry in the DSN before and after heterogeneous handover. First, each interface mounted on the MN is assigned IP address (i.e. locator) from the access network respectively. If IP communication is available, MN's network information is transferred to the DSN by the thin client. As shown Fig. 2, the binding cache entry has sufficient network connection information of the MN for setting up the data path.

Key	ID	Locator	Status	LMA	MAG	ATT	HNP
1	MN-ID	Loc.1	Active	LMA1	MAG1	WiFi	HNP1
2	MN-ID	Loc.2	Standby	LMA2	MAG2	WiMAX	HNP2

(a) Before handover

Key	ID	Locator	Status	LMA	MAG	ATT	HNP
1	MN-ID	Loc.1	Standby	LMA1	MAG1	WiFi	HNP1
2	MN-ID	Loc.2	Active	LMA2	MAG2	WiMAX	HNP2

(b) After handover

Fig. 2 Binding cache entry in DSN

Fig. 3 shows results of an experiment of packet transport between MN and CN. Consistent sized (1Kbyte) TCP packets are sent from CN to MN every 10ms and the time stamps value of received data packets and signal messages are measured in MN using wireshark. MN has two different interfaces; WiFi and WiMAX. As the MN moves, it is switched from WiFi interface to WiMAX interface. In Fig. 3, the packet arrival time is not affected at all by the heterogeneous handover. The time variance is within negligible compare with the normal case. That is, handover of the MN does not affect data transmission. Also there was no packet drop during the handover. This indicates that there is no delay and loss in the situation that the access network is changed.

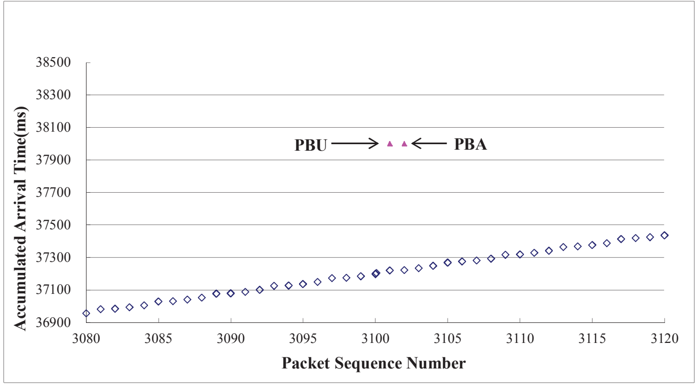


Fig. 3 Packet transport from CN to MN

From the experiment, the session between the MN and CN is maintained without any performance degrade when MN handover between heterogeneous networks. Thus, the proposed architecture is suitable in heterogeneous network environments.

IV. Conclusions

Recent mobile devices including smart-phones are evolving into multi-purpose handsets having multiple/dual interfaces. Most current mobile devices are always on, operating in a state that can receive data any time through these different interfaces. Thus, heterogeneous mobile network environments, in which the service is available through a variety of access networks and a single device, are very common. In line with such trend, heterogeneous mobility is now a basic requirement. However, the network-based mobility technology which is currently preferred by many carriers does not support it. This paper proposes a PMIP based heterogeneous mobility scheme. The proposed scheme is expected to be applied well in this situation. Even though this scheme result in change in the terminal by mounting the thin client, installing the light software is not difficult since the recent mobile devices are already considered as a small computer.

References

[1] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil,"Proxy mobile IPv6," RFC 5213, Aug. 2008.

[2] R. Wakikawa and S. Gundavelli, "IPv4 support for proxy mobile IPv6," RFC 5855, May 2010.

[3] S. Park, E. Lee, M.-S. Jin, and S.-H. Kim, "Inter-domain roaming mechanism transparent to mobile nodes among PMIPv6 networks," IEICE Trans. Commun., Vol.E93-B, No.6 Jun. 2010.

[4] T. M. Trung, Y.-H. Han, H.-Y. Choi, and Y. G. Hong, "A design of network-based flow mobility based on proxy mobile IPv6," IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp.373-378, Apr. 2011.

Smart Electric Vehicle Charging for Smart Home/ Building with a Photovoltaic System

Young-Min Wi, *Student Member*, *IEEE*, Jong-Uk Lee, and Sung-Kwan Joo, *Member*, *IEEE*
The School of Electrical Engineering, Korea University, Seoul, Korea

***Abstract*--In this paper, a smart electric vehicle charging method is proposed for smart home/building with a photovoltaic system. The proposed method is designed to determine the charging schedule of an electric vehicle based on the predicted photovoltaic output and electricity consumption. Numerical results are provided to demonstrate the effectiveness of the method.**

I. INTRODUCTION

Many recent studies have suggested that the use of electric vehicles would be a viable means of increasing the reliability of a power system in a smart grid environment [1]–[3]. If electric vehicle technologies can be deployed at low cost, such electric vehicles can also be useful in improving electrical energy efficiency and penetration of renewable energy.

This paper focuses on cost-effective methods for improving the efficiency of electric vehicle charging in smart home/building environments. Smart electric vehicle charging is one of key technologies used in home energy management systems (HEMS) as well as in building energy management systems (BEMS).

In this paper, a smart electric vehicle charging method for a residential or commercial building with a photovoltaic (PV) system that takes into account electricity price and consumption levels as well as vehicle characteristics is proposed. This method can be divided into two stages: prediction and scheduling. In the prediction stage, PV output and electricity consumption are forecasted by time series model with weather sensitivity. In the second stage, a vehicle charging schedule is determined on the basis of the results from the first stage and on the basis of the electricity price and is optimized subject to various constraints such as vehicle charge level and battery capacity, charging rate, and user preference.

II. SMART ELECTRIC VEHICLE CHARGING

In this section, the characteristics of the smart electric vehicle charging method are discussed. A block diagram for the two-stage smart charging method is shown in Fig. 1, and more detailed explanations of each component are provided in the following sub-sections.

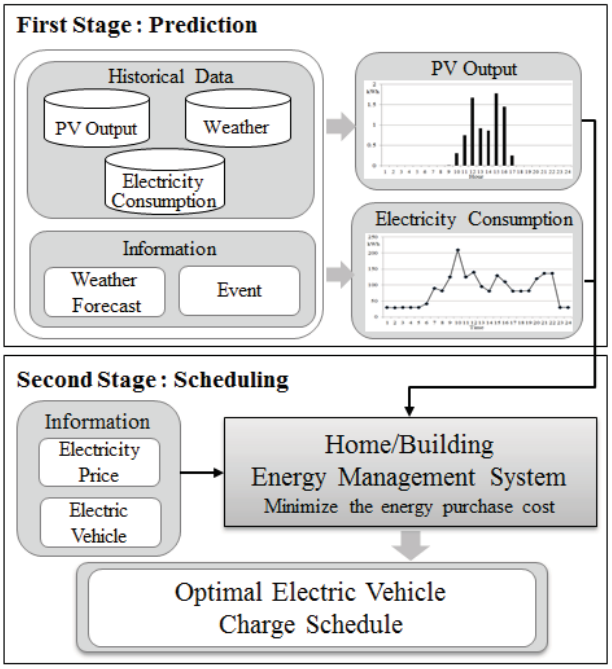


Fig. 1. Overview of the proposed method.

A. Prediction of PV Output and Electricity Consumption

Prediction of PV output and electricity consumption is required to reduce energy purchase cost for smart home/building with a PV system. The proposed method adopts time series model with weather adjustment for predictions of PV output and electricity consumption prediction. The procedure for the prediction is as follows:

Step1. Data Selection: Historical data from the most recent qualifying days are selected.

Step2. Prediction: Using data from Step 1, the PV output and electricity consumption are calculated with an exponential smoothing model.

Step3. Weather Adjustment: The forecasted values are adjusted using weather adjustment multipliers calculated as the ratio of forecasted and actual values of previous time before start time of the scheduling period.

Unlike the PV output, electricity consumption typically varies by type of day (i.e., weekend or weekday). To improve the forecasting accuracy, therefore, the day type needs to be taken into account.

B. Scheduling for Electric Vehicle Charging

The objective of solving the electric vehicle scheduling problem is to determine the best times to charge in order to

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0006346). This work was supported by the Human Resources Development of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Knowledge Economy (No. 20114010203010).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

minimize the energy purchase cost, while satisfying the constraints of the charging level and rate, battery capacity, and user convenience. An objective function for smart electric vehicle charging problem is expressed as

$$minimize\ C = \sum_{t=1}^{T} \{E_{grid}(t) \cdot P(t)\} \qquad (1)$$

where C is the energy purchase cost during the scheduling period; $E_{grid}(t)$ is the amount of electricity bought from power grid at time t; $P(t)$ is the electricity price at time t; and T is the total length of the charge scheduling period.

This optimization problem needs to be solved subject to the following constraints:

-Electricity energy balance constraint

-Battery capacity limits of each electric vehicle

-Charging rate limit of charger

-User preference: i.e., the set of target states-of-charge (SOC) and expected departure times of electric vehicles.

III. NUMAERICAL RESULTS

In this section, the numerical results are presented. For electric vehicle charge scheduling, a commercial building with 50 kW PV panels was considered. Using the assumption that three electric vehicles arrive at 09:00 with an initial SOC of 20%, 30%, and 40% and depart at 18:00 with a target SOC of 80%, predictions of PV output and electricity consumption are shown in Figs. 2 and 3, respectively.

Fig. 4 shows the electric vehicle charging schedule determined by the proposed method. Vehicles are charged during the low-price time period, and each electric vehicle meets its target (user-defined) SOC at the end of the period. From these results, it can be seen that the proposed method is effective in order to minimize the overall charging costs while attaining the target SOC.

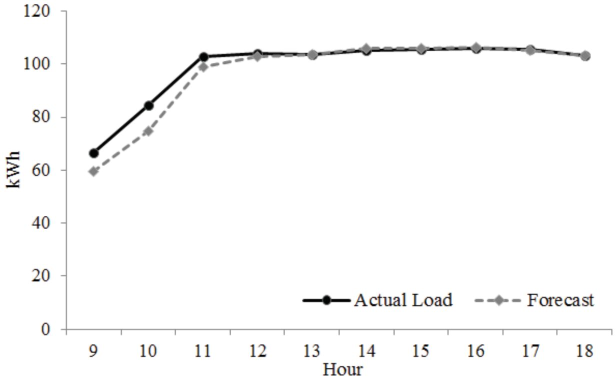


Fig. 2. Results of electricity consumption forecast.

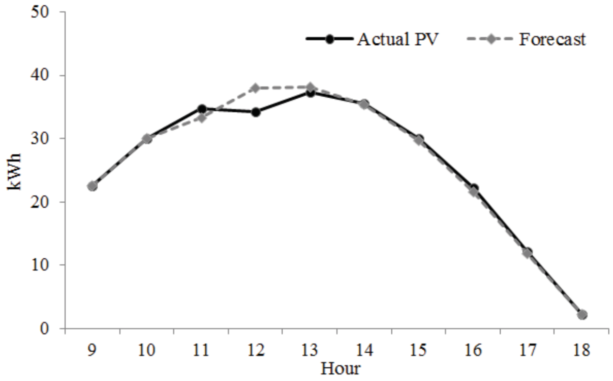


Fig. 3. Results of PV output forecast.

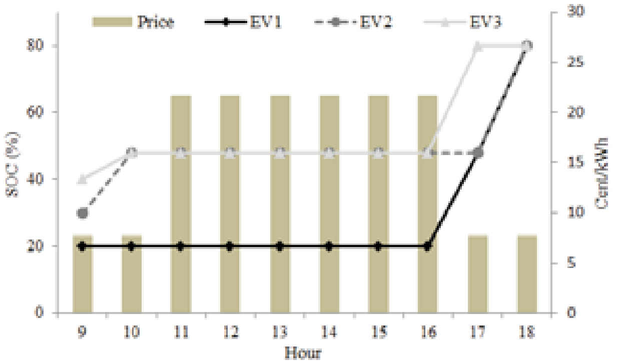


Fig. 4. Change in SOC of electric vehicle and electricity price during the scheduling period

IV. CONCLUSION AND FUTURE WORK

In this paper, a cost-effective and efficient electric vehicle charging method is introduced for smart home/building with a PV system. As described here, numerical simulation results are presented to illustrate the idea of the proposed method. Further work is required to consider the impacts of the PV output and electricity consumption forecast error and vehicle-to-grid on the performance of the proposed method.

REFERENCES

[1] Y. Ota, H. Taniguchi, T. Nakajima, K. Liyanage, J. Baba, and A. Yokoyama, “Autonomous distributed V2G (Vehicle-to-Grid) satisfying scheduled charging,” *IEEE Trans. Smart Grid,* vol. 3, no. 1, pp. 559-564, Mar. 2012.

[2] W. Shi and V. Wong, “Real-time vehicle-to-grid control algorithm under price uncertainty,” *2011 IEEE International Conference on SmartGridComm*, pp. 261-266, 2011.

[3] C. Guille and G. Gross, “Design of a conceptual framework for the V2G implementation,” *in Proc. of IEEE Energy2030,* Atlanta, GA, Nov.2008

Three-Dimensional Mobile[1]User Interface Using Finger Gestures and a Rear-Facing Camera

Byung-Hun Oh[1], Kwang-Woo Chung[2], and Kwang-Seok Hong[1], *Member,IEEE*
[1]School of Information and Communication Engineering, Sungkyunkwan University, South Korea
[2]Department of Railway Operation System Engineering, Korea National University of Transportation, South Korea

***Abstract*--We propose a 3D mobile user interface for 3D space using a mono camera in mobile devices. This paper introduces a novel interaction technique for hand held mobile devices which enables the 3D user interface to be controlled by the motion of the user's hand image taken a rear-facing camera. This proposed fingertip detection algorithm employs the Maximum Morphological Gradient Combination (MMGC) and the AdaBoost algorithm to distinguish skin color for varying lighting conditions and complex backgrounds, a feature that had captured limited in previous systems. Additionally, the proposed 3D mobile user interface can map the point of the fingertip and the estimated area in 3D space. The experimental results indicate that the proposed algorithm is applicable for mobile user interfaces based on finger movements.**

I. INTRODUCTION

In recent years mobile devices have increased in both technology and popularity. Also, mobile devices have been deployed with various new technologies, such as high quality cameras and the ability to support rich multimedia. Users want to increase the efficiency of their interactions with mobile devices and applications [1].

Because of recent research on new types of input interface systems, the latest mobile devices can currently support navigation through direction keys, keypads or scroll bars on touch-sensitive sensors. However, as the number of UI sensors increases, it becomes difficult to integrate the sensors into the hardware of existing small-form-factor mobile devices. In addition, functions cannot be represented exactly in a simplified two-dimensional (2D) form. Vision-based three-dimensional (3D) mobile UIs, however, can serve as an important mechanism for using camera-equipped mobile devices because, with their use, no new hardware is necessary.

Thus, we present a 3D mobile UI based on finger gesture estimation using the AdaBoost algorithm [2]. The proposed systems for mobile UIs require one-handed interaction and were developed for a rear-facing camera. They also estimate the finger movement and operate 3D mobile applications using finger gesture images.

II. SYSTEM ARCHITECTURE

The system architecture can be divided into the finger detection module and the 3D gesture interface module. The finger detection module is composed of two processing procedures. These procedures can respectively estimate skin color based on gradient and color information. An AND image between the morphological gradient combination image and the color-based pre-processed image is used as the input for the AdaBoost algorithm in order to detect the fingertip area. The point of the fingertip and the estimated area are then further refined and fed into the 3D gesture interface module, which determines the necessary commands for the various mobile applications.

III. FINGERTIP DETECTION

A. Finger Detection Processing

For the skin color segmentation in our proposed system, each pixel was classified as being either skin or non-skin and converted into a new binary image using the threshold value analysis defined as follows:

$$SkinColor(x,y)=\begin{cases}1 & if\ (77 \le C_b \le 127)\cap(133 \le C_r \le 178)\\ 0 & otherwise\end{cases} \tag{1}$$

To reduce the effects of small background objects in the binary image, two basic morphological operations were performed, leaving non-skin-colored objects. To remove large objects except to the finger region, we labeled each blob.

A distinctive gradient value cannot be acquired using only a grayscale image because the red, green, and blue (RGB) values in an image are obscured during its conversion to grayscale. Thus, we determined the maximum morphological gradient values in the split R, G, and B planes and then combined them into a single image. This process produced clearer gradient values than those from a grayscale image. An MMGC image is defined in the following equation:

$$MMGC=\sum_{j}^{height}\sum_{i}^{width}\max(MG_R(i,j),\ MG_G(i,j),\ MG_B(i,j)) \tag{2}$$

Finally, we can obtain an AND image using the MMGC image and the resultant image from the skin color segmentation and blob detection. The fingertip detection from this AND image includes clear gradient information and non-skin color subtraction and performs better than the original image.

B. Fingertip Detection based on AdaBoost Algorithm

For our fingertip detection system, we collected images of the half-circle area of the fingertip from the pre-processing results to use as samples. In particular, 2240 positive sample images and 4500 negative sample images were collected from

This work was supported by PRCP through NRF of Korea, funded by MEST(2012-0005861) and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2010-0021411).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

the pre-processing results for the training process. All the samples were collected under various illumination conditions. The fingertip cascade classifier is a 13-stage cascade that is 20 × 10 in size.

IV. 3D MOBILE USER INTERFACE

A. 3D Coordinate Estimation

In order to determine the spatial location of the fingertip, X and Y coordinates can be generated by changing the estimated point. In the proposed system, the center pixels of the detected fingertip were used as the X and Y coordinates. The Z coordinate was then generated by changing the area of the fingertip, which can be mapped into 3D space for implementation.

B. Finger Gesture Commands

In this proposed system, finger gestures can be performed for the operations of click, up, down, left and right. They can play the same role as a directional keypad and mouse. All of the directional commands except click are defined by the chessboard distance.

Assume that we have an $N \times M$ digital image I with $I[i, j] \in 0 \le i \le N-1, 0 \le j \le M-1$ and the current pixel position of the finger point is at (i_2, j_2), while the previous position was at (i_1, j_1). The chessboard distance is defined as:

$$d_{chess} = \max(|i_2 - i_1|, |j_2 - j_1|) \qquad (3)$$

The distance moved and the direction and instant speed of the finger point between two frames determines the directional commands.

TABLE I
Directional Commands by Finger Gesture Recognition

Directional Commands	Conditions
Up	$\lvert j_2 - j_1\rvert > \lvert i_2 - i_1\rvert$, $j_2 > j_1$, $d_{chess} > M/\sigma$
Down	$\lvert j_2 - j_1\rvert > \lvert i_2 - i_1\rvert$, $j_2 < j_1$, $d_{chess} > M/\sigma$
Left	$\lvert i_2 - i_1\rvert > \lvert j_2 - j_1\rvert$, $i_2 < i_1$, $d_{chess} > N/\sigma$
Right	$\lvert i_2 - i_1\rvert > \lvert j_2 - j_1\rvert$, $i_2 > i_1$, $d_{chess} > N/\sigma$

Table.1 shows each of the directional commands. In the table, the variable σ can be varied according to the frame rate of the system properly. In the proposed system $\sigma = 4$ is used with the frame rate of 10fps, 24 bit color, and 320 x 240 resolutions [3].

The click command is generated by the area of the fingertip. The instantaneous rate of change of the fingertip area is applied to the click command by finger gestures, and is defined as the fingertip area of the current frame over the fingertip area of the previous frame. When a user moves his finger back and forth toward the camera similar to the gesture of clicking a mouse, a click command is triggered.

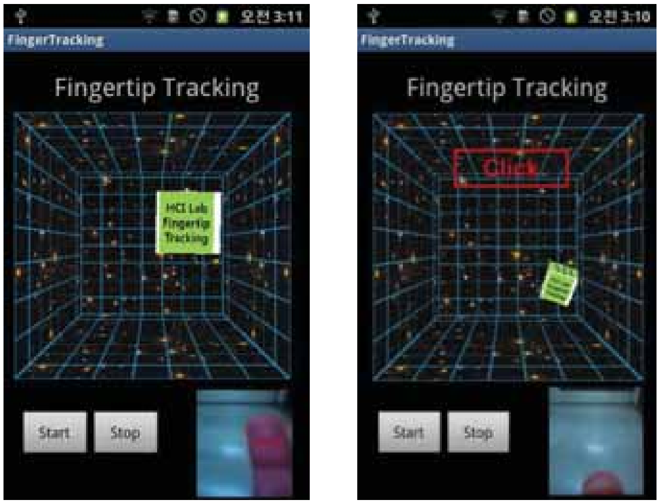
V. EXPERIMENT RESULT AND APPLICATION

A. Experiment

To evaluate the finger interface recognition performance, five participants used a directional command test program. Each finger interface input command was tested 200 times. The experiment result shows the accuracy above 93%.

B. Application

We implemented a few test applications to highlight the strengths and limitations of the proposed system. Figure 1(a) demonstrates that the proposed system can accurately estimate the 3D position and is applicable to 3D user interfaces. Additionally, Figure 1(b) presents click command examples for various finger movements.



(a) (b)

Fig 1. 3D Fingertip Tracking Interface : (a) 3D Application Control (b) Click Command Examples

VI. CONCLUSION

We proposed a 3D mobile user interface for 3D space using a mono camera in mobile devices. The proposed system required one-handed interaction and was developed for a rear-facing camera. Moreover, it estimated the finger movements and then operated 3D mobile applications using images of these finger gestures taken through the rear-facing camera. Consequently, this study confirmed the feasibility of the proposed algorithm for finger-movement-based mobile user interfaces.

REFERENCE

[1] Eunjin Koh, Jongho Won, Changseok Bae, "On-premise skin color modeling method for vision-based hand tracking," Consumer Electronics, 2009. ISCE '09 IEEE 13th international Symposium on, vol., no., pp.908-909, 25~28 May, 2009.

[2] V. Paul, J. Michael, "Rapid Object Detection using a Boosted Cascade of Simple Features", IEEE Conference on Computer Vision and Pattern Recognition, Vol.1, pp. 511-518, 2001

[3] Jun-Ho An, Jin-Hong Min and Kwang-Seok Hong, "Finger Gesture Estimation for Mobile Device User Interface Using a Rear-Facing Camera", Communications in Computer and Information Science, 2011

Real-Time Hand Shape Recognition by Orientation Invariant Data Learning for Smart TV

Jae-Joon Han, Changkyu Choi, ByungIn Yoo, Dusik Park, and Changyeong Kim
Advanced Media Lab., Samsung Advanced Institute of Technology, Samsung Electronics, Korea

***Abstract*—The paper proposes a novel recognition system for hand shapes at a distance for smart TV. Two types of hand shapes are selected for the needs of TV browsing. The proposed method provides robust recognition performance on various hand orientations and guarantees real-time computation.**

I. INTRODUCTION

Smart TV has brought a great attention by providing new experience such as internet and various applications including streaming videos and games. Such experience naturally needs a different type of user interface to support easy browsing rather than button-type remote controllers [1]. Recently, TV manufacturers have started to include new user interfaces which recognize hand gestures. In addition, Kinect provides full body motion tracking capability which enables browsing on the screen by hands. While it provides more natural interface to users, it only utilizes a hand position with temporal information. Therefore interaction such as selecting, rotating, dragging, and dropping cannot be achieved instantly without considering hand shape recognition. Moreover, since TV users may browse while lying on a sofa, hand shape recognition method which is robust to various orientations is of importance.

Hand shape recognition has been studied widely using color and depth images. The most studies consider the contour information of the hand after extracting the hand region as a shape descriptor [2]. Such studies have a limitation when the outlines of the different hand shapes are similar due to the hand orientation. Therefore, the limited range of the orientation is important to ensure the performance. To overcome such limitation, the recent study proposes a 3D volumetric shape descriptor which is a 2D shape descriptor layered along the main PCA axis. While the performance of the proposed 3D shape descriptor outperforms the 2D shape descriptor, it requires averaging over some past image sequences [3]. As for feature extraction methods, a depth comparison feature method proposed by Shotton et al. is used in full-body pose estimation. The depth comparison feature is used with random forest for per-pixel classification [4]. On the other hand, labeling pixels requires recognition processing at each pixel on the entire foreground image.

The paper presents a hand-shape recognition method without any orientation limitation. Instead of using per-pixel classification, the proposed method applies a random forest trained on a data set with large variations in orientation using the depth comparison features for recognizing a shape.

II. THE PROPOSED METHOD

A. Hand Shape Recognition System

The proposed system has three components of a hand region segmentation module, a depth feature extraction model, and a shape recognition module as shown in Figure 1. The segmentation module acquires the hand depth image of 128x128 pixels centered at the 2D pixel position of the hand joint from the skeleton module of Kinect for windows® as shown in Fig. 1 (a), and then segments the 3D hand region within the 3D axis-aligned bounding box centered at the 3D hand position where the size of each axis is 20 cm as shown in Fig. 1 (b).

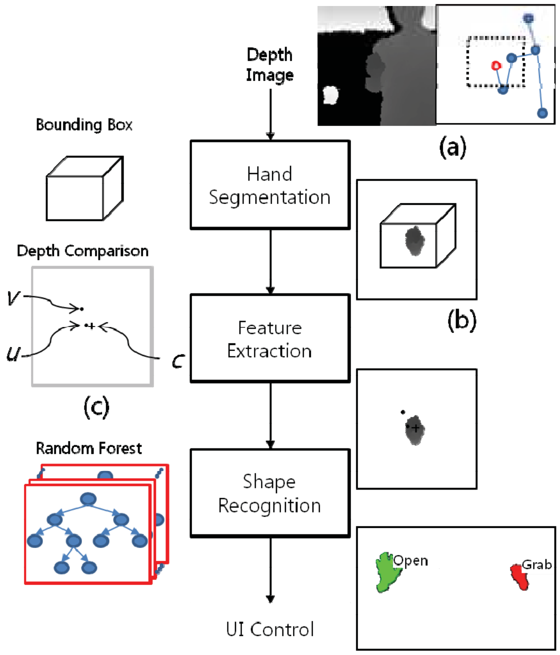


Fig. 1 The proposed hand shape user interface system.

The feature extraction module generates a large number of depth comparison features. Each feature is computed by comparing the depth values of the two randomly selected pixel points on the segmented hand image similar to the features used in [4] as shown in Fig. 1 (c). In detail, each depth comparison feature is obtained as

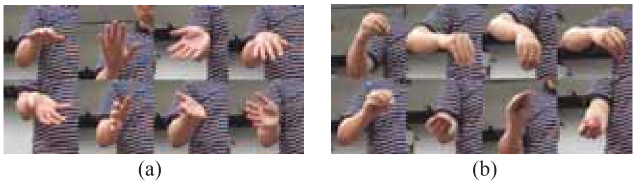
$$f_{(u,v)}(I,c)=d_I\left(c+\frac{u}{d_I(c)}\right)-d_I\left(c+\frac{v}{d_I(c)}\right),$$

where $d_I(x)$ is the depth at pixel x in image I; c is the center pixel position of the image; u and v are two randomly selected pixel points. Note that the two sampled points are generated from Normal distribution (~$\mathbf{N}(c, \sigma^2)$) with its mean which is

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

located at the center of image, because the extracted hand region is already located at the center of the image.

The hand shape recognition module recognizes a hand shape by a random forest classifier learned in advance with ground truth data. The ground truth data contain hand images of various rotations to be robust for hand shapes in different orientations as in Figure 2. Because the classifier is trained with right hand images, the recognition module uses the mirror image when recognizing the left hand.



(a) (b)

Fig. 2 Examples of hand poses with variations in terms of shape and orientation: (a) Open palm and (b) Grab

B. *Training Hand Shape Recognizer*

The paper considers the smart TV user interface such as browsing, selection, dragging, and dropping. Therefore, two hand shapes of "open" and "grab" are selected in that humans "grab" an object to select and "open" to release. The ground truth data are composed of 32,245 depth images from seven different users ranging from 1 to 2.5 meters away from a camera. Each hand shape of "open" and "grab" is captured with small degree of shape variations and large degree of orientation variations to make a robust hand shape recognizer.

The hand region segmentation and feature extraction methods for training are the same as the recognition system. In training random forest, the randomly selected features among all the features are used to find the best split. Here the number of the randomly selected features is set to the square root of the total number of features. As for stopping criteria, the maximum depth of each tree is set to 20; the minimum number of samples for a node to be split is set to four. Training 100 random trees with 2500 depth comparison features extracted from each of 32,245 training depth images requires about 2 hours on a 2.93 GHz Intel Xeon CPU with a C++/OpenCV implementation.

III. Experimental Results

To evaluate the performance of the hand shape recognition, we compared the proposed method with a random forest based on Hu moment feature [5]. As for the proposed method, the number of the features is determined as 2500 by experiment. The total 3720 testing images are captured independent of the training images from three different users. And then, the recognized shape is compared to the ground truth for each image. Figure 3 shows the performance of each method. The proposed method outperforms the Hu moment based method by about 5 percent in recognition rate. As the size of the tree increases, the recognition rates of both methods increase. However, due to the large dimension of the depth comparison feature, the proposed method requires larger number of trees than the Hu moment based method to converge. Table 1 shows the recognition rate on two hand shapes of the proposed method in terms of the number of the trees. The implemented hand shape recognizer with a 2.93 GHz Intel Xeon CPU takes 6.6ms per image on average from the hand segmentation to the hand shape recognition without any GPU acceleration.

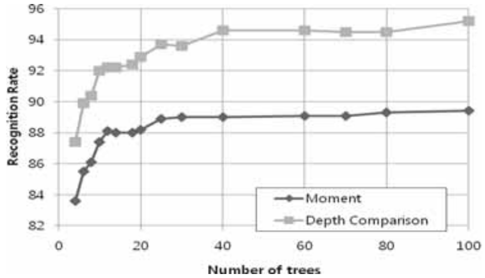


Fig. 3 Performance on the proposed and the Hu moment based methods

IV. Conclusion

TABLE I
PERFORMANCE ON EACH HAND SHAPE WITH DEPTH COMPARISON FEATURE

Number of Trees	Open Palm	Grab	Overall
100	96.7	93.4	95.2
80	96.2	92.6	94.5
70	96.4	92.4	94.5
60	96.3	92.6	94.6
50	96.6	92.6	94.7
40	96.5	92.5	94.6

The paper presented a novel method of recognizing hand shapes. By applying simple low-level depth comparison features on a centralized hand segmented image with random forest, it is possible to provide the robust recognition performance on hand shape with various orientations. Due to the simplicity of the feature computation, it achieves the real-time performance without any additional computational acceleration. The proposed method is applicable to hand-shape based user interfaces for Smart TV which is naturally more intuitive to humans.

References

[1] B. Yoo, J.-J. Han, C. Choi, H. Ryu, D. Park, and C. Kim, "3D Remote Interface for Smart Displays," *In Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, 2011,* pp. 551-560, Apr. 2011, ACM.

[2] Z. Ren, J. Meng, J. Yuan, and Z. Zhang, "Robust Hand Gesture Recognition with Kinect Sensor," *In Proceedings of the 19th ACM international conference on Multimedia*, pp. 759-760, ACM.

[3] P. Suryanarayan, A. Subramanian, and D. Mandalapu, "Dynamic Hand Pose Recognition using Depth Data, " *In Proceedings of the 20th ICPR*, pp. 3105-3108, Aug. 2010.

[4] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-Time Human Pose Recognition in Parts from a Single Depth Image," *In Proceedings of 2011 IEEE CVPR,* Jun. 2011.

[5] S. Conseil, S. Bourennane, and L. Martin, "Comparison of Fourier Descriptors and Hu Moments for Hand Posture Recognition," *In Proceedings of EUSIPCO, 2007.*

An Interactive Toolkit for Designing Vibrotactile Haptic Messages

Anak Agung Gede DHARMA, *Member, IEEE* and Kiyoshi TOMIMATSU

Abstract-- **In this paper, we propose an interactive way and toolkit for designing custom haptic messages. Our proposed toolkit can assist users in designing haptic messages for complicated expressions, as proven by user testing results.**

I. Introduction

Haptic feedback plays important roles in delivering non-intrusive message. Notable examples of its importance can be found in various devices, such as mobile phone, Personal Digital Assistant (PDA), game controllers, and medical instruments. Furthermore, haptic feedback plays a significant role for supporting daily lives of visually impaired persons.

On the other hand, designing haptic feedback still has remaining problems. Even in this age of media and telecommunication, it is still limited to a combination of simple force patterns. The main problem lies in the design of haptic feedback, which requires fine and careful tuning. Designing haptic feedback also involves subjective perception or sensibility. Although physiological tactile response has been reported in several researches, the non-linearity between haptic stimulus and its subjective perception remains an open question [1]. One of the possible solution to overcome this problem is by involving subjective human perception in the design of haptic force patterns (which is also known as hapticon).

One of the earliest usage of the term hapticon was proposed by Rovers et al. (2004). Hapticon can be defined as a small-programmed force pattern that can be used to communicate a basic notion in a similar manner to the icons used in a graphical user interface [2]. Designing hapticon is often described as an abstract design process. At present, haptic feedback is designed based on subjective perception on its designer. As the result, the final design often does not meet the demand of the user. Thus, a thorough study regarding the usability of haptic stimuli is essential before these artificial haptic feedbacks can be successfully implemented into real-world application.

The objective of this study is to propose an alternative system for designing haptic messages, including hapticons. In this study, we have developed a working prototype for displaying various types of haptic messages. We propose an interactive evolutionary computation user interface to design haptic messages [3]. In this method, user's role is to feed simple Boolean value to the computer for multiple iterations until a suitable haptic message can be found. On the other hand, we also provide a user interface for manual editing. To measure the effectiveness of our proposed evolutionary computation model, we performed user testing by comparing pairs of hapticon that are designed by those two different methods. Three expressions were selected as the task for hapticon design, i.e. happiness, sadness, and anger.

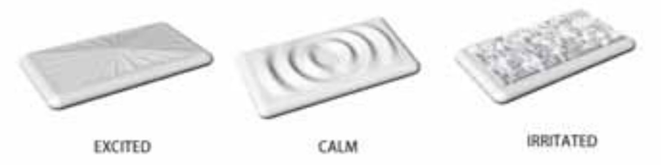


Fig. 1. Possible corresponding textures to certain emoticons, as described by Hunter Sebresos [7]

This paper consists of eight sections in total. Section II discuss about previous researches that are related to this study, especially relevant works regarding haptic communication and subjective perception. Section III explains the principle of physiological properties of haptic sensory, which is used as the basis to create artificial haptic messages and stimuli, as described in section IV. Two different modes of user interfaces are described in section V and the effectiveness of each mode is described in section VI. In section VII, we attempt to describe future application possibilities that can be realized by using our proposed toolkit. Finally, our conclusion and suggestions for future work are described in the last section.

II. Related Works

Various types of haptic perceptions and their emotional correlations have been investigated in preceding researches. One of the earliest researches is done by Rovers et al., which develop haptic instant messaging framework that combines text messages with haptic effects and hapticons [2]. Furthermore, Shin et al. developed tactile emotional interface for instant messenger chat, which consists of vibrating motors, pin actuators, heat coil, pressure sensors, buttons, and LEDs [4]. The device utilizes an intuitive tangible input-output method for displaying hapticons, which also includes a user interface to allow its users to create custom hapticon.

An attempt to apply haptic feedback technology to interpersonal communication has been described by Brave et al. in inTouch, which creates a physical link between distant users by physical analog movement of rollers [5]. Furthermore, Chang et al. develop a device that convey nonverbal signals in ComTouch, i.e. by converting hand pressure into vibration

This work was supported by JSPS KAKENHI Grant Number 245436

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

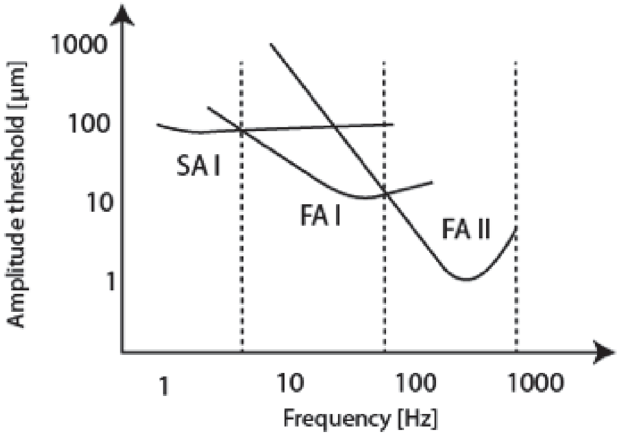


Fig. 2. Detection thresholds of vibratory stimuli (based on Konyo [8])

between users [6].

A hapticon concept design with emphasis to its detailed surface properties is described by Hunter Sebresos (Fig. 1) [7]. Sebresos mentions the possibility of associating certain patterns or textures with certain emotions. This study aims to simulate surface tactile properties and correlates them to their corresponding emotional expression as described by Sebresos.

III. Selective Stimulation Method

The concept of selective stimulation method is based on the fact that tactile receptors in human skin cannot sense physical factors directly. They can only detect inner skin deformation caused by contacting objects [8]. Thus, we may be able to activate tactile receptors' nerves as if physical factors of any material come in contact with the skin by giving artificial tactile stimulation.

Selective stimulation method is based on the direct manipulation of three tactile receptors, i.e. Fast Adapting Afferents Type I (FA I), Fast Adapting Afferents Type II (FA II), and Slow Adapting Afferents Type I (SA I). Each receptor has spatial and temporal response characteristics for physical stimulation, as described in Figure 2. Each receptor also causes subjective sensation that corresponds to inner skin deformation. The utilization of selective stimulation method and selected frequency ranges that are used in this study will be described in next sections.

IV. Artificial Haptic Message Design

As described in section 3, each tactile receptor has its own spatial and temporal properties. Thus, it responds to its corresponding detection threshold frequency. In this research, we stimulate three tactile receptors based on their respective detection thresholds, as described by Konyo et al. [8].

- FA I: most sensitive between 25 – 40 Hz;
- FA II: most sensitive between 200 – 250 Hz;
- SA I: most sensitive between 0.4 – 7 Hz.

In this study, we define 0.4-7 Hz, 25-40 Hz, 200-250 Hz as low frequency, medium frequency, and high frequency, respectively (Fig. 3).

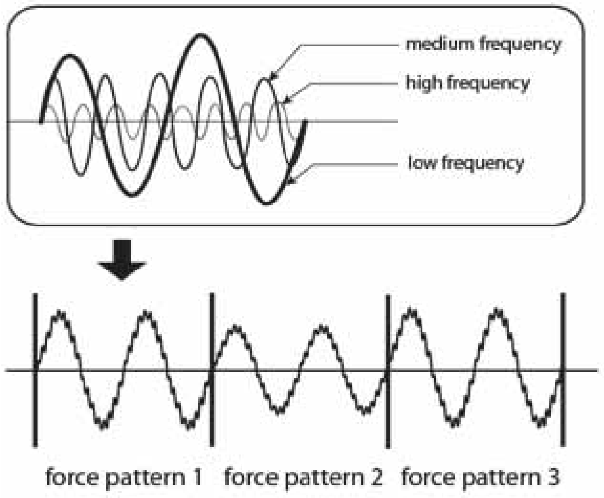


Fig. 3. Hapticon design by superposition

TABLE I
Three Frequency Variables for A Given Force Pattern

Force Pattern	Receptor Target	Frequency Range (Hz)
Frequency_FA1	FA1 (Meissner)	25 – 40
Frequency_FA2	FA2 (Pacinian)	200 – 250
Frequency_SA1	SA1 (Merkel)	0.4 – 7

TABLE II
Three Amplitude Variables for A Given Force Pattern

Force Pattern	Receptor Target	Amplitude Range (micron)
Amplitude_FA1	FA1 (Meissner)	0 – 450
Amplitude_FA2	FA2 (Pacinian)	0 – 120
Amplitude_SA1	SA1 (Merkel)	0 – 600

Our hapticon design concept and the correlation between its variables are illustrated in Figure 3. The hapticon is designed by superpositioning three haptic vibrations of different frequency ranges, i.e. constructive interference among three different frequency ranges.

In this study, a hapticon is constructed of 3 individual force patterns that are played back in sequential order. Each force pattern has 6 variables (i.e., 3 frequency variables and 3 amplitude variables) as described in table I and table II. Furthermore, each force pattern has a fixed duration of one second.

V. User Interfaces for Hapticon Editing

In addition to artificial haptic message design, we provide two kinds of user interfaces to edit the hapticon properties.

One of the main challenges in designing hapticon is complexity, particularly regarding the non-linear correlation

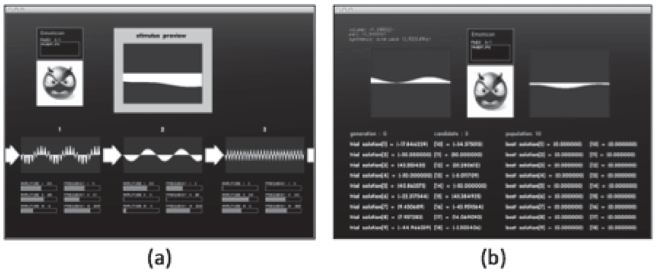


Fig. 4. Two Graphical User Interface (GUI) for editing hapticon, i.e. manual tuning (a) and evolutionary computation (b)

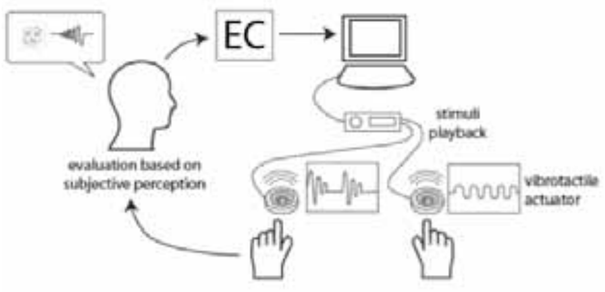


Fig. 5. The schema of hapticon tuning with evolutionary computation algorithm

between haptic perception and physical properties. On the other hand, interactive design techniques can solve this problem by incorporating human cognitive ability in the design process. User actively involves in the hapticon design process by providing input to the computer for a definite number of iterations.

We propose a user interface that utilizes evolutionary algorithm as the intermediary for the optimization process. On the other hand, we also provide an option to manually tune each frequency and amplitude variable for every force pattern (i.e., manual tuning method).

Our main emphasis is on the simplicity and ease of use of the interface. We assumed that haptic message display does not require complicated Graphical User Interface (GUI). Thus, we are using two-dimensional visual representation of haptic feedback and only display important variables in the GUI. The differences and detailed descriptions of two different modes of hapticon editor are described below.

A. Manual Tuning

Manual tuning is a standard method for editing hapticon variables. User manually tunes 3 frequency variables and 3 amplitude variables for each force pattern. A hapticon consists of 3 force patterns, therefore the user has to tune 18 variables in total. The GUI for manual tuning mode is described in Figure 4-a.

B. Evolutionary Computation

In addition to manual tuning, we provide evolutionary computation (EC) user interface. We utilized differential evolution algorithm, i.e. a form of evolutionary computation that is developed by Price et al. [9]. The schema of this algorithm and GUI are described in Figure 5 and Figure 4-b, respectively. User and computer form a closed system that continuously loops. Human cognition is being involved into computing process to determine fitness value, while the algorithm to search for optimal value is executed by Central Processing Unit (CPU). User has to choose either one (left or right) of the stimuli. The iteration process continues until the most suitable hapticon that can represent a certain expression can be found. The sequential steps for the iteration process are described below.

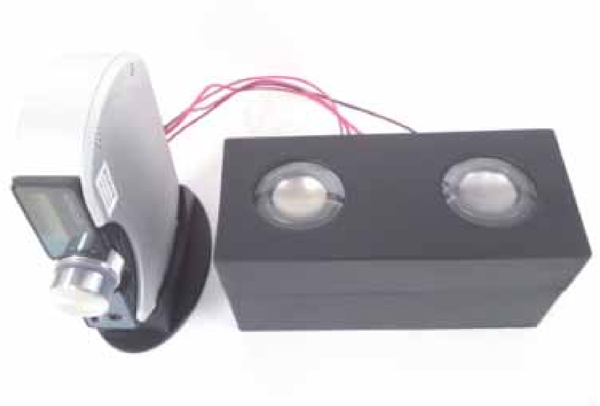


Fig. 6. Hapticon playback toolkit, which consist of a digital amplifier (left) and a pair of vibrotactile actuators (right)

- CPU sends a pair of signal to vibrotactile actuator, which are known as trial vectors;
- These signals will be rendered by digital amplifier, which in turn will be rendered by vibrotactile actuators as hapticon;
- User feeds the system with a feedback by deciding which one of the hapticon pair represents a given expression;
- User feedback determines fitness value, thus better hapticon become the best individual in the next generation and worse hapticon is discarded;
- This process is repeated according to the number of population size, i.e. 10 for this case;
- The iteration process continues to the next generation, the offspring for the next generation are generated based on fitness value;
- The whole iteration process is repeated until user has succeeded in finding an adequate hapticon that represent a given expression.

VI. Discussion

We performed a preliminary user testing to measure the performance of both user interfaces. Three design concepts of haptic expressions are given to six designers, i.e. the expressions of happiness, sadness, and anger. They are requested to design hapticons for these expressions with two different user interfaces as described in section V, i.e. manual tuning and evolutionary computation.

We conducted a subjective test to compare and evaluate the performance of these methods, 20 subjects participated in this study [10]. Eighteen sets of hapticons (3 hapticons made by 6 designers with manual tuning and 3 hapticons made by designers with evolutionary computation) and the comparison

TABLE III
HAPTICON DESIGN TASKS AND USER TESTING RESULT

Hapticon design task	Emoticon counterpart	EC interface is preferred over manual tuning[a]	Manual tuning is preferred over EC interface[a]
Happiness expression	😀	50% (3 out of 6 hapticon designs)	0% (0 out of 6 hapticon designs)
Sadness expression	😢	50% (3 out of 6 hapticon designs)	0% (0 out of 6 hapticon designs)
Anger expression	😠	33% (2 out of 6 hapticon designs)	33% (2 out of 6 hapticon designs)

[a] the preference is based on the result of Wilcoxon signed rank test result. A type of interface is significantly preferred over another if it has a P-value of less than 0.05 ($p < 0.05$).

data were measured by Wilcoxon signed rank test. The result shows that in most of our cases, preferable hapticons were produced by EC interface. Furthermore, it implies that user can differentiate poorly designed hapticon and appropriate hapticon for each given expression.

In most cases, EC interface can help designers to create better hapticons compared to manual tuning method. For happiness and sadness expression, 50% designers could significantly create better hapticon designs by using hapticon design support system ($p < 0.05$). However, the result for anger expression is diverse. Two designers (33%) could create preferable hapticon designs by using hapticon design support system while the other two succeeded in creating better hapticon with manual tuning method ($p < 0.05$) [10].

The results suggest that our evolutionary computation interface is more effective for difficult expression that requires careful tuning such as happiness and sadness. However, simple expression like anger is quite easy to understand and can be designed by maximizing amplitude within the range of high frequency (200-250 Hz), which produces friction sensations and unpleasant feelings.

VII. POSSIBLE APPLICATION SCENARIOS

We intend to utilize haptic messages developed in this study to assist the elderly and visually impaired people in their daily life, for example, by providing haptic feedback for street navigation application. In addition, there is also a potential to expand the scope of the artificial haptic stimuli developed in this study for common users. For instance, there are special scenarios where haptic feedback is the only available way for users, e.g. in a meeting room where all devices have to be set to silent mode. In those kinds of situations, hapticon can be used to provide a prior notice to the user regarding the importance and status of the incoming message (Figure 7).

Other possible applications include simulated experience for virtual environments, information or data haptification, and to provide haptic memory for tangible objects.

VIII. CONCLUSION AND FUTURE WORKS

This study confirms that designing haptic message (especially hapticon) involves subjective perception and requires fine and careful tuning. We have confirmed that user can easily evaluate the quality and suitability of hapticon for a given expression, thus providing the basis for tactile communication. Furthermore, these findings emphasize the possibilities of incorporating it into real-world applications, such as haptic support for the visually disabled or the elderly.

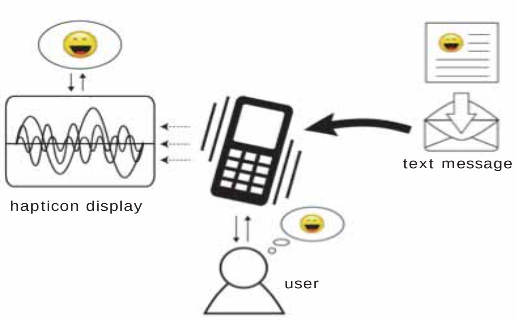


Fig. 7. A possible application scenario where hapticon is used to give prior notification regarding the content of text message

Future works will include expanding the scope of the study to provide a comprehensive guideline for haptic encoding, creating additional features to user interface, and performing user testing to measure haptic messages capability in specific scenarios.

ACKNOWLEDGMENT

We would like to express our gratitude to Dr. Hideyuki Takagi, who provides support for user testing experiment.

REFERENCES

[1] X. Chen, F. Shao, C. Barnes, T. Childs, and B. Henson, “Exploring relationships between touch perception and surface physical properties,” *International Journal of Design*, Vol. 3, No. 2, pp. 67-76, 2009.

[2] A. F. Rovers, H. A. van Essen, “HIM: a framework for haptic instant messaging,” *CHI ’04 Extended Abstracts on Human Factors in Computing Systems* (Vienna, Austria), pp. 1313-1316, April 2004.

[3] H. Takagi, “Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation,” *Proceedings of the IEEE*, vol. 89, no.9, pp.1275-1296, 2001.

[4] H. Shin, J. Lee, J. Park, Y. Kim, H. Oh, and T. Lee, “A Tactile Emotional Interface for Instant Messenger Chat,” *Proc. Int. Conf. on Human-Computer Interaction*, pp. 166-175, 2007.

[5] S. Brave, and A. Dahley, “inTouch: a medium for haptic interpersonal communication,” *CHI ’97 Extended Abstracts on Human Factors in Computing Systems* (New York, USA), pp. 363-364, 1997.

[6] A. Chang, S. O’Modhrain, R. Jacob, E. Gunther, and H. Ishii, “ComTouch: design of a vibrotactile communication device,” *Proc. 4th Conf. on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, pp. 312-320, 2002.

[7] H. Sebresos, (2012, July) “Mobile Hapticons,” [Online], Available: http://concepthunter.com/touch/hapticons.html

[8] M. Konyo, A. Yoshida, S. Tadokoro, and N. Saiwaki, “A tactile synthesis method using multiple frequency vibrations for representing virtual touch,” *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (Alberta, Canada), pp. 1121-1127, August 2005.

[9] K. Price, R. Storn, and J. Lampinen, “Differential Evolution – A Practical Approach to Global Optimization,” Springer, Berlin, 2005.

[10] A. A. G. Dharma, H. Takagi, and K. Tomimatsu, “Emotional Expressions of Vibrotactile Haptic Message Designed by Paired Comparison-based Interactive Differential Evolution,” *2011 Evolutionary Computation Symposium* (Iwanuma, Japan), pp. 247-252, December 2011 (in Japanese).

A Robust Human Pointing Location Estimation Using 3D Hand and Face Poses with RGB-D Sensor

Donghun Kim and Kihyun Hong

Purdue University, School of Electrical and Computer Engineering, West Lafayette, IN 47906, USA

***Abstract*—We present a robust method to estimate human directed target positions in a display. In the proposed method, we utilize 3D hand and face poses instead of hand pose only, to compute human pointing directions and increase accuracy of location estimation. To combine hand and face pointing location estimates, a soft-switching fusion is applied. In the proposed approach, we use a RGB-D sensor which gives more information (depth information) than conventional camera sensor. As a result, we present an experiment that shows the proposed method is more accurate than a single hand or a face pose based target point estimates.**

I. Introduction

As human interactive consumer electronic devices such as gesture recognition based smart TVs are introduced in the market, the needs of remote posture activation related technologies become growing. As enabling technologies, human posture and hand pose estimation methods [1-3] were proposed and applied to many devices. However, these methods provide only limited control signals to devices; some pre-defined postures or gestures can only be allowed to recognize. If an interactive system utilizes a human pointing estimation [4-6] in addition to human gestures, it will give much freedom to a device control methodology. It may enable control pad or mouse-free machine interaction. In this paper, we introduce a robust method for target pointing estimation using 3D hand and face poses. Sometimes, target pointing estimate with a single reference such as hand or face pose may not be reliable because of visual perception and pose estimation errors. To increase estimation accuracy, we use both hand and face information with a soft switching fusion in the presence of human motion along depth change.

On the other hand, we utilize a RGB-D sensor which provides depth information along with a RGB channel camera. It allows robust hand and face detections than conventional camera sensors.

II. Pointed Target Detection

A. Human pointing gesture

The goal is to obtain a target position on a pre-defined plane i.e., a display panel, with respect to the camera system in Fig. 1. To achieve the task successfully, we need three modules of algorithms for head and hand detection, their 3D pose estimation, and robust pointed target estimation as shown in Fig. 2. We utilize color and depth information to detect and estimate 3D poses of both head and hand. Each pose results in the ray of a 3D pointing direction so that we compute the intersection of a ray with a pre-defined plane in a Plücker coordinate [7]. Note that we need to calibrate the pointing action to compensate visual perception error.

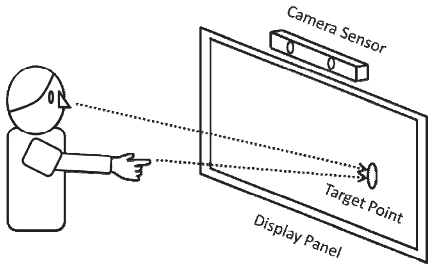


Fig. 1 Target pointing system

B. Head detection and 3D pose estimation

Head detection is completely based on face detection and its 3D pose is estimated by utilizing 3D active appearance model of the face [8-9]. This model is constructed by triangle meshes of structural landmarks to consider shape variation and appearance variation simultaneously. We utilized a PCA model about variations with the pre-trained model provided from Kinect SDK [10] instead of training possible experimental databases that confined with small numbers of samples on poses, races, ages and so on.

The face is initially detected from trained color and features with a depth range filter. After locating the face pose in a coarse, head pose is estimated by fitting a 3D active appearance model known its 3D pose in a camera coordinate.

Using the fitted model at an estimated 3D pose, we obtain the ray of a head direction with a normal vector on a frontal triangle patch defined from the landmarks of meshes.

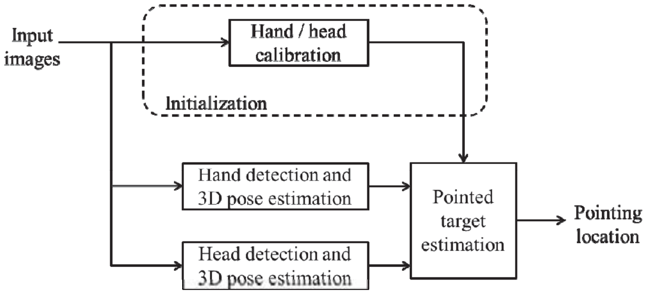

Input images
Hand / head calibration
Initialization
Hand detection and 3D pose estimation
Head detection and 3D pose estimation
Pointed target estimation
Pointing location

Fig. 2 Overview of an algorithm

C. Hand detection and 3D pose estimation

A hand is detected with a help of skeleton data estimated by random forest decision [11]. Then we utilize a depth filter in 3D space and color filter in a 2D image plane to find the hand area in 3D. To compute a hand direction, we apply the 2-mean clustering with initial points of min-max depth values. Then we obtain the ray of a hand direction using two cluster means in a simple and efficient way. Furthermore, we select a tip of

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

hand for more accurate 3D pose by using depth information. This refinement gives us a more robust directional vector regardless of hand shapes like fist, palm, or fingers.

D. *Intersection between a ray and a plane*

The line in 3D can be represented by a 4 x4 skew-symmetric homogeneous matrix, Plücker matrix [7]. Given two 3D points *P*, *Q* in terms of above directional vectors, we can represent a ray with a Plücker matrix, *L*:

$$L = PQ^T - QP^T.$$

Then we can compute the intersection, *M*, of a ray with a plane π by

$$M = L\pi.$$

In this paper, we use the z = 0 plane under the assumption that the interesting plane such as a display panel is on the X-Y plane in the camera coordinate as in Fig. 1.

III. Hand-face calibration and fusion

A. *Hand-face calibration*

The simplest way is pointing out to the landmarks known 3D position as the ground truth. We record the corresponding pairs. Then we can estimate the parameters of a calibration model, homography, from error distribution between the ground truth and target pointing results in a linear least mean square error sense. Then we can apply this calibration model to the hand and face pointing estimates.

B. *Hand-face data fusion*

We do weighted matrix summation to hand and face based estimates. Using calibrated results in terms of hand and head, we determine appropriate weight matrices in inverse proportion to calibration errors. Then we apply the matrices to obtain the better target pointing result by data fusion.

IV. Experimental result

We tested four target points on a display in the presence of distance changes to the camera. Head and face calibrations were done at a distance of 64 inches to the camera. Testing samples were captured within a distance from 54 inches to 72 inches. For calibrations with known four target points, we used about 28 samples each obtained by pointing gesture at a fixed human standing position. Test samples were about 50 samples each to four target points in association with different human position along distances to the camera. As a result, we presented the error and its standard deviation between estimated target position and the ground truth in Table 1. Also, we illustrated the result on a display panel in X-Y axis in Fig. 3. The markers of square, x, +, o represent the ground truths, hand pointing result, face pointing result, and hand-face combined pointing result, respectively. Therefore, we are convinced that our proposed method is improved not only with less 29% error, but also with less 54% standard deviation than pointing location estimation by means of a hand pose only.

TABLE I
Average of estimated point errors and standard deviations

	Error [a]	Error std. [a]
Hand	9.69	7.04
Face	10.02	4.90
Combined	6.89	3.23

[a] Unit: cm

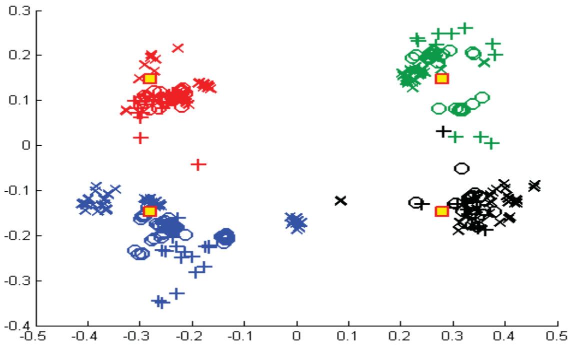


Fig. 3 Estimated target points on a display panel.

V. Conclusion

As the size of a display panel is getting larger recently, the hand pointing gesture is considerable as the emerging natural interface. Hand calibration is necessary to perform an appropriate interaction because of human visual perception error. In this paper, we proposed a robust way of pointing location estimation on a display by means of combined calibration with both hand and face poses. This approach results in robust pointing target estimation in the presence of human motion along the distance to the camera.

Reference

[1] R. Poppe, "Vision-based human motion analysis: An overview," *Comp. Vis. Image Und.*, 2007, pp. 4-17.

[2] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, pp. 607-626.

[3] A. Erol, G. Bebis, M. Nicolescu, and R. Boyle, "Vision-based hand pose estimation: A review," *Comp. Vis. Image Und.,* 2007, pp. 52-73.

[4] D. Kim and Y. Sivathanu, "An information delivery system for visually impaired people in dynamic environment," *IEEE Int'l Conf. on the Sys. Man and Cyber*, 2011, pp. 2062–2067.

[5] P. Matikainen, P. Pillai, L. Mummert, R. Sukthankar, and M. Hebert, "Prop-free pointing detection in dynamic cluttered environments," *IEEE Int'l Conf. on the Automatic Face and Gesture Recognition and Workshops*, 2011, pp. 374–381.

[6] S. Carbini and J. Viallet, "Pointing gesture visual recognition for large display," *Int'l Conf. on Pattern Recognition*, 2004.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. 2nd ed., 2004.

[8] M. Zhou, L. Liang, J. Sun, and Y. Wang, "AAM based face tracking with temporal matching and face segmentation," *IEEE Int'l Conf. on Computer Vis. Pattern Recog.*, 2010, pp. 701–708.

[9] T. Cootes, G. Wheeler, and K. Walker, "View-based active appearance models," *Image and Vision Computing*, 2002.

[10] Kinect for Windows [Online]. Available: http://www.microsoft.com/en-us/ kinectforwindows/

[11] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," *IEEE Int'l Conf. on Computer Vis. Pattern Recog.*, 2011, pp. 1297–1304.

Endowing Existing Desktop Applications with Customizable Body Gesture-based Interfaces

Fabrizio Lamberti, *Member, IEEE*, Andrea Sanna, Gianluca Paravati, and Claudio Demartini, *Member, IEEE*
Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy

***Abstract*-- In this paper, a framework allowing to extend the applicability of natural user interaction techniques to existing programs is presented. Body gestures captured by a depth camera are mapped to application commands, and a wide set of common desktop applications can be controlled without any code rewriting.**

I. Introduction and Background

Gestures represent a key aspect of conversation between humans by enabling a direct expression of mental concepts [1] and a large number of human-computer interaction (HCI) paradigms inspired to such a communication means were developed since early eighty's [2]. Resulting interfaces are generally categorized in the class of natural user interfaces (NUI). Because of the naturalness and variety of interaction possibilities offered by this kind of interfaces, hand and body gestures were progressively introduced in many application scenarios encompassing, for instance, the inspection of virtual environments, the browsing of multimedia contents, etc. [3]

Indeed, the design of ever more pervasive gesture-based systems will play an important role in the future of the HCI. This trend is today witnessed by the mass-market diffusion of a number of consumer products integrating touch/multi-touch and depth sensor based interaction solutions. However, though touch/multi-touch interaction can be already found in a variety of appliances and software products, the development of body gesture-enabled general purpose applications (basically, different than video games) capable of exploiting off-the-shelf depth camera sensors devices are still in an embryonal phase. As a matter of fact, further stimuli to a widespread exploitation of such technologies in everyday applications will be probably provided by the evolution of standards (like those governed by the OpenNI organization) and the release of development frameworks able to ease the necessary implementation steps.

Nevertheless, even at the present time, some research prototypes supporting an easy transformation of the graphics interfaces of traditional desktop applications into gesture-aware user interfaces are already available. As a matter of example, a solution allowing users to control applications using multi-touch devices is presented in [4]. A comparable solution named Flexible Action and Articulated Skeleton Toolkit (FAAST) but tailored to body gestures is discussed in [5]. Specifically, FAAST has been designed to add gesture-based control possibilities into 3D graphics applications and video games on personal computers. FAAST interfaces with OpenNI-compliant depth cameras and extracts user's skeleton joints location/orientation. Pose information can then be used to control virtual avatars in ad hoc 3D environments. Moreover, an integrated input emulator is able to translate recognized gestures into mouse and keyboard commands to be later exploited to control existing programs.

Despite its incredible flexibility, a severe limitation of FAAST is that applications the user may want to control (like, for instance, virtual reality-based ones) have to be designed to support this kind of input. In fact, though mouse-based interaction is quite common in most desktop applications, keyboard shortcuts are not always implemented. In these cases, the FAAST toolkit alone would be of little or no help.

In this work we present a framework that, by specifically leveraging on the FAAST toolkit, aims at enabling gesture-based control of a largely extended set of existing desktop applications. The developed framework relies on a formal description of the application's graphics interface. Such a description can either be made available by developers, or extracted through operating system calls, or obtained by image processing techniques, or generated manually. Customizable mapping rules can then be defined to link a given user's gesture (or a sequence of them) to a specific operating system event to be fired on a particular graphics component of the considered interface. The designed approach allows to transparently transform the traditional interfaces of common desktop applications into true natural interfaces that can be controlled by using one of today's consumer depth camera sensors. Indeed, other APIs could also be used, e.g. those providing access to automation/accessibility features, at the cost of higher coding efforts and reduced flexibility.

II. Proposed Architecture

The overall architecture of the designed framework is shown in Fig. 1. The key role is played by the Gesture-based interface controller which, on the one side, communicates with the Application wrapper while, on the other side, interacts with the FAAST module. The Application wrapper component is responsible for handling the interface description of the application being controlled, which is obtained with the "reverse engineering-oriented" approach described in [6] (that was originally designed to migrate a desktop-based application onto a mobile device). In particular, the above component can be either used to automatically extract a detailed description of graphics elements constituting the interface, or to support users in the manual creation of such a description. In automatic mode, the Application wrapper moves the mouse (through operating system calls) over imaginary horizontal lines spanning the whole interface, hence simulating mouse-based user interaction. Interface elements react by changing their appearance or by forcing mouse pointer updates. The Application wrapper continuously grabs the graphics content of the frame buffer just before and after the update. Differences identified by performing an exclusive OR between

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

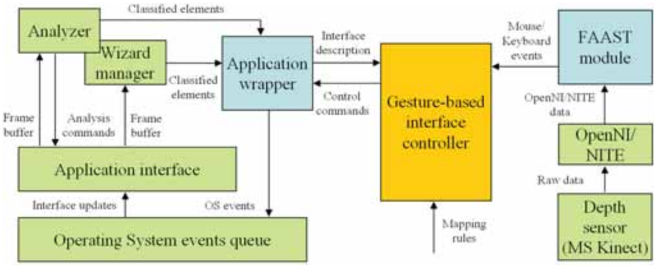


Fig. 1. Overall architecture of the proposed framework.

the captured images indicate the exact position of a the particular interface element. By means of ad hoc template matching rules, the Application wrapper is able to classify the particular element as one of the interface components supported by the designed framework (including combo boxes, menus and menu items, buttons, text boxes and text areas, sliders, check boxes, scrollbars and other custom widgets). Once the graphics elements have been classified, a description of the interface is generated and stored into the XUL (XML User Interface Language [7]) format. In manual mode, the Application wrapper provides the user with a set of wizards that let him or her manually locate interface elements and assign them to a particular class. When the user selects the particular application to interact with, the Application wrapper delivers the corresponding interface description to the Gesture-based interface controller, which allows the user to specify a mapping between gestures and control commands targeted to specific interface elements. The mapping for the given user/application is stored to be possibly re-used at a later time. Once a mapping has been defined, the Gesture-based interface controller associates emulated mouse motion and pre-configured key press and release stimuli coming from the FAAST module with specific actions to be carried out on the desktop application. Such information is delivered to the Application wrapper, which is responsible for translating it into suitable instructions to be inserted in the events queue where they will be processed by the operating system.

III. Sample application scenario and remarks

The developed framework has been evaluated by exploiting a commercial depth camera to endow some common desktop programs with configurable gesture-based interfaces. In the following, the application to a particular use case represented by a 3D viewer plug-in for web browsers will be discussed. Since the considered application natively embeds only a very limited set of keyboard shortcuts (namely, the arrow keys, which allow the user to interact with the scene only once a specific navigation mode has been selected using interface buttons), the Application wrapper component was first exploited to generate a complete description of available interface elements. Then, since some commands could be only accessed via a right-click context menu, the automatically generated description was enriched with manual annotations. Classified interface elements are highlighted in Fig. 2, while a portion of the XUL interface is reported in Fig. 3.

As a matter of example, the two leftmost elements in the lower left corner of the interface correspond to the "Walk" and "Fly" buttons. During the experimental tests, they have been mapped over the `Walk` and `Wave` gestures of the FAAST toolkit, respectively. Because of space limitations, it is not possible to report all the translation rules. Indeed, one-to-one translations are straightforward, whereas more complex mappings that can be used e.g. to access hidden application functionalities requiring, for instance, the selection of a sub-menu item, may need a further example. This is the case, for instance, of the "Speed/Faster" menu option in Fig. 2, which has been mapped over the `Crouch` gesture of FAAST and which is activated by pushing three events in the operating system queue (open the menu/sub-menu and select an item).

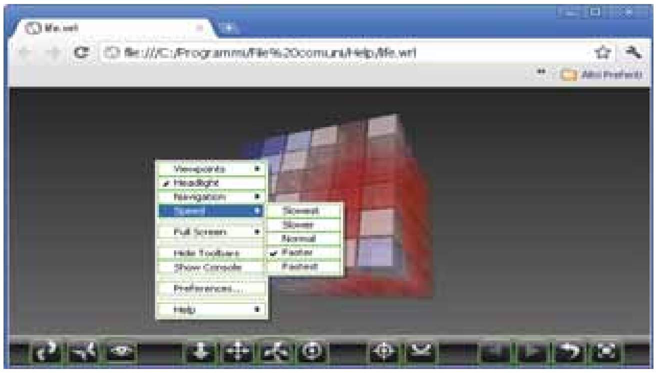


Fig. 2. Elements in the interface of the selected application.

```
<hbox id="buttons">
   <button id="walk" left="27" top="570" width="32"
    height="32" isVisible="true" icon="sel0.jpg"/>
   <button id="fly" left="64" top="570" width="32"
    height="32" isVisible="true" icon="sel1.jpg"/>
</hbox>
```

Fig. 3. Portion of the XUL-based interface description.

IV. Future works

Future works will be focused on in increasing the interaction means supported by the framework, e.g., by incorporating gaze tracking techniques, vocal stimuli, etc. Additional tests will be carried out to gather also users' feedback on system usability.

References

[1] V.I. Pavlovic, R. Sharma, and T.S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 677-695, 1997.

[2] R.A. Bolt, "Put-That-There: Voice and Gesture at the Graphics Interface," *ACM Computer Graphics*, vol. 14, no. 3, pp. 262-270, 1980.

[3] T. Selker, "Touching the Future," *Communication of the ACM*, vol. 51, pp. 14-16, 2008.

[4] G. Paravati, M. Donna Bianco, A. Sanna, and F. Lamberti, "A Multitouch Solution to Build Personalized Interfaces for the Control of Remote Applications," *2nd Int. Conf. on User-Centric Media*, 2010.

[5] E.A. Suma, B. Lange, A. Rizzo, D.M. Krum, and M. Bolas, "FAAST: The Flexible Action and Articulated Skeleton Toolkit," *IEEE Virtual Reality Conf.*, pp. 247-248, 2011.

[6] F. Lamberti, and A. Sanna, "Extensible GUIs for Remote Application Control on Mobile Devices," *IEEE Computer Graphics and Applications*, vol. 28, no. 4, pp. 50-57, 2008.

[7] V. Bullard, K.T. Smith, and M.C. Daconta, *Essential XUL Programming*. John Wiley & Sons, Chichester, England, 2001.

Non-mating Connector for USB
A Quality Waterproof Connection

Joshua S. Benjestorf, *Student Member*, *IEEE*, and Xiaoguang Liu, *Member*, *IEEE*

Abstract – **This paper presents the design of a Non-Mating Connector (NMC) based on capacitive coupling. Proof of concept USB 3.0 connectors are presented with excellent performance. NMCs hold great promise as waterproof high speed connectors.**

1. Introduction

By definition, an electrical connector is an electro-mechanical device for joining electrical circuits as an interface using a mechanical assembly [1]. This type of connector relies on ohmic contacts as the primary way to “join” the two connector parts in electrical circuits. In addition, most connectors also relay on mechanical latches in order to secure these ohmic contacts in place for optimal signal performance. In the world of consumer electronics, all connectors are understood to work this way. However, this methodology has three main disadvantages that diminish performance and life expectancy.

The first disadvantage with the typical connector is with the ohmic contacts [2]. Over time they experience contact corrosion which diminishes signal integrity and quality. Second are the mechanical latches on the connector contacts and housing. These are also prone to wear over time as they become weaker due to fatigue failure [3], the material property that measures material stress. Effectively this reduces the down force responsible for making the electrical connection. Third is exposure to ambient moisture due to exposed ohmic contacts. The primary focus of this paper is to present a new connector concept that has remedies for these disadvantages.

All three disadvantages experienced by typical connectors can be solved with a new type of connector called the Non-Mating Connector (NMC). NMCs rely on capacitive coupling for signal transfer using high-k dielectric materials. The NMC will eliminate the need for internal ohmic contacts, mechanical latches on the contacts and will create a connector impervious to ambient moisture. A high-level single NMC is illustrated in Figure 1. NMCs in their essence make waterproof connectors possible which will ultimately help advance the effort for creating waterproof consumer electronics for the consumer.

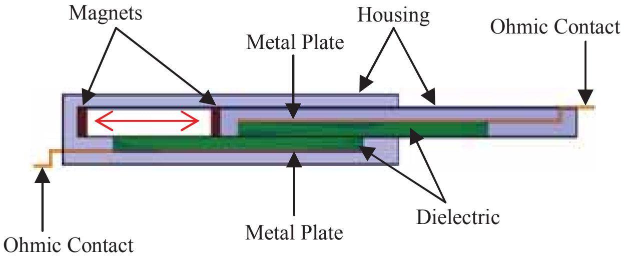


Figure 1: high-level single NMC illustration

Any connector currently on the market today can be made into an NMC. An example of such an application is the USB 3.0 which is shown in Figure 2. The figure shows the female A-receptacle and B-receptacle for the typical USB 3.0 connector [4] (a) and its NMC equivalent (b). Internally, the NMCs in Figure 2b have no mechanical latches or exposed ohmic contacts. Instead, isolated thin metallic plates are inside the NMC behind a dielectric material. Since they are internally isolated from the environment, they are not susceptible to ambient moisture like the typical USB 3.0 is in Figure 2a.

This work presents the NMC concept along with the results demonstrating proof of concept for the NMC equivalent for the USB 3.0 A-receptacle and B-receptacle connectors.

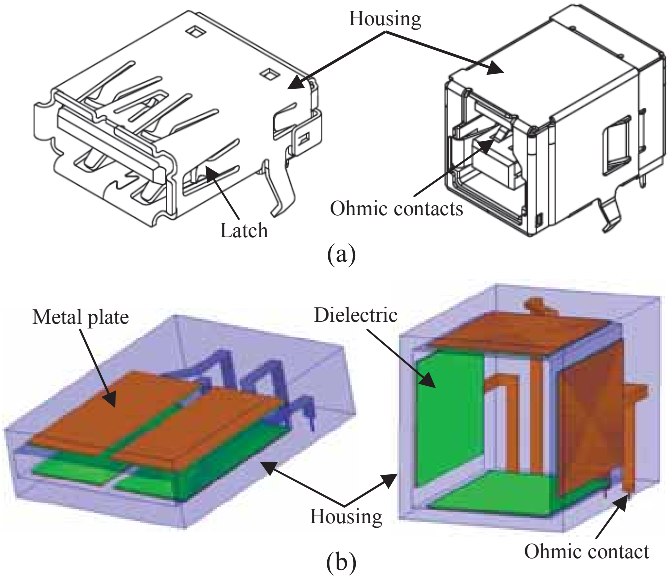


Figure 2: female A-receptacle and B-receptacle for the typical USB 3.0 connector (a) and its NMC equivalent (b)

2. Review and Overview

In order to convey the methods and reasons for the design choices used for the USB 3.0 non-mating connector, it is first necessary to review the present state of knowledge for the USB. The two main reasons behind developing the USB connector were ease of use and port expansion [5]. The original USB had only two speeds which were 12Mb/s and 1.5Mb/s later to be surpassed by the USB 2.0 that had a sample rate of 480Mb/s. The current USB operates 10 times that speed with a sample rate of 5Gb/s. This speed is absolutely essential in order to keep up with the demands for faster computational power in the consumer electronics industry.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

In order to realize a higher sample rate, the USB 3.0 uses two differential signals instead of only one as was in previous generations. Figure 3 shows the channel model for the USB 3.0. Its electrical physical layer uses mated connectors and can be used with or without a cable as shown in the figure (top figure and bottom respectively). It is important to take notice of the presence of AC capacitors which play a vital role in the operation of the USB 3.0.

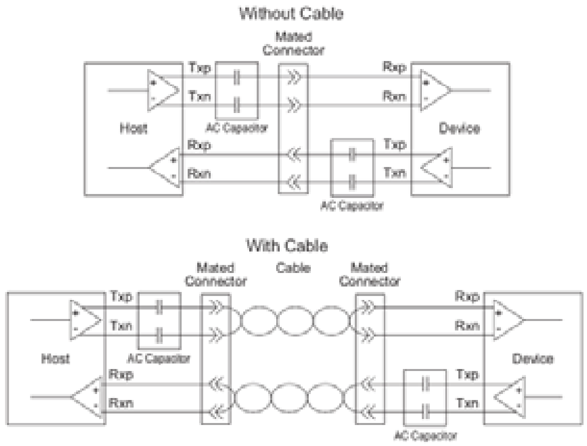


Figure 3: USB 3.0 channel models for mated connectors

These AC capacitors are placed on the transmitter side of the model (T_{xp} and T_{xn} lines). Their purpose is to couple the Host to the Device as well as block any DC offset that might exist. Thus they are essential in passing the data propagating through the USB 3.0 connector in the form of pulses.

It is important to point out that there is nothing restricting the position for where these AC capacitors are placed. The mated connectors in the channel model, for example, could themselves encapsulate the AC capacitors instead of them having to be soldered into place. Extending this idea even further, the connectors themselves could be made to function as the AC capacitors by making each one used in the channel model into an NMC like the one that is illustrated in Figure 1. This would eliminate the need for internal ohmic contacts which would be replaced by parallel plates.

In order accomplish this task the parallel plates must be made mobile instead of being fixed in position like regular capacitors. When it is said the two plates of the connector are made mobile it is meant that the connector insertion action is accomplished by sliding horizontally the two parallel plates in and out of position as depicted in Figure 1. Effectively, this would be the same as the mating or unmating in connectors. Notice the presence of the magnets in Figure 1. Their purpose is to lock into place the two parts of the NMC connector. By doing so, they eliminate the need for any mechanical latches which are extensively used in the USB 3.0 mated connector.

Since the USB 3.0 channel model accommodates two differential signals, the high-level NMC in Figure 1 must be duplicated four times; one for each of the four data lines. This gives rise to the new channel model which is called the NMC USB 3.0 channel models for non-mating connectors. This model is illustrated in Figure 4.

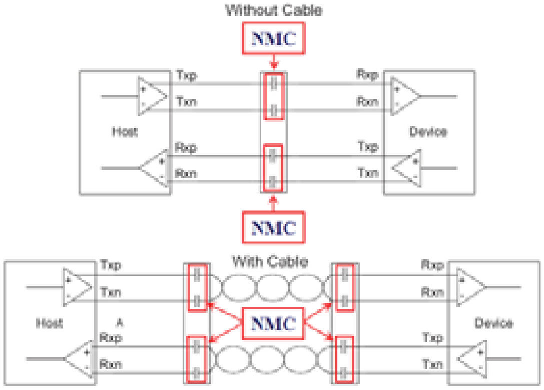


Figure 4: NMC USB 3.0 channel models for non-mating connectors

In order to ensure the same performance of the new channel model in Figure 4, the capacitance value for each NMC in it will have to be the same as the AC coupling capacitors used in the original channel model of Figure 3.

The AC capacitors only pass the incident leading and trailing edges of the digital pulses. To ensure optimal performance of data transfer, the values for the AC coupling capacitors depend on the sample rate for each application they are used for. To calculate an approximate value for an AC capacitor to an accuracy of 3%, the following equation is used [6].

$$C = \frac{7.8 * \text{Run Length} * \text{Bit Period}}{\text{R}}$$

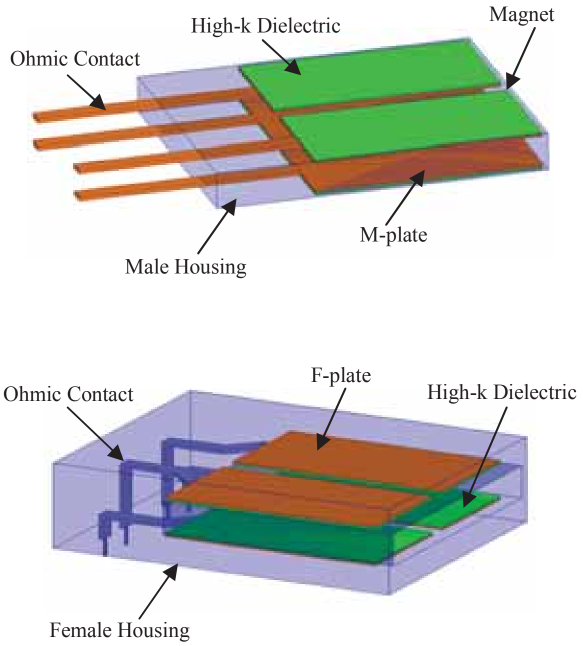
Where C is the capacitance in farads, R is resistance in ohms, Run Length (RL) in meters and Bit Period in seconds. The values for these AC capacitors (referred to as PHY capacitors) in the LVDS methodology are commonly chosen to be around 10nF [7] to 0.1uF [8]. Therefore, each NMC for the USB 3.0 must be designed with this capacitance value range in mind.

3. Development Method and Procedures

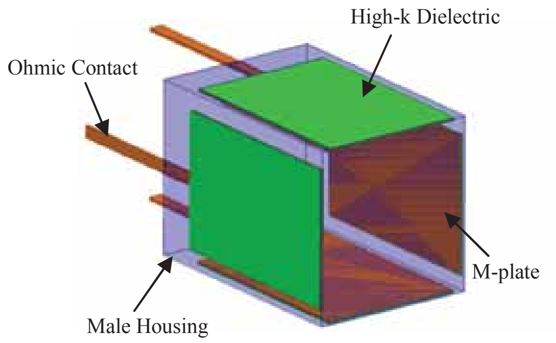
In every connector there exists a male and female receptacle and both are needed to make one connector pair. NMCs are no exception. Figure 5 shows the NMC equivalent drawings for the USB 3.0. Both male and female A-receptacle pairs (a) and B-receptacle pairs (b) are shown. One of the design requirements for the NMC equivalent of the USB 3.0 connectors was to design it with the exact same outside dimensions as the regular USB 3.0 connector receptacles. The same pin configurations for each are also maintained in order to maintain continuity with all consumer electronics that currently uses the USB 3.0. Notice the construction of each NMC for the USB 3.0 resembles very closely the high-level single NMC illustration in Figure 1. The only difference in these designs for both receptacles is they have four NMCs since there are, again, two differential pairs.

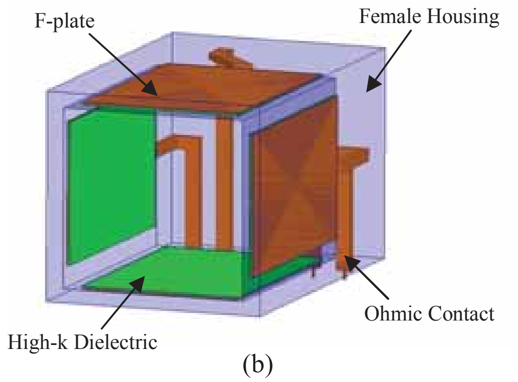
The connector housing for the designs shown in Figure 5 is made from a very good insulating material such as PPC. The M-plates and F-plates in Figure 5a and Figure 5b stand for the male and female plates respectively. Their function is the same as the two parallel plates in a capacitor and are made from very thin, highly conductive metal such as copper or gold. All plates in this design are sandwiched between a high-k dielectric and the connector housing as illustrated in the drawings. Since they are not exposed, this makes all plates impervious to the environment. Attached to each metallic plate is a trace which extends out of the housing mold of the connector.

Finally, at the center end of each connector are magnets which are placed toward the back end of each NMC. By using magnets to lock the male and female parts of the connector together, mechanical latches can be eliminated. In Figure 5, the location for each magnet was difficult to point out. For the B-receptacle that is shown in Figure 5b, they were excluded from the drawing. However, the idea of how these magnets are interconnected with the NMC USB 3.0 is conveyed in the high-level example in Figure 1.



(a)



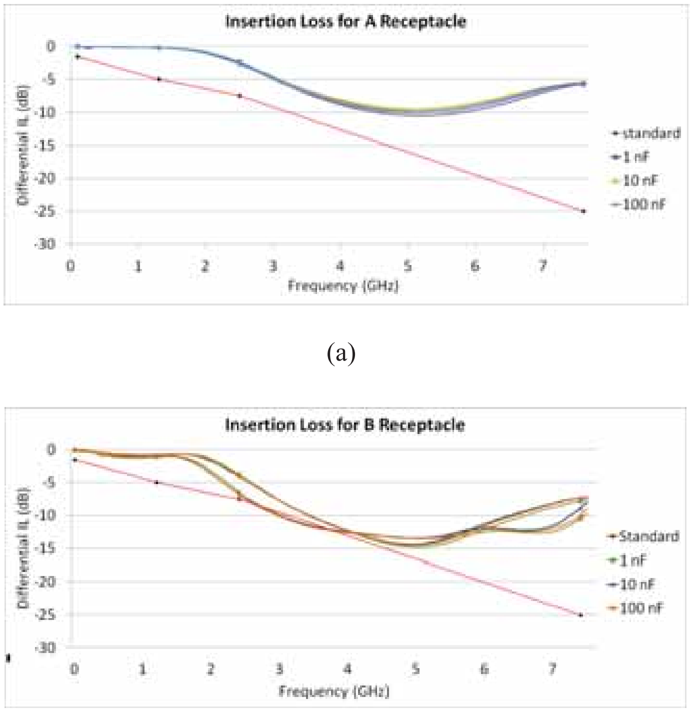


(b)

Figure 5: NMC USB 3.0 male and female A-receptacle pair (a) and B-receptacle pair (b)

4. Results

The NMC equivalent connector pairs for the USB 3.0 in Figure 5 have been simulated for differential insertion loss (IL) in HFSS. The simulation results were compared to the industry standard given in the Universal Serial Bus 3.0 specification handbook. This specification is based on measurements of differential signal energy transmitted through a mated cable assembly with loss limits which are specified by four vertices: (100 MHz, -1.5 dB), (1.25 GHz, -5.0 dB), (2.5 GHz, -7.5 dB) and (7.5 GHz, -25 dB) [9].



(a)

(b)

Figure 6: HFSS simulations for differential insertion loss for the NMC USB 3.0 A-receptacle mated pair (a) and B-receptacle mated pair (b)

Figure 6 above shows the HFSS simulation plots for both the mated A-receptacle and B-Receptacle pairs. Included in the two plots are the four vertices that define the loss limits in the specification handbook for comparison. For each NMC receptacle mated pair, insertion loss for six values of capacitance were measured: 1 pF, 10 pF, 100 pF, 1 nF, 10 nF and 100 nF. Only the last three IL measurements were plotted for the sake of clarity.

For the simulations of the A-receptacle mated pair in Figure 6a, all three IL simulations exceeded the minimum loss requirements. The vertices for the A-receptacle at the limit frequencies with a capacitance value of 10 nF were (100 MHz, -0.01 dB), (1.25 GHz, -0.23 dB), (2.5 GHz, -2.6) and (7.5 GHz, -5.5 dB). The simulations of the B-receptacle mated pair in Figure 6b also exceeded the IL loss limits with a capacitance value of 10nF. These vertices were (100 MHz, -0.05 dB), (1.25GHz, -0.97 dB), (2.5 GHz, -6.7 dB) and (7.5GHz, -8.9dB). Thus, as frequency increases the difference in performance between the two NMC mated pair receptacles decreases.

Insertion loss measurements for the same six capacitance values were taken with the Agilent E8361A network analyzer. These results are plotted in Figure 7 against the insertion loss limit vertices, again for comparison purposes. According to the measured data, the insertion loss for a 10 nF ceramic capacitor at the IL limit frequencies were (100 MHz, -0.01 dB), (1.25 GHz, -0.04 dB), (2.5 GHz, -0.08 dB) and (7.5 GHz, -0.28 dB) which far exceeds the IL loss limits.

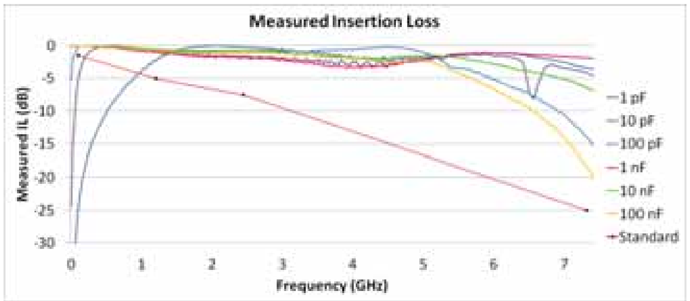


Figure 7: insertion loss measurements with Agilent E8361A network analyzer for six different capacitance values

5. Conclusion

The work in this paper has introduced the concept, methodology and has shown proof of concept for the non-mating connector. The concept used for the NMC uses capacitive coupling and has the potential for unlimited applications. This work also introduced the first application for the NMC which was for the USB 3.0. Both A-receptacles and B-receptacles have been designed, tested and shown to perform far above industry standards according to the Universal Serial Bus 3.0 specifications. The life expectancy of the NMC is expected to be much greater than the typical connector due to the elimination of mechanical contacts by using magnets as the primary way of locking the NMC in place. An increase in signal performance is also realized by eliminating contact corrosion by eliminating the need for internal ohmic contacts found in typical connectors as the method of interfacing the two connector parts. The NMC equivalent for the USB 3.0 serves two functions; one being the connector itself and the other as the ac coupling capacitors. Results for both simulated mated pairs and measurements for IL of several capacitor values verified the optimal capacitance value for optimal NMC performance to be between 10 nF to 100 nF for the USB 3.0 application.

The vision for the NMC is to bring consumer electronics one step closer to becoming waterproof. Up until now this has not been possible until connectors are made waterproof. The NMC will make this possible.

REFERENCES

[1] Robert S. Mroczkowski, *Electrical Connector Handbook Theory and Applications*, McGraw Hill, 1998.

[2] M. Antler, *Electrical effects of fretting connect or contact materials: A review*, AT&T Bell Laboratories, Columbus OH, OH 43213 USA.

[3] Beer, Ferdinand P.; E. Russell Johnston, Jr., *Mechanics of Materials* 2nd ed., McGraw-Hill, Inc., 1992, pp. 51.

[4] Universal Serial Bus 3.0 Specification (including errata and ECNs through May 1, 2011), Revision 1.0, Hewlett-Packard Company, Intel Corp., Microsoft Corporation, NEC Corporation, ST-Ericsson, Texas Instruments, pp. 5-21, 2011.

[5] Universal Serial Bus 3.0 Specification (including errata and ECNs through May 1, 2011), Revision 1.0, Hewlett-Packard Company, Intel Corp., Microsoft Corporation, NEC Corporation, ST-Ericsson, Texas Instruments, pp. 1-1, 2011.

[6] LVDS Owner's Manual Including High-Speed CML and Signal Conditioning, 4th ed., 2008, p. 34-35.

[7] Kal Mustafa and Chris Sterzik, *AC-Coupling Between Differential LVPECL, LVDS, HSTL, and CML*, Texas Instruments, Application Report SCAA059C – March 2003 – Revised October 2007, pp. 2.

[8] John Guy, *Tap AC-Coupling for LVDS Signals*, Maximum Integrated Products Inc.

[9] Universal Serial Bus 3.0 Specification (including errata and ECNs through May 1, 2011), Revision 1.0, Hewlett-Packard Company, Intel Corp., Microsoft Corporation, NEC Corporation, ST-Ericsson, Texas Instruments, pp. 5-49, 2011.

Generation of Efficient Bitstreams for Functional Tests of Video Decoders

Soonwoo Choi, Seunggyu Jeoung, Jicheon Kim and Soo-Ik Chae
Department of Electrical Engineering and Computer Science
Seoul National University, Seoul, Korea

Abstract-- **We propose a new method to generate bitstreams for efficient functional tests of video decoders. The proposed method employs a 3-way covering combinatorial method in finding efficient bitstreams for high-level syntax elements (SEs) while using a constrained-random method in generating bitstreams for low-level SEs. Furthermore, it fills up the coverage holes that are found after analyzing the SE coverage. The bitstreams generated with the proposed method are three times more efficient than those with the existing methods.**

I. Introduction

Developing video decoders for embedded applications requires much time and effort in eliminating design errors. In an early design stage of video decoders, we need to focus on functional tests instead of conformance tests. The H.264 conformance test suite [1] consists of 204 test bitstreams, which includes totally about 15 million macroblocks (MBs). This test suite is not suitable for debugging because it takes more than a year for its RTL simulation assuming that it takes about five hours for a 1080p picture. Furthermore, it is almost impossible to test all possible high-level configurations within a reasonable time budget. For example, a sequence parameter set (SPS) and a picture parameter set (PPS) in H.264, which determine the main high-level features of the H.264 standard, include 15 and 10 binary flags respectively. More than 33 million configurations are required to test all these possible configurations. Therefore, it is necessary to develop a set of efficient test bitstreams that can test the video decoders with RTL simulation in a reasonable time and cover most of these configurations.

In this paper, we propose a new method to generate efficient test bitstreams, which has three unique features. First, we try to reduce the redundancy in the bitstreams for both high-level and low-level SEs in order to reduce the bitstream length as much as possible. The proposed method employs a combinatorial method [2] in finding efficient bitstreams for high-level SEs while it uses a constrained-random (CR) method in generating bitstreams for low-level SEs. Second, we sort the bitstreams in an efficient order so that higher coverage can be obtained as early as possible. Third, we improve the test bitstream set by finding and removing coverage holes in the bitstream set. Consequently, the resulting shorter bitstreams can increase substantially the coverage while detecting design errors in a short time.

The rest of the paper is organized as follows: Section II describes the proposed method that constructs the efficient test bitstream set. Section III includes the experimental results, which is followed by the conclusion in Section IV.

II. Efficient bitstream generation

In this section, we describe how to generate efficient test bitstreams as shown in Fig. 1. The proposed method first selects the high-level features of the video standard with profiles and levels. After that, configurations for the selected features are determined with an n-way covering combinatorial method. Then, we select the size and number of pictures of a video sequence for each configuration. It is also necessary to determine low-level SE values as well as unselected high-level SE values. After generating test bitstreams, it finds coverage holes and modifies configurations, if needed, to generate additional bitstreams to fill the coverage holes.

For a video standard, the generated bitstreams should cover all the important combinations of the selected features that depend on specific tools. Therefore, we determine a set of configurations satisfying high cross coverage of the SEs with an n-way covering combinatorial method [2] for the selected features. Here, we employed a 3-way covering simply because it provides good enough test coverage and most of important cases are affected by less than or equal to three high-level SEs.

In generating some bitstreams, we selected pictures of 5x4 MB array because they include all possible cases that the current MB and the neighbor MBs are positioned at 9 different regions where each MB in the same region has equal

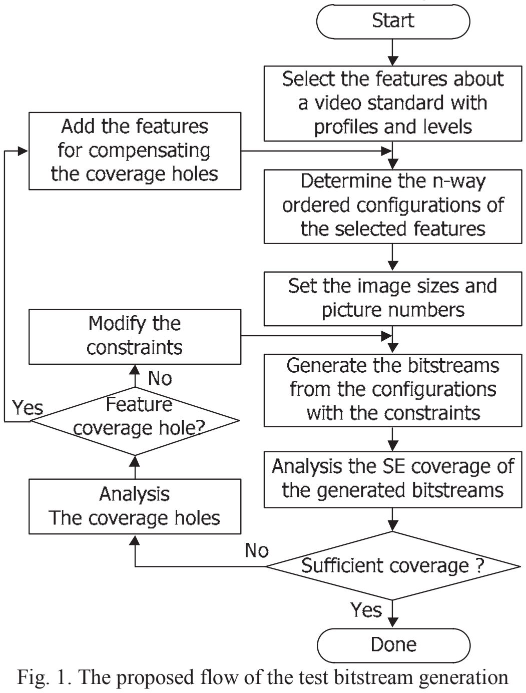

Start
Select the features about a video standard with profiles and levels
Add the features for compensating the coverage holes
Determine the n-way ordered configurations of the selected features
Set the image sizes and picture numbers
Modify the constraints
No
Yes
Feature coverage hole?
Generate the bitstreams from the configurations with the constraints
Analysis the SE coverage of the generated bitstreams
Analysis The coverage holes
No
Sufficient coverage ?
Yes
Done

Fig. 1. The proposed flow of the test bitstream generation

This work was supported by Samsung Electronics Ltd., Inter-university Semiconductor Research Center in Seoul National University, and BK21.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

availability of its neighbor MBs. The number of pictures in a sequence should be determined by the number of required configurations and the number of MBs required for obtaining high enough coverage of low-level SEs. To cover the dependency of line buffers, we employed pictures with maximal MB width and 4 MB height. We also selected a larger number of pictures with small sizes such as one MB or 2x2 MBs to cover the test cases for the features related to the picture order count (POC) or the group of picture (GOP).

The values of SEs that are not determined in a configuration are randomly selected among the values that satisfy their constraints. It is practically impossible to cover all the possible spatial and temporal combinations for MB-level SEs. Therefore, the CR method [3] without using input picture sequences was employed for the MB-level SEs to overcome their limitation of selecting the values of SEs due to their spatial and temporal correlations.

In this paper, we adopted the average coverage of SEs as well as the code coverage. Note that we used subsampling for some SEs with a wide range because it is practically impossible to reasonably cover all their ranges. In analyzing the SE coverage of the generated bitstreams, we tried to find two different kinds of coverage holes. One type of coverage holes corresponds to a region that is not covered due to missing features. In this case, we should add one or more additional bitstreams to cover those features. The other type corresponds to a region that is not covered due to over-constraint. In this case, we should relax the constraints for its SEs to include the uncovered test cases.

III. Experimental Result

For MPEG-4 simple and advanced simple profiles, we selected seven features including binary flags for short header, data partitioning, reversible VLC, interlace, quant type, sprite and quarter sample. For H.264/AVC baseline and high profiles, we similarly selected six features including five binary flags for frame, MBAFF, CABAC, transform 8x8, spatial direct, and a ternary flag for POC type, of which value can be 0, 1 and 2. We constructed test sets by using ACTS tool for the selected features and generated a test bitstream for each configuration.

We generated a set of 17 bitstreams with the proposed method, which were compared with another set of 17 bitstreams obtained with the FFmpeg encoder for 32 pictures of 5x4 MB sizes for the MPEG-4 standard. The bitstreams from the proposed method achieve about 90 % of the average SE coverage while those from the FFmpeg encoder reaches to about 40%.

Similarly, we also compared SE coverage of the bitstreams from the proposed methods and the CR methods together with the conformance bitstreams for the H.264 standard, which are summarized in Table I. Here, the bitstream set for H.264.1 [1] is limited to a subset that includes 179 bitstreams only for 8-bit 4:2:0 YUV format. The average SE coverage for the proposed method's bitstreams is substantially higher than that of the subset of H.264.1, although the length of the bitstream set from the proposed method is only about 0.27 % of that of the H.264.1. For comparison, we also generated another set of bitstreams from the CR method that reaches to about the same level of the average SE coverage of the bitstreams from the proposed method. We found that the length of the test bitstream set from the proposed method is a little less than a third of that from the CR method.

Table I. Comparison of SE coverage among the H.264.1 conformance test suite, the CR bitstream set and the proposed bitstream set

Level	H.264.1 [1] (179 ea)	CR [3] (75 ea)	Proposed (20 ea)
SPS	75.5 %	81.6 %	81.4 %
PPS	70.6 %	92.5 %	92.5 %
SLH	76.0 %	89.2 %	89.2 %
MB	91.3 %	98.0 %	97.4 %
Average	78.4 %	90.2 %	90.1 %
# of MBs	≈ 6.35 x 10^6	52416	17216
Coverage/MB (Norm)	0.00001 (1 x)	0.00172 (139 x)	0.00524 (424 x)

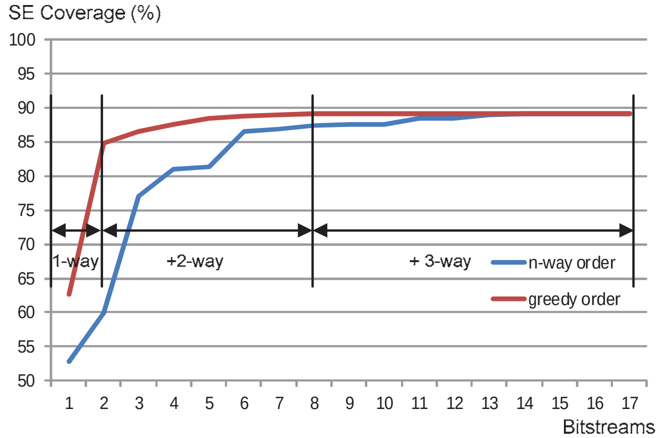


Fig. 2. Comparison of the average SE coverage of the n-way convering's order and the greedy algrorithm's order for the MPEG-4 standard

We confirmed that the bitstream set sorted in the n-way covering's order is almost as good as that in the order sorted with the greedy algorithm [4], which is shown in Fig. 2 for the MPEG-4 standard.

IV. Conclusion

In this paper, we described a new method to generate efficient test bitstreams for the functional test of the video decoders. With experimental results for MPEG-4 and H.264/AVC standards, we confirmed that the test bitstream sets of the proposed method have substantially higher efficiency than that of the CR method [3].

References

[1] ITU-T Rec. H.264.1, "Conformance specification for ITU-T H.264 advanced video coding", April, 2010.

[2] R. Kuhn, Y. Lei, and R. Kacker, "Practical combinatorial testing: beyond pairwise", IT Professional, vol. 10, no. 3, pp.19-23, May. 2008.

[3] J. Cho, S. Choi, and S.I. Chae, "Constrained-Random Bitstream Generation for H.264/AVC Decoder Conformance Test", IEEE Transactions on Consumer Electronics, vol.56, pp.848-855, May. 2010.

[4] J. Kim, S. Choi, and S.I. Chae, "Efficient Test Bitstream Set Selection Algorithm for Verifying H.264 Decoder", IEEK Fall Conference 2011, pp.151-152, November, 2011.

Compression Friendly Medical Image Encryption based Order Relation

Ganzorig Gankhuyag, Soongi Hong, and Yoonsik Choe

Abstract—**In this paper a novel encryption methodology for improving compression in medical images is presented. It is shown that with the proposed scheme the pixel discontinuities are reduced which reduced the approximation error during the compression process. Permutation re-ordering is used to obtain the encryption key. The simulation results show that the proposed scheme can achieve 7-15dB improvement in PSNR values when used with JPEG2000 compression standard. It is shown that the proposed encryption scheme is independent of compression method used and can achieve real-time performs for encryption and decryption process.**

I. INTRODUCTION

With significant increase in security threats the role of encryption has increased significantly in medical image processing domain also. Conventionally, the problems associated with encryption and compressions are dealt independently. Hence, most of encryption algorithms like Advanced Encryption Standard (AES), Data Encryption Standard (DES) plan for improvement in compression efficiency.

The concept of pixel ordering was introduced to improve compression efficiency in [1]. Various encryption methods have been introduced for medical images [2]-[5]. The principal disadvantages of these encryption schemes have been their requirement for modification of inherent encoder or decoder used for compression. Additionally some of these schemes are limited for RAW images.

The basic idea of order relation is decreasing discontinuity of pixel values in image. Reduction in discontinuity results in reduction of approximation error originating from compression applied in spatial domain. Image consists of co-related pixel values. In various transform based lossy compression methods the performance of compression is determined by the extent of approximation done during the process of compression.. It has been shown that increase in discontinuities in pixel values increases approximation error. This in turn reduces the compression efficiency. Hence, reduction in pixel value discontinuities will reduce approximation error and will thereby increase compression efficiency.

In this paper we recommend novel order relation encryption algorithm that bound up with compression. Proposed encryption algorithm is compression friendly which decreasing discontinuity from image. Remaining part of this paper is following Section 2 introduces order relation encryption algorithm. In the Section 3 simulation results are discussed. In the Section 4 we will analyze security strength. We make conclude in Section 5.

II. ORDER RELATED ENCRYPTION

From information theory: Let A and B be two sets of natural values. The product set of A and B is the set of all ordered pairs (a, b) such that $a \in A$ and $b \in B$, as in (1).

$$A \times B = \{(a, b): a \in A, b \in B\} \tag{1}$$

For a given image, let P be a pixel and the cardinalities of P is Q. The relation of I from P, Q set such as (p, q), as in (2). The P is the set of pixel and Q is the set of pixel intensity.

$$I = \{(p, q): p \in P, q \in Q\} \tag{2}$$

In other words, for each pixel processed row-wise, the pixel values are stored row-wise in set P while their frequency of occurrences is stored row-wise in set Q. For an 8 bit grayscale image, the image values range from 0 to 255. In the proposed scheme, the frequencies of various pixel values are combined to form I sets. The range of pixel values for each I set and hence, the number of I sets per image is decided by the user key. The boundaries of I set are determined using Lloyd-Max method. Each column is processed independently and hence, the number of I sets and their boundary vary for each column. The ordered sets of I are permuted and re-ordered again to reduce the pixel correlation. Once ordered, the un-correlated behavior of I sets provides a form of encryption. The permutation of I sets also reduces the discontinuities between various I sets. Fig. 1.a and 1.b shows, the original grayscale medical image and the order related encrypted image respectively.

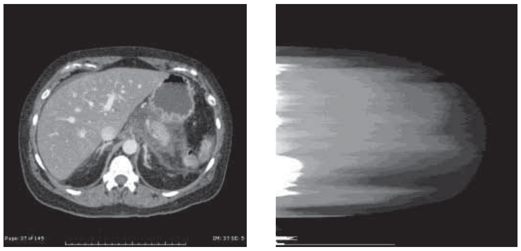


Fig. 1(a) original image, (b) order related encrypted image

III. SIMULATION

The proposed scheme was implemented as an extension to JPEG2000 compression. The proposed scheme is inserted as pre-processing and post-processing block as shown in Fig. 2. Using the simulation environment shown in Fig. 2, 5 grayscale medical images were processed. After perform in the proposed order related permutation the JPEG2000 compression was performed. For each image compression ratio of 5, 10 and 20 was used. The proposed design was simulated using Intel i5 processer (3.3 GHz) with 4 Gb RAM.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

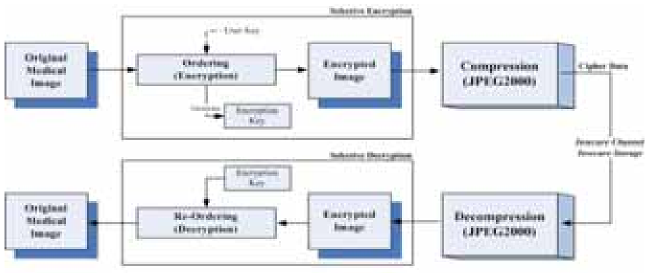


Fig. 2 Proposed Algorithm simulation scheme

Using the user key and the ordering information the encryption key is generated. After image decompression, the decoder uses the encryption key to obtain the original medical image.

Table I shows the comparison of medical images compressed using JPEG2000 technique with and without the use of proposed encryption scheme. The value of user key was set to 120. .

TABLE I
COMPARE PSNR BETWEEN PROPOSED AND NOT ENCRYPTED

		JPEG2000 COMPRESSION RATIO: PSNR(dB)		
TEST MEDICAL IMAGE		CO-RATIO 5	CO-RATIO 10	CO-RATIO 20
IMAGE – A	ORIGINAL	52.14	45.06	40.77
	PROPOSED	INFINITY	64.33	56.88
IMAGE – B	ORIGINAL	49.89	42.71	37.26
	PROPOSED	67.90	56.85	53.54
IMAGE – C	ORIGINAL	42.74	37.56	34.42
	PROPOSED	62.52	55.39	51.80
IMAGE – D	ORIGINAL	49.18	40.64	35.35
	PROPOSED	INFINITY	61.63	56.83
IMAGE – E	ORIGINAL	35.80	31.79	29.41
	PROPOSED	57.93	53.59	49.91

From the results it can be seen that for given set of simulation conditions the use of proposed encryption scheme improves the PSNR by more than 15dB. It should be noted that this improvement is dependent on the boundary values which in-turn depends on user key.

Table II shows the time required for the encryption and decryption in the proposed scheme. User keys values of 40, 80, 120 and 160 were used for simulation. It is evident that the timing values are dependent of the user key values.

TABLE II
SPEED TEST SIMULATION RESULT

IMAGE	SIZE	AVERAGE ENCRYPTION TIME(SEC)	AVERAGE DECRYPTION TIME(SEC)
IMAGE – A	512×512	0.13	0.05
IMAGE – B	480x480	0.14	0.05
IMAGE – C	600x432	0.20	0.07
IMAGE – D	600x560	0.15	0.06
IMAGE – E	224x600	0.16	0.06

IV. SECURITY ANALYSIS

The choice of user key is determined by the factors such as encryption strength, speed of computation and compression efficiency desired. As mentioned earlier, P and Q sets are obtained from user key and the encryption key is obtained from permutation ordering of P and Q sets. The value chosen for encryption is based on the number of Q set. The number of Q sets varies for each row. For $m \times n$ pixel image, the total numbers of Q sets are given by

$$\text{Total number of Q sets} = \sum_{i=1}^{m}\sum_{j=1}^{n} Q \qquad (3)$$

Let b bits be used for representing total number of Q sets. Hence, the possible value of encryption key range from 0 to 2^b. Since, the value of Q is dependent on image, it is not possible to obtain it from any random image. As the image size increases, or the user key is changed the number of Q sets increases significantly. Therefore, the number of possible values of encryption key also increases significantly. This factor is extremely critical against the brute force attack in decryption. In other words, as the image size increases the un-authorized decryption will become more complicated for proposed scheme.

V. CONCLUSION

In this paper compression friendly encryption methodology for medical images is presented. The permutation ordering removes the position information to achieve encryption. Using pixel re-ordering the discontinuities in image is also reduced. This improves the compression efficiency. The simulation results shows that with the use of proposed encryption method, the PSNR value for the compressed image improves by more than 7 to 15 dB for JPEG2000 compression method. The proposed encryption is independent of compression methodology and can be achieve real time processing speed with any existing compression standard.

ACKNOWLEDGMENT

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the “ITRC(Information Technology Research Center)” support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2012-H0301-12-1008)

REFERENCE

[1] Soongi Hong, Minyoung Eom and Yoonsik Choe, “Variable-length code based on an order complexity,” *Picture Coding Symposium*, 2009.

[2] A.Mitra, Y. V. Subba Rao, S. R. M. Prasanna, “A New Image Encryption Approach using Combinational Permutation Techniques,” *International Journal of Electrical and Computer Engineering*, 1:2 2006.

[3] Yicong Zhou, Karen Panetta, “A Lossless Encryption Method for Medical Images Using Edge Maps,” *International Conference of the IEEE EMBS*, 3-6 Sept, 2009.

[4] Zahia Brahimi, Hamid Bessalah, “A new selective encryption technique of JPEG2000 codestream for medical images transmission,” *Systems, Signals and Devices*, 2008.

[5] K. Usman, H. Juzoji, “Medical Image Encryption Based on Pixel Arrangement and Random Permutation for Transmission Security,” *e-Health Networking, Application and Services*, 2007.

Address Generation for Lossless Frame Memory Compression in an H.264/AVC Encoder

Hyun Kim, Chae Eun Rhee, and Hyuk-Jae Lee
Inter-university Semiconductor Research Center,
Department of EECS, Seoul National University, SEOUL, KOREA
Email: {snusbkh0, chae, hyuk_jae_lee}@capp.snu.ac.kr

***Abstract*-- An inter-prediction in an H.264/AVC encoder requires a large amount of access to the external memory. Thus, the power consumption and encoding latency caused by the external data transactions are significantly high. One approach to reduce the external data transactions is the frame memory compression (FMC). For lossless FMC, the compressed data size of each macroblock depends on the contents of the macroblock, and consequently, an address table is required to locate the start address of the compressed data of each macroblock. In this paper, a novel address calculation scheme is proposed for lossless FMC. Due to regular memory allocation, the memory addresses of all macroblocks in a frame are calculated using 7 additions. For 1280x720 videos, 2KB of internal memory is enough to store all information to calculate the memory address.**

I. INTRODUCTION

An increasing demand of the high resolution videos makes it important to compress videos for transmission and storage. The H.264/AVC standard is widely used because of high compression efficiency. However, a large amount of computation is required and also the power consumption is considerable. In particular, for an inter-prediction which shows the high coding efficiency, the reconstructed frame is stored and used as a reference frame in the encoding for the next frame [1]. The size of frame data is too big to be stored in the internal memory. Thus, the frame data is generally placed in the external memory. The data transactions between the encoder and the external memory cause large power consumption and the encoding delay. The portion of the external power consumption in the total amount of power consumed by the system is significant. One approach to reduce the external data transactions is the frame memory compression (FMC) which compresses the frame data to be stored in the external memory. A number of FMC algorithms have been proposed and, among them, an algorithm with discrete wavelet transform (DWT) followed by set-partitioning in hierarchical tree (SPHIT) coding is one of the best algorithms [2]. This paper adopts this algorithm with a 16x16 block size for the lossless compression as shown in Fig. 1. In the lossless FMC, the compressed data size of each 16x16 macroblock (MB) is different. Thus, for random accessibility, an address table is required where the memory address is stored for the corresponding MB. However, the straightforward address table is too large to be stored in the internal memory. If the address table is located in the external memory, the amount of the external memory access increases. This paper proposes a novel address calculation scheme for the lossless FMC. Seven types of the memory size are allocated to store the compressed MB data. Based on these predetermined memory allocation sizes, the memory address is calculated using 7 additions. For high definition (HD) videos, 2 Kilobyte (KB) of internal memory is enough to store all information to calculate the memory address. The rest of this paper is organized as follows. Section II describes the proposed address table design. In Section III, simulation results are shown and conclusions are presented in Section VI.

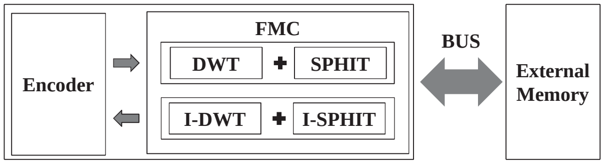


Fig. 1. Hardware Platform

II. ADDRESS TABLE DESIGN

A. Regular Memory Allocation

Table I shows the method to store the compressed MB data to the external memory. The first row represents the size of the compressed MB data. The size of uncompressed MB data is 256 bytes in Y component. Thus, the size of compressed MB data ranges from 0 to 256 bytes. In the ninth column, the size of compressed MB data is sometimes over 256 bytes because a lossless FMC does not guarantee the reduction of data size. The second row represents the allocated memory size for the data which have data size in the first row. When the size of the compressed data is over 256 bytes, the MB data is stored to the external memory without compression. The type in the third row is to identify the memory size allocated to each MB and 3 bits are used to represent 8 types of memory size. This regular memory allocation lowers the utilization of memory space. However, in general, the extreme saving in the external memory is not critical. In terms of the external memory bandwidth, the increase in bandwidth is small because only the valid data is stored and loaded from the external memory.

TABLE I
PREDETERMINED SIZES FOR MEMORY ALLOCATION

Compressed MB Data Size(byte)	0 - 64	65 - 96	97 - 128	129 - 160	161 - 196	197 - 224	225 - 256	256+
Allocated Memory Size(byte)	64	96	128	160	196	224	256	256
Type	0	1	2	3	4	5	6	7

This work was supported by "Samsung Electronics Semiconductor Business".

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. *Address calculation*

For the motion estimation in an H.264/AVC encoder, the search range data need to be loaded from the external memory. Thus, the memory address of arbitrary MBs should be calculated in real-time. If all MBs have their types in Table 1, the memory address of a particular MB is calculated by adding the allocated memory size of previous MBs from the starting address of the current frame. However, it takes a significant amount of time to calculate the address to access the latter MBs in the frame. Thus, the encoding time also increases due to the long address calculation time. To speed up the address calculation, two auxiliary tables are designed. In one table, the memory address for the leftmost MB in each MB line is stored while the compressed MB data is written in the external memory. Now, the address calculation starts not from the address of the first MB in the frame but from the address of the leftmost MB. However, the calculation time is still long because the maximum number of addition operation is proportional to the width of videos. To further reduce the calculation time, pre-calculated multiplication results are stored in the other table utilizing the regularity of the allocated bit size as shown in Table II. The first column represents the number of additions. In case of HD (1280x720) sequences, one MB line consists of 80 MBs. Thus, maximum 79 additions are pre-calculated. From the second to the eighth columns, the multiplication results are shown according to the allocated memory size. This multiplication table is used as follows. To access the Nth MB in the Mth MB line, the number of the occurrences of each type is counted from the 1st to (N-1)th MBs in the Mth MB line. For seven types, the value which is the allocated memory size multiplied by the number of occurrence is obtained from Table II. The sum of the seven values obtained from Table II and the address of the leftmost MB of the Mth MB line is the memory address for the Nth MB in the Mth MB line. In real applications, the multiplication table in Table II is downsized by reusing the overlapping multiplication results.

TABLE II
MULTIPLICATION TABLE

	Allocated Memory Size (byte)						
Numbers	64	96	128	160	196	224	256
0	0	0	0	0	0	0	0
1	64	96	128	160	196	224	256
2	128	192	256	320	392	448	512
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
79	5056	7584	10112	12640	15484	17696	20224

III. SIMULATION RESULTS

In this section, the change of an external memory size, a bandwidth and the internal buffer size are shown when the lossless FMC with the proposed address calculation scheme is applied to an H.264/AVC encoder. For simulation, a reference software model which gives exactly the same results with a hardware-based encoder is used. An encoding configuration is set to base profile [1]. One hundred frames of ten HD test sequences are used for the simulation and the quantization parameter values vary from 20, 24, 28 to 32.

TABEL III shows the external memory size, the external data bandwidth and the internal buffer size for the frame data. The external data bandwidth for storing the reference frame to the external memory is only considered. The second column represents simulation results without the FMC, whereas the third column represents the simulation results with FMC. Here, the compressed data is stored contiguously without empty space [3]. The simulation results with the proposed scheme in Section II.A are shown in the fourth column. For the external memory size, the FMC in the third column and the proposed scheme in the fourth column save 53.97% and 45.34% of memory space, respectively. The memory saving is sacrificed by 8.63%. However, the difference of the external data bandwidth between the FMC in the third column and the proposed scheme is marginal. In the third column, the internal buffer size to store the straightforward address table is about 9KB which is calculated as 9000 (=20x(1280x720)/(16x16)/8) bytes for HD video sequences [3]. An internal memory size for the proposed address calculation scheme is analyzed as follows. First, 1350 (=3x(1280x720)/(16x16)/8) bytes are required to represent the type of each MB. Second, 101.25 (=18x720/16/8) bytes are required to store the memory address of the leftmost MB in each MB line. Lastly, 587.5 bytes are necessary for the multiplication table. In total, about 2KB of internal memory is necessary for the proposed scheme. Compared to the straightforward address table, the proposed address calculation scheme achieves 77% of internal memory size saving.

TABLE III
COMPARISON OF THE EXTERNAL MEMORY SIZE, EXTERNAL MEMORY BANDWIDTH AND INTERNAL BUFFER SIZE

KB per one frame	Without FMC	FMC with the straightforward address table	FMC with the proposed address calculation scheme
External memory size	900	414.24	491.95
External memory bandwidth	900	414.24	420.41
Internal buffer size	0	9	2

IV. CONCLUSION

A large amount of data transaction between the encoder and external memory increases the power consumption and the encoding latency. When the lossless FMC is used, the external data transaction decreases significantly. To support the random accessibility in the FMC, a novel address calculation scheme is proposed with a small hardware overhead. The proposed scheme is fast and thus applicable for real time encoders.

EXAMPLES OF REFERENCE STYLES

[1] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] Yongseok Jin and Hyuk-Jae Lee, "A Block-based Pass-Parallel SPIHT Algorithm", IEEE Transactions on Circuits and Systems for Video Technology, 2012

[3] S. Lee, M. Chung, S. Park, and C. Kyung, “Lossless frame memory recompression for video codec preserving random accessibility of coding unit,” IEEE Trans. Consumer Electro., vol. 55, no. 4, Nov. 2009.

A 0.67nJ/S Time-domain Temperature Sensor for Low Power On-chip Thermal Management

Young-Jae An, Kyungho Ryu, Dong Hoon Jung, Seung-Han Woo,
and Seong-Ook Jung, *Senior Member, IEEE*
VLSI System Laboratory, School of Electrical & Electronic Engineering, Yonsei University, Korea

***Abstract*—This paper presents a time-domain temperature sensor with process variation tolerance for low power on-chip thermal management. To achieve a suitable on-chip composition, the proposed sensor uses the externally applied code to save the area- and power-consumption. The proposed sensor is implemented using a 0.13μm CMOS process technology and its core area is 0.031mm^2. The measurement results show that the energy per conversion rate is 0.67nJ/S at 1.2V supply voltage, conversion rate is 430k samples/sec, and sensing error is -0.63 ~ +1.04°C with 2nd order master curve and one-point calibration over the temperature range of 20 ~ 120°C.**

I. Introduction

Recently, portable devices require more various functions with higher performance. To meet these requirements, application processors (APs) for the portable device have been developed [1], [2] and even quad-core processors are used for these APs in recent [3]. On the other hand, since the required performance and functions for APs increase, the heat problem of APs becomes getting worse. In the high temperature, the circuit performance and reliability are getting worse; hence, the temperature sensor for on-chip thermal management becomes important. Also since the high performance AP can cause the temperature to change dynamically and the portable device has the limited power source, a high conversion rate with low energy consumption per conversion is required for the temperature sensor.

This paper presents the time-domain temperature sensor for low power on-chip thermal management. To achieve a suitable low power on-chip composition, the proposed sensor uses the code difference between the temperature-dependent code and externally applied reference code instead of the reference generator [4] which is the area- and power-consuming block. Consequently, the proposed sensor achieves a high conversion rate of 430k samples/sec and a high energy efficiency of 0.67nJ/S

II. Proposed Sensor Design

The proposed temperature sensor is designed with a variable ring oscillator (VRO), a time-to-digital converting block (counter), and controller blocks, as shown in Fig. 1. The proposed sensor senses the temperature using the code difference between the temperature-dependent binary code (*Temp_code*), which is quantized from the output of VRO (*VRO_out*), and the externally applied temperature-independent binary reference code (*Ref_code*). However, since *VRO_out* is changed with not only temperature but also process variation, the calibration for process variation is performed before the sensing.

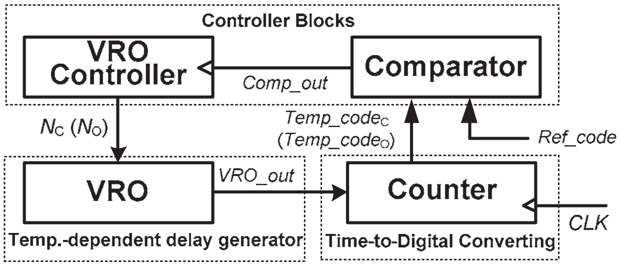


Fig. 1. The schematic of proposed temperature sensor

The proposed temperature sensor operates with two procedures: calibration and sensing. The following subsections describe the operation procedures and blocks of the proposed sensor in detail.

A. Calibration Procedure

In the calibration procedure, the proposed sensor performs the calibration to remove the effect of process variation in *Temp_code* for improving the sensing accuracy. Also, since the complexity of calibration increases the high-volume production cost, the proposed sensor uses the one-point calibration instead of the multi-point calibration.

When NMOS driving strength is equal to PMOS, *VRO_out* can be modeled as the product of the number of activated delay cells in VRO, a temperature term and a process term using [4]. The *VRO_out* is quantized to *Temp_code* by the counter and *Temp_code* can be expressed the following equation.

$$Temp\_code = N \cdot T^{-\alpha} \cdot P \tag{1}$$

where N is the number of activated delay cells, T is temperature, α is the temperature coefficient which depends on the process technology, and P is process-only-dependent term. The comparator generates *Comp_out* when the *Temp_code* and *Ref_code* are different. Then, VRO controller adjusts N until *Temp_code* is the same as *Ref_code*. Then, the calibrated *Temp_code* (*Temp_code*$_C$) can be written as follows,

$$Temp\_code_C = Ref\_code = N_C \cdot T_C^{-\alpha} \cdot P \tag{2}$$

where T_C is the calibration temperature, N_C is the number of activated delay cells at T_C. From (2), the proposed sensor can

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0011037). Chip fabrication was supported by IC Design Education Center (IDEC)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

TABLE I
THE PERFORMANCE COMPARISONS OF RECENT ONE-POINT CALIBRATION TEMPERATURE SENSOR

Paper	Temperature Range (ºC)	Sensing Error (ºC)	Conversion Rate (Samples/sec)	Resolution (ºC/bit)	Area (mm^2)	Technology (μm)	Energy (J)
[4]	0~100	-4.0~4.0	5k	0.78	0.12	0.13	240n
[5]	0~60	-5.1~3.4	10k	0.139	0.01	0.065	15n
[6]	-40~100	-2.7~2.9	366k	0.043	0.0066	0.065	1.09n
This work	20~120	-0.63~1.04	430k	0.595	0.031	0.13	0.67n

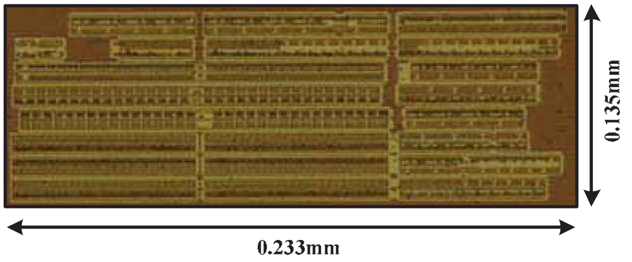


Fig. 2. The microphotography of proposed temperature sensor

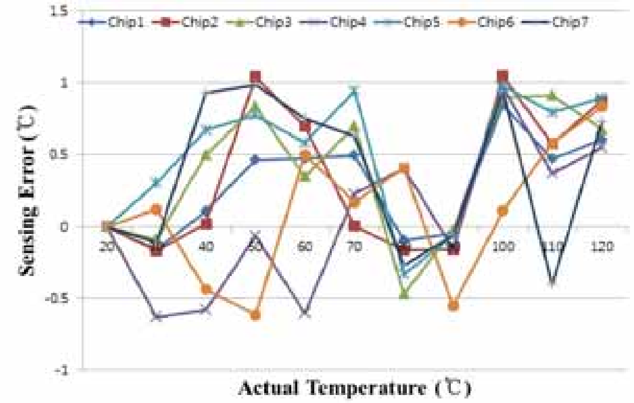


Fig. 3. The temperature sensing error and measured temperature of 7 test chips

achieve the constant value of *Temp_code*$_C$ by adjusting N_C according to any various P due to process variation.

Since *Ref_code* replaces the reference generator [4], the proposed sensor can save the area and power consumption. Furthermore, to find the code difference, the sensor compares each bit of the code concurrently. Thereby the sensor can achieve the high conversion rate with high energy efficient.

B. Temperature Sensing Procedure

After the calibration procedure, the proposed sensor starts the temperature sensing. If the operation temperature (T_O) changes from T_C, *Temp_code* at T_O (*Temp_code*$_O$) becomes different from *Temp_code*$_C$ due to the temperature term in (1). In the sensing procedure, *Temp_code*$_O$ is adjusted to be equal to *Ref_code* with the same as calibration procedure. The equation of *Temp_code*$_O$ is shown in following.

$$Temp\_code_O = Temp\_code_C = N_O \cdot T_O^{-\alpha} \cdot P = N_C \cdot T_C^{-\alpha} \cdot P \qquad (3)$$

where N_O is the number of activated delay cells at T_O. Thus,

$$T_O = T_C \left(N_O / N_C \right)^{\alpha} \qquad (4)$$

As a result, the process variation related to the term P is removed and T_O can be obtained independently using known parameters, T_C, N_C, and N_O.

III. MEASUREMENT RESULTS

The proposed sensor was fabricated using 0.13μm CMOS process technology. The microphotography of the chip is shown in Fig. 2 and the active area of the proposed sensor is 0.031-mm^2. The measurement results of 7 test chips measured over 20 to 120°C with 1.2V supply voltage. The measured energy per conversion rate is 0.67nJ/S with 430k samples/sec of conversation rate and the range of sensing resolution is 0.595 °C/bit. Fig. 3 shows the sensing error is from -0.63 to +1.04°C with a 2nd order master curve fitting for curvature correction. Table I shows that the performance comparison of the proposed sensor with recently published temperature sensors which adopts the one-point calibration.

IV. CONCLUSION

In this paper, the time-domain temperature sensor for on-chip thermal management is proposed with one-point calibration. The measurement results show that the proposed sensor achieves low sensing error, high energy efficiency, and the high conversion rate compared with other sensors by using the code difference between temperature-dependent code and reference code.

REFERENCES

[1] S. Hartwig, M. Luck, J. Aaltonen, R. Serafat, and W. Theimer, “Mobile multimedia – challenges and opportunities,” *IEEE Trans. Consumer Electron.*, vol. 46, no. 4, pp. 1167-1178, Nov. 2000.

[2] Z. Wei, K. L. Tang, and K. N. Ngan, “Implementation of H.264 on mobile device,” *IEEE Trans. Consumer Electron.*, vol. 53, no. 3, pp. 1109-1116, Aug. 2007.

[3] “Whitepaper The Benefits of Quad Core CPUs in Mobile Devices” Internet: www.nvidia.com/content/PDF/tegra_white_papers/tegra-whitepaper-0911a.pdf, [2011]

[4] Ha D., Woo K., Meninger S., Xanthopoulos T., Crain E., Ham D., “Time-Domain CMOS Temperature Sensors With Dual Delay-Locked Loops for Microprocessor Thermal Monitoring”, *IEEE Trans. Very Large Scale Integration (VLSI) Systems*, pp. 1-12, Aug. 2011

[5] Ching-Che Chung, Cheng-Ruei Yang, “An autocalibrated all-digital temperature sensor for on-chip thermal monitoring”, *IEEE Trans. Circuits and Systems II: Express Briefs*, vol. 58, pp.105-109, Feb. 2011

[6] Kisoo Kim, Hokyu Lee, Sangdon Jung, Chulwoo Kim, “A 366kS/s 400uW 0.0013mm^2 frequency-to-digital converter based CMOS temperature sensor utilizing multiphase clock”, *IEEE Custom Integrated Circuits Conference*, pp. 203-206, Sep. 2009

A Fast Adaptive Power Allocation for Maximizing a Discrete Utility Function of OFDMA System

Sungho Hwang, Youn Seon Jang, and Ho-Shin Cho, *Member, IEEE*

Abstract--**In this paper, we propose a fast adaptive power allocation algorithm for an orthogonal frequency division multiple access (OFDMA) system that employs an adaptive modulation and coding (AMC) scheme to significantly reduce the processing time. We show that the proposed algorithm reduces the complexity by *M* -the number of AMC levels- times compared with greedy adaptive power allocation (APA) which is known as the optimal algorithm for power allocation problem while still maintaining the utility at the same level as greedy APA.**

I. INTRODUCTION

The orthogonal frequency division multiple access (OFDMA) system is a multi-carrier system. Thus, the system capacity is obtained by the sum of individual subcarrier capacity which highly depends on the amount of power allocated to each subcarrier. Obviously, the subcarrier capacity becomes higher when more power is allocated. To maximize the overall system capacity, we should optimally control and assign each power of multiple carriers within a limited total transmission power. This power allocation has been a difficult and challengeable issue to many researchers. As one of the solutions for the capacity maximization, Lagrangian methods have been used. Water-filling, in which more power is allocated to stronger subcarriers, is known as the optimal solution for the capacity maximization using the Lagrangian methods [1].

However, water-filling is not appropriate to obtain the best performance when adaptive modulation and coding (AMC) scheme is employed, in which the power allocation should be performed with some discrete power levels. Regarding the issue, several adaptive power allocation (APA) schemes, which are sometimes called the "bit-loading" problem, have been investigated to increase the system utility in an OFDMA system with discrete AMC levels [3]-[5]. The APA schemes have developed with slightly different objectives and constraints. In [3], the maximum utility rate has been obtained under a limited power constraint while, in [5], total transmission power is minimized to achieve a target data-rate of system which is a sum of individual data-rates of all users. These schemes have iterative procedure to successfully achieve each objective by using optimal [3], heuristic [4], or sub-optimal [5] algorithms. Especially, as optimal algorithm, greedy APA (Modified Levin-Campello algorithm) maximizes the utility argument of power at each step of bit assignment. However, because zero power is initially allocated to all subcarriers, the processing time for iterative allocation of power to whole subcarriers until the total allocated power reaches to the limit is too long. In this paper, we propose a fast-converged power allocation method to significantly reduce the processing time for an OFDMA system employing AMC.

II. SYSTEM MODEL

We assume that AMC has M discrete channel states and the channels are already allocated to users by means of one of existing sub-carrier allocation methods such as maximum C/I or the proportional fairness (PF). In this paper, we only deal with the power allocation to each sub-carrier. The channel state of sub-carrier i is identified by the signal-to-noise ratio (SNR) ρ_i which the corresponding user experiences. And it is defined as $\rho_i = |h_i|^2/N_i$, where N_i and h_i are the noise power density and the frequency response of the sub-carrier i, respectively. [2]. Thus, the channel state vector is defined by

$$c = [\rho_1 \, \rho_2 \cdots \rho_S]^{\mathrm{T}} \tag{1}$$

where S denotes the number of sub-carriers.

III. PROBLEM FORMULATION

Let $U_i(\cdot)$ be the function of utility for sub-carrier i and P_{total} be total transmission power. If sub-carrier i has an allocated power P_i, the sub-carrier's utility is $U_i(P_i)$. Thus, we are interested in the following problem:

$$\text{Problem } (\mathbf{P})\text{: } \max_{P_i} \sum_{i=1}^{S} U_i(P_i) \tag{2}$$

$$\text{Subject to } \sum_{i=1}^{N_c} P_i \le P_{total} \; , \qquad P_i \ge 0 \tag{3}$$

where $U_i(P_i)=a_i\log(1+b_iP_i\rho_i)$, $U_i(P_i)$ is a differentiable non-decreasing concave function, a_i is the weight vector for the allocated user at sub-carrier i to generalize the problem and b_i is the factor for controlling bit error rate of the allocated user at sub-carrier i. By changing the value of a_i, Problem (**P**) can cover throughput maximization and also proportionally fair (PF) allocation, and b_i is determined by $b_i = -1.5/\ln(5\cdot\mathrm{BER}_i)$ [8], where BER_i is the bit error rate of the allocated user at sub-carrier i. Then, the optimal power allocation for problem (**P**) has the following solution [1]:

$$P_i^* = \left(\frac{a_i}{\lambda} - \frac{1}{b_i\rho_i}\right)^+ \tag{4}$$

where $(x)^+=\max(x, 0)$, and λ is chosen such that the power constraint is met:

$$\sum_{i=1}^{S}\left(\frac{a_i}{\lambda} - \frac{1}{b_i\rho_i}\right)^+ = P_{total} \tag{5}$$

And we know that the performance improvements are marginal even though APA is employed in continuous utility system under following assumptions [2]:

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

A1) The variation of each sub-carrier channel state is sufficiently low.

A2) The variation of weighting factor a_i gets smaller for maximizing long term utility.

That is, under these assumptions,

$$P_i^* \approx P_{total}/S \tag{6}$$

Using this property, we propose a new algorithm for a discrete data rate. The proposed algorithm allocates power to maximize the system utility.

To present the proposed algorithm, we explain one function. $d(P)$ is the lowest power in the level which contains P for the discrete system utility function. The proposed algorithm is described in following procedures:

1) For initialization, $P_i = d\left(P_{total}/S\right)$.

2) The remain power is allocated using greedy APA.

For the iteration part, greedy APA algorithm finds the sub-carrier that has the maximum value of system utility per power. Thus, the computational complexity of once iteration is $O(S)$. Because other parts of the iteration and the initialization have much smaller computational complexity than $O(S)$, we can ignore those parts. Therefore, if A_i is the number of power allocations of sub-carrier i, the computational complexity of greedy APA is

$$\sum_{i=1}^{N_c} A_i O(S) \tag{6}$$

Because max A_i is M, we can rewrite Eq. (6) as the following:

$$O(M S^2) \tag{7}$$

For the proposed scheme, the computational complexity of the initialization is $O(S)$. And, because the iteration part is the same as greedy APA, the computational complexity of the proposed scheme is $O(S)$ during once iteration. However, because of the initialization, remain power is given by $\Sigma_i \Delta P_i^*$ which is always less than $\Sigma_i\ \Delta P_i^N$. Therefore, the power allocation of each sub-carrier occurs only once or not at all. Finally, the maximum computational complexity of the proposed scheme is

$$O(S^2) \tag{8}$$

IV. NUMERICAL RESULTS

The OFDMA system, proposed by the IEEE 802.16 WMANS standard [6], is considered herein with 20 sub-carrier and 7 AMC levels. The probability density function of the SNR is assumed to be an exponentially distributed random variable with a mean of 12 dB [7]. And we use the overall system capacity as the system utility for a simple simulation.

Fig. 1 shows the average of system utility for 80000 symbol times. In Fig. 1, the proposed algorithm shows the same performance as greedy APA.

V. CONCLUSION

In this paper, we proposed a fast APA algorithm to maximize the discrete system utility for an OFDMA system employing AMC. The proposed scheme showed the same performance compared to greedy APA, which is the optimum solution for a discrete system utility function, and reduced the complexity by maximum M times compared with greedy APA.

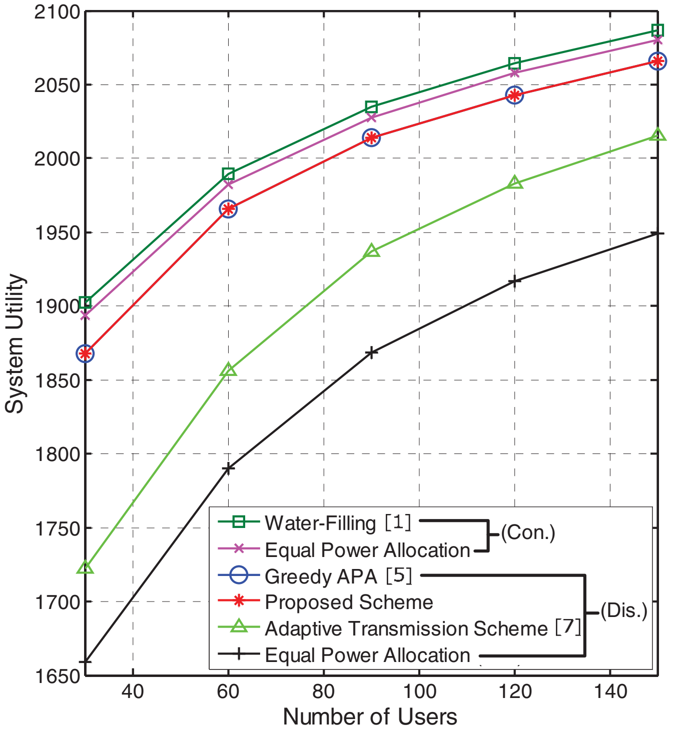


Fig. 1. Average system utility versus number of users

REFERENCES

[1] D. Tse, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[2] G. Song, and Y. (G.) Li, “Cross-Layer Optimization for OFDM Wireless Networks-Part I: Theoretical Framework,” *IEEE Trans. Commun.*, vol. 4, no. 2, pp. 614 – 624, March 2005.

[3] G. Song, and Y. (G.) Li, “Cross-Layer Optimization for OFDM Wireless Networks-Part II: Algorithm Development,” *IEEE Trans. Commun.*, vol. 4, no. 2, pp. 625 – 634, March 2005.

[4] K. Kim, “Efficient Adaptive Modulation and Power Allocation Algorithm for OFDMA Cellular Systems,” in *Proc. 2005 Wireless Telecommunications Symposium*, pp. 169 – 173, April 2005.

[5] Y.-F. Chen, and J.-W. Chen, “A Fast Subcarrier, Bit, and Power Allocation Algorithm for Multiuser OFDM-based Systems,” *IEEE Trans. Veh. Tech.*, vol. 57, no. 2, pp. 873–881, Jan. 2008.

[6] S.H. Ali, Lee Ki-Dong, and V.C.M. Leung, “Dynamic resource allocation in OFDMA wireless metropolitan area networks,” *IEEE Wireless Communications*, vol. 14, issue 1, Feb. 2007, pp. 6-13.

[7] J.G. Proakis, *Digital Communications, second edition*, McGraw-Hill, New York, 1989.

[8] X. Qiu and K. Chawla, “On the performance of adaptive modulation in cellular systems,” *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 884-895, Jun. 1999.

Enhanced Min-Sum Decoding for DVB-T2 LDPC Codes

Sung Ik Park[1)], *Member, IEEE*, Young Min Choi[2)], Heung Mook Kim[1)], *Member, IEEE*, Jeongchang Kim[3)], *Member, IEEE*, and Wangrok Oh[4)], *Member, IEEE*

[1)] Electronics and Telecommunications Research Institute, Daejeon, Korea
[2)] Cleverlogic, Daejeon, Korea
[3)] Korea Maritime University, Busan, Korea
[4)] Chungnam National University, Daejeon, Korea

***Abstract*-- In this paper, we propose an enhanced min-sum decoding algorithm for DVB-T2 LDPC codes. It is obtained by two step approximation of the function $\ln\cosh(x)$ on the sum-product algorithm. Simulation results show that the proposed algorithm does not cause serious performance degradation, as compared with sum-product algorithm. In addition, it shows better performance than conventional min-sum algorithm and its modifications.**

I. Introduction

Digital Video Broadcasting - Second Generation Terrestrial (DVB-T2) has adopted low density parity-check (LDPC) codes for powerful error-correction code [1]. Recently, for commercial use of LDPC codes, efficient decoding algorithms have been studied extensively. The sum-product algorithm (SPA) which is a conventional decoding algorithm for LDPC codes can be simplified as the min-sum algorithm (MSA) and its modifications [2], which greatly reduce the implementation complexity. However, the MSA and its modifications may lead to considerable performance degradation.

This paper proposes an enhanced MSA for DVB-T2 LDPC codes. The proposed algorithm may be efficiently implemented because it consists of integer additions, comparisons, and bit-shift operations without multiplications (divisions) and memories. Moreover, the proposed algorithm doesn't need any channel information such as signal to noise ratio (SNR).

II. Sum-Product and Modified Min-Sum Algorithms

Kschischang et al. [3] showed that the SPA operates in a factor graph. They briefly described the SPA for binary variable and check nodes of degree 3 in the factor graph of an LDPC code. Note that the variable nodes represent codeword symbols and the check nodes enforce the constraint that the adjacent symbols should have even overall parity.

The vectors (p_0, q_0) and (p_1, q_1) denote probability mass functions (PMF) for binary random variables, and each binary PMF can be parameterized by a single value. Among several different parameterizations, we will deal only with log-likelihood ratio (LLR) values. The following SPA for binary variable and check nodes is well-known [3].

Sum-Product Algorithm Based on LLR

For $\Lambda_1 = \ln(p_0/p_1)$ and $\Lambda_2 = \ln(q_0/q_1)$,

$$\mathrm{VAR}(\Lambda_1,\Lambda_2) = \Lambda_1 + \Lambda_2, \tag{1}$$

$$\mathrm{CHK}(\Lambda_1,\Lambda_2) = \ln\cosh\left(\frac{\Lambda_1+\Lambda_2}{2}\right) - \ln\cosh\left(\frac{\Lambda_1-\Lambda_2}{2}\right). \tag{2}$$

The VAR and CHK functions can be extended to more than two arguments recursively as follows:

$$\mathrm{VAR}(\Lambda_1,\Lambda_2,\ldots,\Lambda_n) = \mathrm{VAR}(\Lambda_1,\mathrm{VAR}(\Lambda_2,\ldots,\Lambda_n)), \tag{3}$$

$$\mathrm{CHK}(\Lambda_1,\Lambda_2,\ldots,\Lambda_n) = \mathrm{CHK}(\Lambda_1,\mathrm{CHK}(\Lambda_2,\ldots,\Lambda_n)), \tag{4}$$

In general, $\ln\cosh(x)$ has been implemented by a look-up table instead of floating point arithmetic, because of its high complexity [4]. Since $\ln\cosh(x) = |x| - \ln 2$ for $|x| \gg 1$, an approximation to (2) may be given by

$$\mathrm{CHK}(\Lambda_1,\Lambda_2) \approx \mathrm{sgn}(\Lambda_1)\,\mathrm{sgn}(\Lambda_2)\min(|\Lambda_1|,|\Lambda_2|), \tag{5}$$

which are precisely the min-sum update rule at a check node [3]. Since the differences between $\ln\cosh(x)$ and $|x| - \ln 2$ at near $x = 0$ can be increased up to $\ln 2 (\approx 0.693)$, they may induce a serious performance degradation. To improve accuracy of check node update in the MSA, the following two check node update rules have been proposed [2].

Scaled Check Node Update Based on LLR

$$\mathrm{CHK}(\Lambda_1,\Lambda_2) = \alpha\cdot\mathrm{sgn}(\Lambda_1)\,\mathrm{sgn}(\Lambda_2)\min(|\Lambda_1|,|\Lambda_2|), \tag{6}$$

Offset Check Node Update Based on LLR

$$\begin{aligned}&\mathrm{CHK}(\Lambda_1,\Lambda_2)\\ &\quad= \mathrm{sgn}(\Lambda_1)\,\mathrm{sgn}(\Lambda_2)\max\big(\min(|\Lambda_1|,|\Lambda_2|)-\beta, 0\big),\end{aligned} \tag{7}$$

Here, $\alpha (< 1)$ is scaling constant and β is correction factor, and good α and β values can be found by density evolution [5]. Note that the MSA and its modifications do not require any channel information, such as the signal to noise ratio (SNR), and work just the received values as inputs.

III. Enhanced Min-Sum Algorithm

The MSA may result in relatively large errors in the case of small LLR values. However, it may work well in high SNR region since $\ln\cosh(x)$ can be well approximated to $|x| - \ln 2$

This research was supported by the Korea Communications Commission (KCC), Rep. of Korea,under the support program supervised by the Korea Communications Agency (KCA) [Development of Wakeup Emergency Alert Broadcast System Using DTV].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

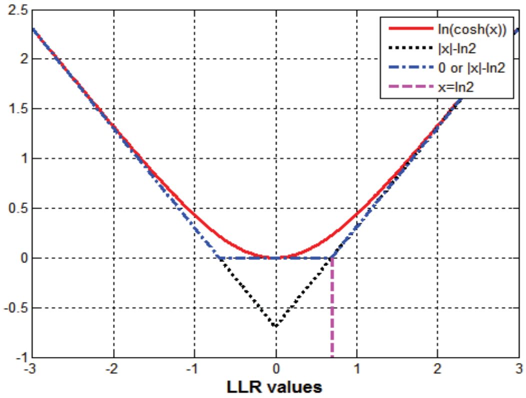


Fig. 1. Step approximation of $\ln\cosh(x)$

for large LLR values, as shown in Fig. 1. Our goal is to find a more accurate approximation that is valid even for low SNR. As shown in Fig. 1, $\ln\cosh(x)$ can be approximated to the following two step functions:

$$\ln\cosh(x) \approx \begin{cases} 0, & \text{if } |x| \le \ln 2 \\ |x| - \ln 2, & \text{otherwise} \end{cases}. \tag{8}$$

If we assume binary-input AWGN channel, then $\Lambda_1 = 2/\sigma^2 \cdot r_1$ and $\Lambda_2 = 2/\sigma^2 \cdot r_2$ where r_1 and r_2 are the received values from channel and σ^2 is the variance of the channel noise. Moreover, if we assume that LLR values from channel are uniformly distributed, then, the equation (2) can be approximated as follows:

$$\mathrm{CHK}(\Lambda_1, \Lambda_2) \approx$$

$$\frac{2}{\sigma^2} \cdot \begin{cases} 0, & \begin{array}{l} |r_1 + r_2| \le \sigma_{op}{}^2 \cdot \ln 2, \\ |r_1 - r_2| \le \sigma_{op}{}^2 \cdot \ln 2 \end{array} \\ \mathrm{sgn}(r_1)\mathrm{sgn}(r_2)\min(|r_1|, |r_2|) - C_{offset}, & \begin{array}{l} |r_1 + r_2| > \sigma_{op}{}^2 \cdot \ln 2, \\ |r_1 - r_2| \le \sigma_{op}{}^2 \cdot \ln 2 \end{array} \\ \mathrm{sgn}(r_1)\mathrm{sgn}(r_2)\min(|r_1|, |r_2|) + C_{offset}, & \begin{array}{l} |r_1 + r_2| \le \sigma_{op}{}^2 \cdot \ln 2, \\ |r_1 - r_2| > \sigma_{op}{}^2 \cdot \ln 2 \end{array} \\ \mathrm{sgn}(r_1)\mathrm{sgn}(r_2)\min(|r_1|, |r_2|), & \begin{array}{l} |r_1 + r_2| > \sigma_{op}{}^2 \cdot \ln 2, \\ |r_1 - r_2| > \sigma_{op}{}^2 \cdot \ln 2 \end{array} \end{cases}, \tag{9}$$

where $C_{offset} = \sigma_{op}{}^2 \cdot \ln 2/4$ and $\sigma_{op}{}^2$ is the noise variance at the operating point. $\sigma_{op}{}^2$ is constant value, and it can be found by computer simulation. The proposed algorithm based on a step approximation of function $\ln\cosh(x)$ can be efficiently implemented because it consists of integer additions (subtractions), comparisons, and bit-shift operations. Note that the proposed algorithm not only does not need a significant amount of memory for storing the look-up table and a processor for accessing the memory, but also does not require any channel information such as SNR.

Simulations have been conducted to compare the performance of the proposed algorithm with those of the SPA, MSA, scaled MSA, and offset MSA. In the simulations, we used DVB-T2 LDPC code with length of 64,800 and rate of 1/3 [1]. Fig. 2 shows the bit error rate (BER) of LDPC code according over additive white Gaussian noise (AWGN) channel. In Fig. 2, value of σ_{op} was found by computer simulation. The number of iterations is set to be 50. The simulation results in Fig. 2 show that DVB-T2 LDPC code with the proposed MSA has just 0.02 dB poor performance as those of LDPC code with SPA at the BER = 10^{-5}. While the performance gap between the proposed MSA and Offset MSA is about 0.14 dB at BER=10^{-5}, that between the proposed MSA and SMSA is about 0.23 dB at the same BER. These results indicate that there is a tradeoff between performance and decoding complexity.

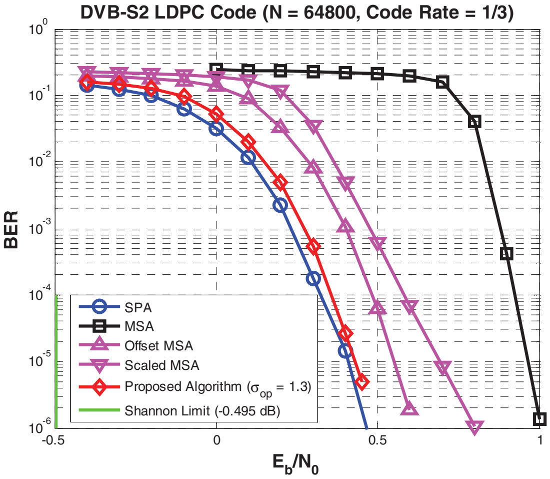


Fig. 2. BER of DVB-T2 LDPC code with length of 64,800.

IV. CONCLUSION

In this paper, we proposed an enhanced min-sum decoding algorithm for DVB-T2 LDPC codes. Simulation results show that the proposed decoding algorithm not only shows much better performance than MSA and its modifications, but also does not cause serious performance degradation, as compared with the SPA.

REFERENCES

[1] ETSI, "Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)," ETSI EN 302 755 V1.2.1, Feb. 2011.

[2] M. P. C. Fossorier, M. Mihaljevic, and H. Imai, "Reduced complexity iterative decoding of low density parity check codes based on belief propagation," *IEEE Trans. Commun.*, vol. 47, pp. 673-680, May 1999.

[3] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graph and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498-519, Feb. 2001.

[4] X.-Y. Hu, E. Eleftheriou, D.-M. Arnold, and A. Dholakia, "Efficient implementation of the sum-product algorithm for decoding LDPC codes," in *Proc. IEEE Globecom*, San Antonio, TX, Nov. 2001, pp. 1036-1036E.

[5] J. Chen and M. P. C. Fossorier, □Density evolution for two improved BP-based decoding algorithm of LDPC codes,□ *IEEE Commun. Lett.*, vol. 6, pp. 208-210, May 2002

A Dual-Code-Rate Memoryless Viterbi Decoder for Wireless Communication Systems

Chu Yu[1], *Yu-Shan Su*[1], *Bor-Shing Lin*[2], *Po-Hsun Cheng*[3], and *Sao-Jie Chen*[4]

[1]Department of Electronic Engineering,
National ILan University, Yilan, Taiwan, R.O.C.
[2]Department of Computer Science and Information Engineering,
National Taipe University, Taipe, Taiwan, R.O.C.
[3]Department of Software Engineering,
National Kaohsiung Normal University, Kaohsiung, Taiwan, R.O.C.
[4]Graduate Institute of Electronics Engineering,
National Taiwan University, Taipei, Taiwan, R.O.C.

***Abstract*--This paper presents a novel dual-code-rate memoryless Viterbi decoder with a 4-level soft decision for wireless communication systems. Based on the proposed architecture, the survivor memory can be eliminated totally, which will significantly reduce 50% of the total power dissipation. As shown, the proposed design uses approximately 27.5K gates in 0.18 μm CMOS technology, and its power consumption is approximately 9.8mW at 80MHz.**

I. Introduction

Software defined radio (SDR) is one of the most emerging research topics in mobile and personal communications. SDR can be viewed as a reprogrammable radio system, which supports a multi-mode radio. Therefore, it is required to provide a run-time reconfiguration capability for the system to reduce the time of deploying new communication standards. This technique is advantageous to international roaming for potable mobile devices. In this paper, either a rate-1/2 or a rate- 1/3 memoryless Viterbi decoder (VD) with a 4-level soft decision is proposed for wireless communication systems used in SDR systems. The proposed design can be applied to IEEE 802.11 a/g, long term evolution (LTE) systems, or other wireless baseband systems.

Based on the proposed VD architecture, the survivor memory unit (SMU) can be eliminated. Since the SMU almost consumes 50% of the total power, the proposed design can gain lower power dissipation than the classical one. Moreover, the proposed design can directly generate the output stream without using an SMU, it thus has a low latency of two clock cycles. This value is also far lower than that of the classical VD hardware.

II. Proposed Architecture

Based on the modified VD algorithm, a novel dual-code-rate VD architecture is proposed, as shown in Fig. 1. The architecture is composed of a reconfigurable branch metric unit (RBMU), a path metric memory (PMM), an add-compare-select unit (ACSU), a path selection unit (PSU), and a path tracer unit (PTU). Compared with the classical VD hardware, the proposed design does not need a survivor memory unit. More detailed functions of these units will be described in the following paragraphs.

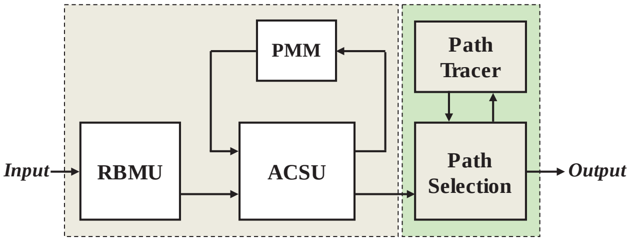


Fig. 1. Block diagram of proposed Viterbi decoder.

The branch metric in a Viterbi decoder is given by the squared distance between the received noisy sequence and the output signals. Assume that two soft decision inputs R_0 and R_1 with rate-1/2 were sent to the decoder, the branch metric $BM(\gamma_{0,} \gamma_1)$ can be obtained by:

$$BM(\gamma_0, \gamma_1) = \sum_{i=0}^{1} \left(R_i - s(\gamma_i)\right)^2, \tag{1}$$

where $\gamma_i \in \{0,1\}$, and the output binary symbol $s(\gamma_i)$ has the symmetric symbol ρ and -ρ. Since $s(\gamma_i) \in \{0,1\}$, after (1) is subtracted by a constant and divided by 2ρ, it can be simplified as [2]:

$$BM(\gamma_0, \gamma_1) = -\sum_{i=0}^{1} \left(R_i \frac{s(\gamma_i)}{\rho} \right) \tag{2}$$

Similarly, a rate-1/3 branch metric can be extended by the above derived equation as follows:

$$BM(\gamma_0, \gamma_1, \gamma_2) = -\sum_{i=0}^{2} \left(R_i \frac{s(\gamma_i)}{\rho} \right) \tag{3}$$

Based on (2) and (3) as well as sharing the common terms, a reconfigurable branch metric unit (RBMU) is proposed, as shown in Fig. 2, which provides either a rate-1/2 or a rate-1/3 branch metric. Through the multiplexers, the unit can compute the branch metrics of the four/eight states from $BM(0,0)/BM(0,0,0)$ to $BM(1,1)/BM(1,1,1)$ for every input data at a 1/2 or 1/3 code rate. Compared with [1] and [2], our design reduces 25% of full adders.

This work was supported by the National Science Council, R.O.C., under Grant NSC100-2220-E-197-001.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

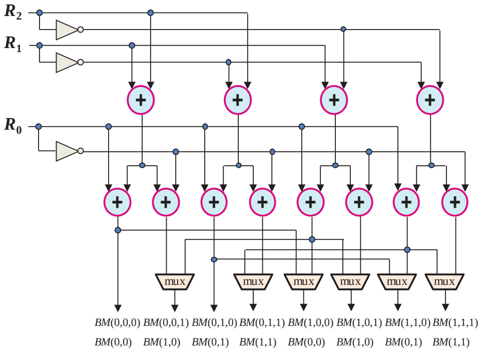

R_2
R_1
R_0
+
mux
$BM(0,0,0)$ $BM(0,0,1)$ $BM(0,1,0)$ $BM(0,1,1)$ $BM(1,0,0)$ $BM(1,0,1)$ $BM(1,1,0)$ $BM(1,1,1)$
$BM(0,0)$ $BM(1,0)$ $BM(0,1)$ $BM(1,1)$ $BM(0,0)$ $BM(1,0)$ $BM(0,1)$ $BM(1,1)$

Fig. 2. Proposed RBMU hardware structure.

The original VD algorithm for constraint length k = 3 can be described as:

$$PM_{0,t=i+1} = \min\{PM_{0,t=i} + BM_{0,t=i}, PM_{2,t=i} + BM_{3,t=i}\} \tag{4}$$

$$PM_{1,t=i+1} = \min\{PM_{0,t=i} + BM_{3,t=i}, PM_{2,t=i} + BM_{0,t=i}\} \tag{5}$$

$$PM_{2,t=i+1} = \min\{PM_{1,t=i} + BM_{1,t=i}, PM_{3,t=i} + BM_{2,t=i}\} \tag{6}$$

$$PM_{3,t=i+1} = \min\{PM_{1,t=i} + BM_{2,t=i}, PM_{3,t=i} + BM_{1,t=i}\}, \tag{7}$$

where $PM_{n,\ t=i}$ denotes the n-th path metric at the i-th stage, and $BM_{r,\ t=i}$ stands for the branch metric of the state r at the i-th stage. Based on (4)-(7), the ACSU with a PMM structure is shown in Fig. 3, which is only a portion of ACSU for k = 7. The CMP circuit used in our ACSU stands for a comparator.

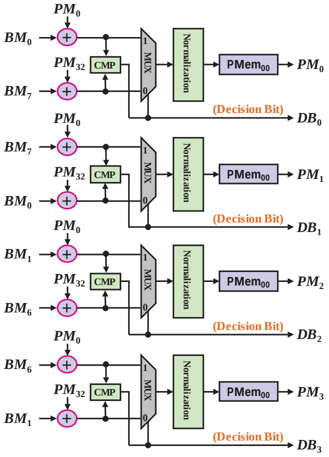


Fig. 3. ACSU with a PMM structure.

Generally, the longer a survivor path is the better. However, for implementation, this will make the hardware cost huge, because much more memory space is required. The path thus needs a proper truncation. Nonetheless, since the ACSU needs accumulating a number of input data under a finite data length, overflow still may occur in the computation process. Here, a data length of 9 bits is used in our design. To avoid overflow occurrence during computing the new path metric, a normalization mechanism is required to be used in the ACSU. According to our simulation result, the proposed normalization mechanism is only used one time every four operations on average. Moreover, this normalization mechanism also has the advantage of increasing the length constraint, thus increasing the received data fidelity.

The PSU is just a simple multiplexer, which is responsible for selecting a proper decision bit (DB_i) of ACSU as the input data of the next stage. Based on the Hamming distance, the most probable survivor path is chosen.

The final stage of our design is a path tracer unit. According to the decision bit from the previous ACSU, the unit produces the wanted output of the VD. State transitions are based on the Trellis diagram of the VD.

III. Performance Evaluation

Table I summarizes of the performance evaluation of various VD architectures, respectively. The designs listed in Table I are mainly implemented in 0.18 μm CMOS technology, except for [2] in 0.5 μm. From the table, because the SMU is not used in from our design, the storage space is smaller. Moreover, the SMU is a power-hungry part, which consumes approximately 50% of the total power. Therefore, our design significantly saves substantial power consumption and hardware area. In addition, the bit error rate performance of the proposed VD approximates to that in [1].

TABLE I.
Comparison of Various Viterbi Architectures

Design	Type	SMU Included	Code Rate	Power (mW)	Gates
[2]	soft 3-bit	yes	1/3	9.8 @2Mhz	-
[3]	soft 4-bit	yes	1/2	68 @72MHz	49K
[4]	soft 5-bit	yes	1/2	58 @100MHz	-
Ours	soft 4-bit	no	1/2 or 1/3	**9.8** **@80MHz**	**27.5K**

IV. Conclusion

A dual-code-rate memoryless VD architecture with low latency has been presented in this paper for SDR systems. The proposed architecture can provide either rate-1/2 or rate-1/3 Viterbi decoding. Based on the proposed architecture, all the probable survivor paths can be identified individually. This result shows that our design can eliminate the survivor storage space, and produce the output bit-steam directly. Since the SMU is eliminated, our design consumes lower power and spends lower chip area than the classical designs. These foregoing features make our design suitable for applications in portable wireless devices.

Reference

[1] D. A. El-Dib and M. I. Elmasry, “Memoryless Viterbi decoder,” *IEEE Transactions on Circuits and Systems-II*, vol. 52, no. 12, pp. 826-830, Dec. 2005.

[2] Y. N. Chang, H. Suzuki, and K. K. Parhi, “A 2-Mb/s 256-State 10-mW Rate-1/3 Viterbi Decoder,” *IEEE J. Solid-State Circuits*, vol. 35, no. 6, pp. 826-834, Jun. 2000.

[3] C. C. Lin, Y. H. Shih, H. C. Chang, and C.Y. Lee, “Design of a Power–Reduction Viterbi Decoder for WLAN Application,” *IEEE Trans. Circuit and Systems I*, vol. 52, no. 6, pp. 1148-1156, Jun. 2005.

[4] Y. C. Tang, D. C. Hu, W. Y. Wei, W. C. Lin, and H. C. Lin, “A Memory-Efficient Architecture for Low Latency Viterbi Decoders,” in *Proc. IEEE International Symposium on VLSI Design, Automation and Test*, VLSI-DAT '09, Apr. 2009, pp. 335-338.

Software-Defined DVB-T2 Receiver Using Coarse-Grained Reconfigurable Array Processors

Navneet Basutkar, Ho Yang, Peng Xue, Kitaek Bae, and Young-Hwan Park
Samsung Advanced Institute of Technology, Samsung Electronics Co., Ltd.
Yongin-si, Gyeonggi-do, 446-712 Korea

***Abstract*—This paper describes the feasibility of software implementation of DVB-T2 receiver with DTG-106 [1] mode using the coarse-grained reconfigurable array (CGRA) based processor. This paper focuses mainly on DVB-T2 system design and implementation of major software functions of DVB-T2 demodulator: FFT, frequency interpolation, multi-level deinterleaving, and soft-demapper. By implementing the full chain DVB-T2 software and measuring the cycle performance, we demonstrate the software implantation of DVB-T2 on dual core CGRA processor running at 400MHz.**

I. Introduction

DVB-T2 is a standard for second generation digital terrestrial television broadcasting system, offering significant benefits compared to the first generation of DVB. DVB-T2 provides nearly 50% increased capacity [1], compared to the current UK mode of DVB-T and it has greater tolerance of multipath and impulsive interference. In order to meet high data rate requirement, the high order OFDM modulation schemes up to 256QAM have been introduced in DVB-T2 with very demanding functions such as various sized FFTs up to 32K-FFT, rotated constellation mappings, 4-level interleaving and hybrid channel coding of LDPC and BCH. The implementation of DVB-T2 receiver thus requires intensive computation of signal processing algorithms, which makes it support up to 36Mbits/s data transmission [1]. Furthermore, many different operational modes defined in DVB-T2 standard inevitably require more chip area when implemented in hardware, and substantially increases the design cost when supporting multiple broadcasting standards. Software implementation of the broadcasting receivers provides attractive solution by supporting the different operational modes in software on a fixed hardware processor. In particular, software-defined function blocks can be easily extended to support many different broadcasting modes defined in the standard, even covering other broadcasting standards with minimum modifications.

This paper describes the software implementation of DVB-T2 demodulator using the coarse-grained reconfigurable array (CGRA) [2] processor consisting of 16 functional units with 4-way SIMD vector processing on each functional unit. CGRA based systems are well-known in the field of software-defined radio (SDR) mainly due to its reconfigurability of the function arrays through software programming [3]. In particular, the combination of SIMD and the CGRA architecture is computationally very efficient for iterative vector data processing of the most communication algorithms.

Considering the complexity of the DVB-T2 standard, we focus on FFT, frequency interpolation, deinterleaving schemes and soft-demapper, since these are the major function blocks in DVB-T2 system. By optimizing software functions we could satisfy the design requirement for the typical mode of DTG-106 on dual core CGRA processor running at 400MHz.

II. System Architecture

The conceptual data flow of the DVB-T2 receiver consists of 4 major blocks: signal detection and extraction, channel equalization, demodulation & bust error protection and error correction, as shown in Figure 1.

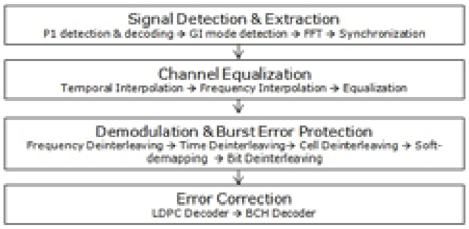

Signal Detection & Extraction
P1 detection & decoding → GI mode detection → FFT → Synchronization
Channel Equalization
Temporal Interpolation → Frequency Interpolation → Equalization
Demodulation & Burst Error Protection
Frequency Deinterleaving → Time Deinterleaving → Cell Deinterleaving → Soft-demapping → Bit Deinterleaving
Error Correction
LDPC Decoder → BCH Decoder

Fig. 1: Typical Data Flow of DVB-T2 Receiver.

The system design for the typical data flow of Figure 1 looks straight forward. However, the real working scenario of the DVB-T2 using DSP is more complicated and should consider following three implementation issues. The first one is shift in the processing data size; the minimum signal processing unit before Time-Deinterleaver (TDI) is the OFDM symbol, while other blocks are being processed based on the size of FEC block. For the typical mode of DTG-106 the sizes of OFDM symbol and FEC block are 3.61msec and 1.06msec, respectively. Because of the different data unit, two CGRA cores are considered for processing separately in two different threads which are run on two different cores in parallel. The second one is the data buffering between major functions; the buffering scheme for Time-Deinterleaver (TDI) defines the

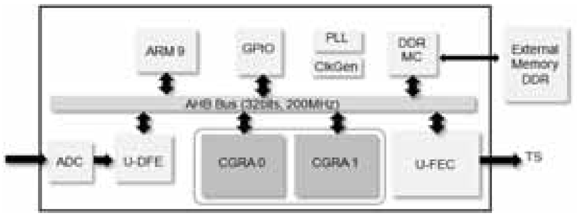


Fig. 2: Baseband Platform Architecture for DVB-T2 Receiver.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

size of memory and affects the memory architecture of the receiver because of the large buffer size requirement of Time-Interleaving (TI) block. The third one is the hierarchy of the data storage; besides the TDI block the temporal interpolation requires at least three OFDM symbols for the worst case scenarios, which requires about 384KB for 32K-FFT mode. The data memories for FFT and stream adaptation are also considered when designing the memory hierarchy.

Figure 2 shows the CGRA processor based platform architecture for DVB-T2 demodulation. The system uses ARM CPU for system control and stream adaptation, and multi-layer AHB bus of 32bit data path. Other important dedicated hardware modules are the universal-Digital Front-End (U-DFE) for Automatic Gain Controller and flexible filter, and the universal-Forward Error Correction (U-FEC) for LDPC/BCH channel decoders.

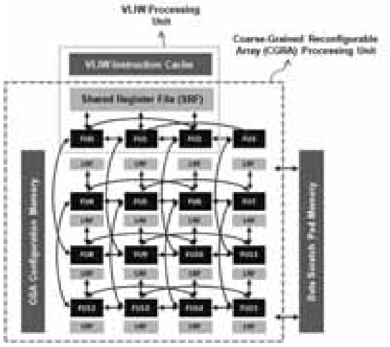


Fig. 3: Coarse-Grained Reconfigurable Array(CGRA) processor.

Figure 3 shows the CGRA core architecture used for the major functions of demodulator. The CGRA core has 16 functional units of 64bits (4-way SIMD) vector processing. Other important features include complex vector multiplication, scalar division and special intrinsic used for various types of interleaving schemes.

III. Software Design and Technical Challenges

One of the major challenges was to implement FFT with 16bit precision. A low complexity algorithm [4] specially proposed for DVB-T2 is used for software implementation of FFT. Aggressive scaling mechanism is used for selected FFT stages to avoid data overflow in output due to successive additions in FFT, thus maintaining high SQNR of 51 dB on average of all FFT sizes. Furthermore, an universal FFT is designed to compute all the FFT modes (1k, 2k, 4k, 8k, 16k and 32k).

For MISO channel estimation, the received pilots are partitioned into two subsets, sum and difference pilots. After temporal interpolation, the sum and difference of the channel response on each subcarrier are obtained from the frequency interpolation of the sum and difference channel response subsets, respectively. Therefore, the computational complexity is doubled when compared with the SISO case.

DVB-T2 has 4-level interleaving scheme for protecting burst errors of channel impairment. Special intrinsics are designed to optimize different deinterleaving schemes of DVB-T2 demodulator.

For the current DVB-T2 implementation, only soft-demapping for the non-rotated QAM is considered. The conventional soft-demapping for rotated QAM in DVB-T2 [1] is a highly complicated full-search algorithm which consumes high amount of processing delay. A new low complexity soft-demapping algorithm to reduce its complexity as low as that of soft-demapping for the non-rotated QAM while maintaining good BER performance is under research.

IV. Discussion

This paper discussed software implementation of the DVB-T2 standard focusing on DTG-106 mode. Figure 4 shows the software mapping results for the DTG-106 mode which is most popular and adopted by UK. Top four major blocks are Frequency Interpolation, Soft-Demapping, Bit-Deinterleaver and 32K-FFT.

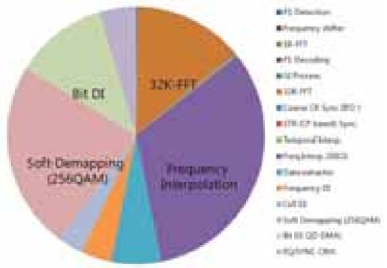


Fig. 4: Cycle distribution of kernels for DTG-106 mode.

Currently, the full chain implementation of DVB-T2 receiver results in the performance of 516 M cycles/sec. This performance can be implemented with 2 CGRA cores having 400MHz clock frequency with 67% load on CGRA0 and 33% load on CGRA1.

In spite of good results of current mapping of DTG-106 mode, we observed two major limitations in terms of cycle performance for FIR filtering and rotated QAM soft-demapping when covering other mode of operations. The first one is easily resolved by introducing new intrinsic for vector inner product, which results in about 30% cycle performance enhancement in early experimentation. However, the rotated QAM soft-demapper requires more fundamental design enhancement including either architecture and/or algorithm modification, which is the future research topic of the current DVB-T2 software implementation.

References

[1] *Digital Video Broadcasting (DVB); Implementation guidelines for a second generation digital terrestrial television broadcasting system (DVB-T2)*, ETSI Std. DVB Document A133, 2010.

[2] B. Mei, S. Vernalde, D. Verkest, H. D. Man, and R. Lauwereins, "ADRES: An architecture with tightly coupled VLIW processor and coarse-grained reconfigurable matrix," in *In Field-Programmable Logic and Applications*, 2003.

[3] D. Novo, "Mapping a multiple antenna SDM-OFDM receiver on the ADRES coarse-grained reconfigurable processor ," in *IEEE Workshop on Signal Processing Systems Design and Implementation*, 2005, pp. 473–478.

[4] K. Jung and H. Lee, "Low-Complexity Multi-Mode Memory-Based FFT Processor for DVB-T2 Applications," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E94-A, no. 11, pp. 2376–2383, Nov. 2011.

GPU based Software DVB-T Receiver Design

Kyu-Hyung Lee and Seo Weon Heo, *Member, IEEE*
School of Electronic and Electrical Engineering, Hongik University, Seoul, Korea
seoweon.heo@hongik.ac.kr

***Abstract*--This paper presents the GPU based software DVB-T receiver design. We first propose the software DVB-T receiver using just one CPU core and investigate the gap between the required time budget and the actual simulation time. Based on the simulation results we partition the algorithm such that some of the algorithm corresponding to the critical path can be processed by the GPU which adopts massively parallel processing elements. To increase the thread usage we design software algorithms which reduce the FFT and Viterbi decoding processing time by the ratio of 10~20 times compared with the CPU based processing.**

Keywords--**GPU, DVB-T, Software modem, SDR**

I. INTRODUCTION

A graphics processing unit (GPU) contains massively parallel processing elements so it has recently been widely applied to various problems such as a signal processing or communication system software design [1-4]. Although it consume large processing power and accompanies latency or not-exactly controlled timing issues, GPU based accelerators provides an increased flexibility and scalability. Recently, GPU based signal processing is widely adopted in the field of wireless communication system design, some of which are in [5, 6].

DVB-T is a European standard for the broadcast transmission of digital terrestrial television signal [7]. This system adopts the OFDM modulation combined with the Reed –Solomon and convolutional coding scheme. Due to both the power limitation and high computation complexity, most of the DTV receivers use dedicated hardware to process the received signal. But recently, research on the software based DTV receiver is active including works using the general purpose GPU accelerator. However, most of the previous works focus on the optimization of some of the specific algorithms such as a FFT, Filtering, LDPC [1-7] and results that show how near we approached in achieving the goal of real time processing using only the software module is not published yet.

In this paper, we first describe the block diagram of the DVB-T receiver briefly and show the processing time of each algorithm block to determine blocks corresponding to the critical path. The results show that FFT and Viterbi decoding algorithm consume significant portion of the processing time so those two algorithms are transferred to and processed by the GPU. Then we describe basic architecture of the GPU. In this paper we adopted the well-known CUDA developed by the NVIDA [8-10]. It contains massively parallel processing elements called as threads. Although it contains a large number of processing elements, it is not an easy work to fully exploit the parallelism due to the limited memory bandwidth and dependency of the processing order. To get more parallelism, we design the receiver software for both the FFT and Viterbi decoding algorithm. The simulation results show that we can reduce the processing time by the ratio of 10~20 times compared with the CPU based processing.

II. DVB-T RECEIVER ARCHITECTURE AND PROCESSING TIME

In Fig. 1, the block diagram of the DVB-T receiver is shown. The baseband signal from the RF demodulator is converted to the digital signal with fixed frequency clock, for example 27 MHz reference clock. Then using the timing information from the CP (cyclic prefix) autocorrelation, the input signal is re-sampled by the Farrow filter to synchronize the sampling timing with the transmitter. Using the frequency offset information from either the CP or from the continuous pilot signal, the re-sampled signal is de-rotated, followed by the low-pass filtering to remove the adjacent channel signal. Then the time-domain signal is converted to the frequency domain signal by the FFT block. DFT based channel estimation algorithm is applied to compensate for the channel frequency response. To disperse the fading channel effect, symbol or bit level interleaving is applied followed by the Viterbi decoding. Finally byte deinterleaver and Reed-Solomon decoder outputs the error-corrected transport packet stream.

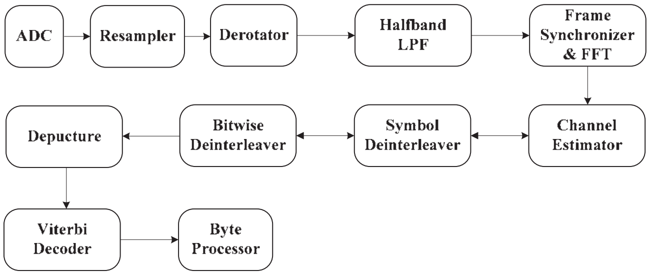


Fig. 1. DVB-T receiver block diagram.

The duration of a single OFDM symbol is 280 μsec in DVB-T receiver system in 2K-mode and 8 MHz bandwidth. To estimate the computation time using high performance CPU (INTEL I7 2600K 3.4GHz) and determine the block corresponding to the critical path, we programmed the receiver software using C language and measured the processing time of each block. The results are shown in Table.1. As shown in the table, the total processing time is

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012-0001368).

approximately 4.55 msec. Even though we consider the fact that only a single CPU core is used among the quad-core, the result is far beyond the time budget of 0.28 msec. So we decided to adopt the GPU which has massively parallel processing elements. Among several receiver blocks, FFT (notice that the DFT based channel estimator uses FFT and IFFT processing) and Viterbi decoder blocks seems to take significant portion of the processing time so those two blocks are transfer to and processed by the GPU.

Table. 1. DVB-T receiver software simulation time.

BLOCK	Processing time (μsec)
Resampler	208.136
Derotator	175.011
Halfband LPF	253.706
Frame synchronizer & FFT	785.287
Channel estimator	1300.405
Symbol deinterleaver	10.807
Bitwise deinterleacer	21.313
Depucture	7.804
Viterbi decoder	1717.787
Byte processor	66.942
Total	4547.198

III. GPU ARCHITECTURE

The CUDA introduced by NVIDIA is a programming environment for writing and running general-purpose applications on NVIDA GPU. CUDA allows programmers to develop GPU applications using extended C programming language instead of graphics API. In CUDA, threads are organized in a hierarchy of grids, blocks, and threads, which are executed in SIMT (Single Instruction Multiple Thread) manner. Threads are virtually mapped to an arbitrary number of streaming multiprocessor (SM). In CUDA, complex memory hierarchy is deployed where each of component memory has different scope and bandwidth so it is critical to decide the proper thread/block/grid partitioning and memory management to exploit the full usage of the massively parallel processing elements.

The GPU Hardware architecture is described in Fig.2. GPU has 256 cores and 8 streaming multi-processors (SM). Each SM has 32 cores. GPU's Memory architecture is described in Fig.3. Each SM has 32,768 registers and single shared memory. The Shared memory is L1 cache and composed of 32 banks. Global Memory is DRAM which is shared among all the SMs. GPU specification used in the simulation is shown in Table. 2.

Table. 2. CUDA GPU specification.

Number of cores	256
GPU clock	900 MHz
Memory clock	1700MHz
Memory interface	GDDR5
Memory interface width	256-bit
Memory Bandwidth	100.9GB/sec
Register / Block	32768
Shared Memory / Block	48KB
L2 Cache Size	384KB

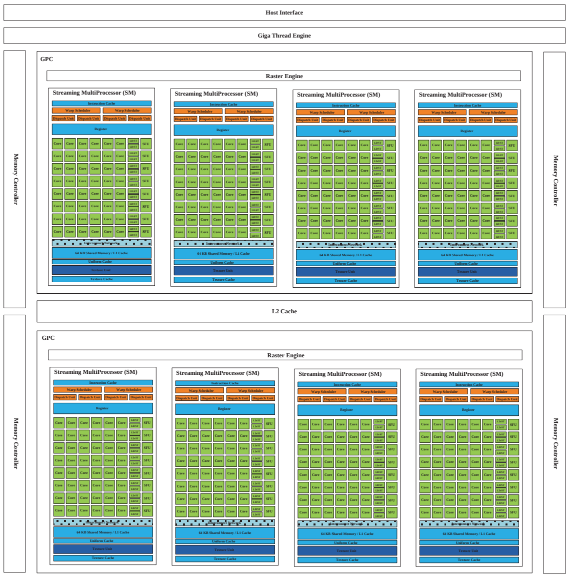


Fig. 2. GPU hardware architecture.

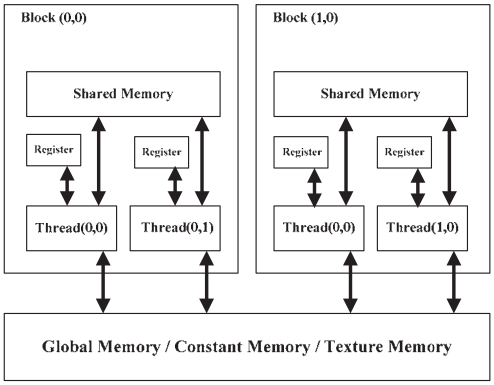


Fig. 3. GPU memory architecture.

IV. DESIGN OF FFT AND VITERBI DECODER ALGORITHM

A. Parallel radix-2 FFT algorithm with CUDA

In order to optimize FFT algorithm, we use 3 methods which are coalescing, streaming and using shared memory. First, global memory loads and stores by threads of a warp are coalesced by the device into as few as single transaction when they access consecutive memory addresses. Second, applications manage concurrency by streaming. A stream is a sequence of commands that execute in order. Different streams, on the other hand, may execute their commands out of order. By streaming we make the GPU processes data while we transfer data from the CPU to GPU simultaneously. Shared memory is much faster than global memory and threads in a SM can access them concurrently. However, due to the limited size and bank structure and considering the fact that cache memory is used in the global memory processing, the proper partitioning of the global and shared memory usage is crucial in reducing the processing time.

First we generate N threads where N is the FFT size. Two threads execute a single butterfly operation as shown in Fig. 4 and a single thread access two global memory addresses. If two threads involved in a butterfly operation access the same global memory address, the access speed is reduced because one thread must wait while the other one accesses the data in the same global memory. To avoid this, upward butterflies are computed first, and then downward butterflies are computed.

FFT processing is composed of multiple stages of butterfly operations (11 stages for 2K-FFT processing). From stage1 to 5, we locate data in a global memory since memory coalescing can be easily applied and shared memory bank conflicts happen as shown in Fig. 5. From stage 6 to 10, we locate data in a shared memory since it is much faster than the global memory and we can avoid the bank conflict as shown in Fig. 6. For example, thread number 0 accesses index number 0 and 32 in shared memory. These data are located in bank 0. So, thread number 0 van access two data at once. In stage11, it is operated in global memory again, because sharing of data between blocks is only possible in global memory and the amount of data in one block is larger than maximum capacity in shared memory.

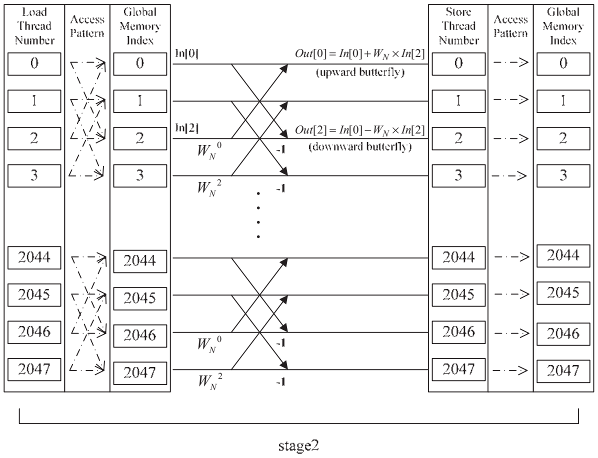
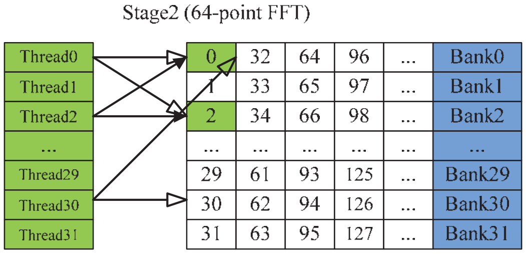
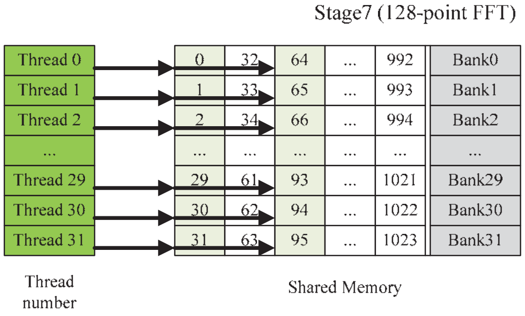


Fig. 4. Threads access and store data scheme & Definition of upward and downward butterfly.



Thread0	0	32	64	96	...	Bank0
Thread1	1	33	65	97	...	Bank1
Thread2	2	34	66	98	...	Bank2
...	...	...	...	...	...	...
Thread29	29	61	93	125	...	Bank29
Thread30	30	62	94	126	...	Bank30
Thread31	31	63	95	127	...	Bank31

Fig. 5. The occurrence of bank conflict in stage2.



Thread 0	0	32	64	...	992	Bank0
Thread 1	1	33	65	...	993	Bank1
Thread 2	2	34	66	...	994	Bank2
...	...	...	...	...	...	...
Thread 29	29	61	93	...	1021	Bank29
Thread 30	30	62	94	...	1022	Bank30
Thread 31	31	63	95	...	1023	Bank31

Fig. 6. Data access structure in stage7.

B. *Parallel Viterbi algorithm with CUDA*

DVB-T system use rate 1/2, 64 states convolution encoder with puncturing operation. For the Viterbi decoding, the previous stage processing should be finished before the next stage processing where one stage is composed of 64 ACS (accumulate compare and select) operations. If we assign a thread to each ACS operation we cannot fully exploit the large number of threads in a GPU. To increase the parallelism, we adopt the well-known sliding block method [11] combined with the parallel ACS unit. Sliding block method divides input sequence into sequence of blocks.

The length of the input sequence gets changed by modulation and code rate. Block length is decided by a common multiple of all of the input sequence length which is 504. In the sliding block method, we have to overlap blocks since the initial decisions are unreliable due to the lack of previous histories. According to reference [12] we need to decide correct outputs after about five times the constraint length of stages are processed. There are 3 conditions to decide the overlap length. First one is improving accuracy, second one is reducing the amount of computations, and finally it has to be the factor of block length. So, 36 is an appropriate number that satisfies above conditions. 64 ACS units operate in parallel in a block and several sliding blocks operate in parallel again. See Fig. 7 and Fig. 8 which shows the thread partitioning and sliding block operation.

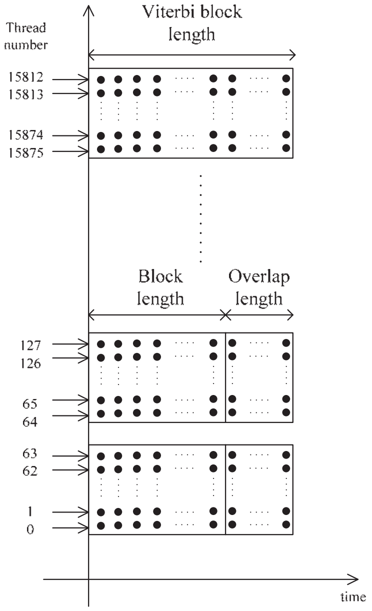


Fig. 7. Thread partitioning for the sliding block Viterbi decoding.

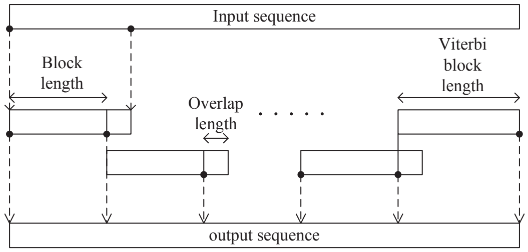


Fig. 8. The procedure of sliding block method.

V. SIMULATION RESULT

A. FFT simulation Result

We simulate 2K-point FFT and 8K-point FFT. A 2K-point FFT unit with CUDA is processed in 12.45 μsec including the data copy time. Computation time is 2.529 μsec. In CPU based code, 2K-point FFT processing time is 270.488 μsec. An 8K-point FFT performs in 32.466 μsec. Computation time is 6.24 μsec. The result is shown in Table. 3. Parallel FFT algorithm is faster about 22 times in GPU than CPU based code.

Table. 3. FFT elapsed time comparison (GPU's case include data copy time).

N-point	CPU processing time(μsec)	GPU processing time(μsec)	CUFFT Library processing time(μsec)
2048-point	270.488	12.45	22.246
8192-point	930	32.466	56.128

Designed Radix-2 FFT achieves up to 44.539Gflops in 2K-point FFT and 85.333Gflops in 8K-point FFT. CUFFT on 1 batch size achieve 9.591Gflops in 2K-point FFT and 30.199Gflops in 8K-point FFT. Designed FFT algorithm is faster than CUFFT provided by NVDIA.

B. Viterbi decoder simulation result

We simulated Viterbi decoder under 64-QAM modulation and 7/8 code rate. Parallel Viterbi decoder by the GPU processing takes 158.546 μsec in 2K-mode. On the contrary, Viterbi decoder by the CPU takes 1717.787 μsec in 2K-mode. Parallel Viterbi decoder speed by GPU is 11 times faster than the one by CPU. Trace back unit is computed in serial. So, processing speed gets lowered. If trace back process is handled by CPU, the unit can be processed a little more quickly. The result is shown in Table. 4.

Table. 4. Viterbi decoder processing time comparison.

Block name (2K-mode)	CPU processing time(μsec)	GPU processing time(μsec)	Improvement performance
Viterbi decoder	1717.787	158.546	About 11 times

VI. CONCLUSION AND FUTURE WORKS

This paper presents the GPU based software DVB-T receiver design. We process the FFT and Viterbi decoding, which occupy a significant portion of the total processing time, in GPU. After the algorithm optimization we reduced the processing time by approximately 22 and 11 times lesser than CPU processing in the case of FFT and Viterbi decoding, respectively. Our ultimate goal is to make it feasible to process the whole DVB-T system in real time only by the software processing. To achieve this, we are planning on implementing Resampler, Halfband LPF and Derotator block in CUDA.

REFERENCES

[1] Y. Chen, X. Cui, and H. Mei, "Large-scale FFT on GPU clusters," *in Proceedings of the 23rd International Conference on Supercomputing*, 2010.

[2] Zhao Lili, Zhang Shengbing, Zhang Meng and Zhang Yi, "Streaming FFT asynchronously on graphics processor units," *Information Technology and Applications (IFITA)*, vol. 1, pp. 308-312, July 2010.

[3] R. de Beer, D. van Ormondt, "Accelerating batched 1D-FFT with a CUDA-capable computer," *Imaging System and Techniques (IST)*, pp. 446-451, July 2010.

[4] Nicholas Hinitt and Taskin Kocak, "GPU-based FFT computation for multi-gigabit wireless HD baseband processing," *EURASIP Jounal on wireless communications and Networking*, vol. 2010, June 2010.

[5] G. Wang, M. Wu, Y. Sun and J. R. Cavallaro, "A massively parallel implementation of QC-LDPC decoder on GPU," *2011 Application Specific Processors Symposium*, pp.82-85, 2011.

[6] M. Wu, Y. Sun, S. Gupta, and J. Cavallaro, "Implementation of a high throughput soft MIMO detector on GPU," *Journal of Signal Processing Systems*, vol. 64, no. 1, pp. 123-136, 2011.

[7] L. Vangelista, N. Benvenuto, S. Tomasin, C. Nokes, J. Stott, A. Filippi, M. Vlot, V. Mignone, and A. Morello, "Key technologies for next-generation terrestrial digital television standard DVB-T2," *IEEE Communization Magazine*, vol. 47, no. 10, pp. 146-153, Oct. 2009.

[8] NVIDIA corp. , "CUDA C Programming Guide 4.1," 2012

[9] NVIDIA corp. , "Nvidia CUDA Programming guide 4.1," 2012

[10] NVIDIA corp. , "Nvidia CUDA CUFFT Library," Feb. 2011.

[11] P. J. Black and T. H.-Y. Meng, "A 1-Gb/s, four-state, sliding block Viterbi decoder," *IEEE Journal of solid-state circuits*, vol.32, no. 6, pp. 797-805, June 1997.

[12] Andrew J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *Information Theory, IEEE Transaction on*, vol. 13, pp. 260-269, April 1967.

A Novel QR Code Guided Image Stenographic Technique

Md. Wahedul Islam, Student Member, IEEE, and Saif alZahir, Member, IEEE

University of Northern British Columbia, Prince George, BC, Canada

ABSTRACT

A novel DWT-SVD based stenographic technique is presented. A QR-code payload guides the alternation of the singular-values of DWT blocks. This method can be used by smart phones, smart video-cameras for business and consumer applications. This method has high PSNR and SSIM and survived chi-square test with 100% message recovery.

Index Terms — Digital Image Steganography, Steganalysis, DWT, SVD, QR Code

I. INTRODUCTION

Multimedia content and systems are rapidly increasing in the industrial electronics area. Information secrecy of such content is of a high significance. Cryptography comes with different solutions; however, it has been proven that it can be broken by the steady progress of the art [1]. Therefore, considering robust and cheaper alternatives are unavoidable. One possible practical alternative is digital steganography. Steganography is an art of hiding secret data in an innocent looking container called cover data. This cover data may be any digital media such as digital image, audio, movie file etc. Usually the embedded secret data is called payload. Once the payload has been embedded into a cover media it may be transmitted to the receiver or posted in public place from where intended receiver can download it. Multimedia message passing in cell and iPhone are getting more popular day by day and sending secret message with stego-image would be an interesting addition.

Steganography algorithms have been developed using the different digital image file format. In the special domain, the most popular approach is the least significant bit (LSB). There exist several variations of this approach [2]. Chung et al. [3] for instance, proposed a singular value decomposition and vector quantization based image steganography with 37.002 dB PSNR. However, spatial domain steganography is venerable to blind steganalysis, meaning easily detected by statistical analysis such as chi-square test. On the other hand, DWT based methods are still in its infancy for steganography. Chen et al. [4] proposed a DWT based image steganography scheme where they embed their secret message in the high frequency components of the DWT using 2 LSB substitutions with wavelet coefficients of *LH*, *HL*, and *HH* Subbands. They obtained stego-image with PSNR difference in the range 39.0033 dB to 54.94 dB, however, they did not report any steganalysis on their method. Driskell et al. [5] achieve high image fidelity using Daubechie wavelet filter by substituting wavelet coefficients that fall below a threshold with coded letters. They also did not report any steganalysis on their method. The literature is scarce on methods that combine DWT-SVD on image steganography and has no QR-code steganographic methods.

II. QUICK RESPONSE (QR) CODE

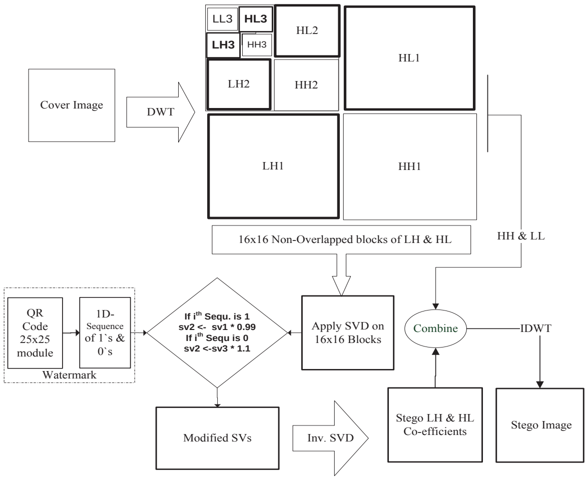
QR code was introduced by Denso Wave in 1994. It is a 2-D code with control points that makes it easier to be interpreted by scanning equipment such as smart phones, digital camera and hand held scanner. Moreover, QR code error correction capability makes it ideal for steganography. For different version of QR code, there are different module configurations where modules refer to the black and white dots which construct the QR Code. The largest standard QR Code is V-40 symbol, which is 177x177 modules in size and can hold up to 4296 characters of alphanumeric data. For more information on QR code, you may refer to: http://www.denso-wave.com.

III. PROPOSED METHOD

A steganographic technique using DWT and SVD transform is proposed. We use a QR code generator to produce a payload (secret message) which is converted to one dimensional vector with a sequence of 1's and 0's. To embed the payload in DWT subbands (especially the *LL* subband), a degradation of the quality of the image is imminent. Hence, we chose to decompose the *LL* sub-band up to three levels to capture the mid-range frequency elements in those sub-bands (i.e., *HL* and *LH*). *HH* typically contains more edge related information of the image, then, the *HL* and *LH* pair is a better subband selection. Each subband *LH* and *HL* are divided into a numbers of non-overlapping 16x16 blocks. There is evidence that any modification on image has less effect on 1st few SVs of the SVD decomposed image [6]. In this research we used the 1-D sequence of the QR code to alter the values of σ_2 with either σ_1 or σ_3 according to the polarity (1 or 0) of the QR code sequence. Fig 1 is the flow chart of the proposed technique. The SVD of a matrix A_{mxn} is given in equation 1 below.

$$A = USV^T, \qquad (1)$$

where *U* and *V* are the left and right singular vectors of size $m \times m$ and $n \times n$ respectively and $S = diag\ (\sigma 1, \sigma 2, .., \sigma r)$ with size $m \times n$. The diagonal matrix *S*, has rank *r* equal to the rank of *A*. The nonzero elements are called singular values of A and are in descending order. The singular values (SVs) are the square roots of the eigenvalues of $A^T.A$ or $A.A^T$. The proposed method can be used in iterative manner, meaning, if we cannot recover the exact payload in the first time, we re-implement the process on stego-image until we obtain the payload. We take the stego-image and apply the following process again to ensure 100% extraction of the payload.



LL3
HL3
LH3
HH3
HL2
LH2
HH2
HL1
Cover Image
DWT
LH1
HH1
16x16 Non-Overlapped blocks of LH & HL
HH & LL
QR Code 25x25 module
1D-Sequence of 1`s & 0`s
Watermark
If i^th Sequ. is 1 sv2 <- sv1 * 0.99 If i^th Sequ is 0 sv2 <-sv3 * 1.1
Apply SVD on 16x16 Blocks
Combine
IDWT
Modified SVs
Inv. SVD
Stego LH & HL Co-efficients
Stego Image

Fig 1 Block Diagram of Payload embedding

The extraction process can be formulated as follows:

1. Apply up to 3rd level DWT to the stego image I_g.
2. Apply SVD on each 16x16 block of $LH_{1\text{-}to\text{-}3}$ and $HL_{1\text{-}to\text{-}3}$ subband of image I_g.
3. Find the ratio between the 1st and 2nd singular value of each block of I_g. Extracted values are then map into extracted one dimensional payload using the following equation:

$$\mathrm{EP_i} = \begin{cases} 1 & \text{if } \frac{\sigma_1}{\sigma_2} > T \\ 0 & \text{if } \frac{\sigma_1}{\sigma_2} < T \end{cases}, \qquad (2)$$

where the value of T (threshold) value is in the range 1.1. Finally one-dimensional $\mathrm{EP_i}$ is converted into 25x25 matrixes to construct the version-2 QR code.

IV. RESULTS AND DISCUSSION

In order to validate the image fidelity, and steganalysis of our steganography method, we have performed extensive simulation with MATLAB 7.9 on image database of Signal and image processing Institute (SIPI) of University of Southern California of which (137 images) we included only 6 test image's result. The QR code we have used as payload is shown in fig. 2 below.

QR-Code	Message
	UNBC Computer Science

Fig. 2 Version 2 QR Code with left-side message.

Table 1 shows the PSNR in dB and SSIM values of the very well known test images after embedding payload using the proposed method. These results show that the stego-image is always greater than 30 dBs. Moreover, we also achieved more than 0.94 SSIM values which is more accurate measurement than PSNR. However, in blind-steganography, the receiver does not have the original cover image, hence; measuring PSNR between stego and cover images would not be possible. Only for 'Lena' image two iterations are needed to extract the embedded payload (QR code).To further verify that our proposed method can withstand a Chi-square attack, we used a Chi-square steganography test program by Guillermito [7] to perform steganography analyses. It has been always observed close to 0 for all stego-images, so the probability for a random embedded message is low. In other words, nothing is hidden in our stego-image.

Table 1 SSIM and PSNR of stego-images

Image	**SSIM index**			**PSNR in dB**		
Iteration # ->	1st	2nd	3rd	1st	2nd	3rd
Baboon	0.994	0.994	0.994	40.31	40.31	40.31
Barbra	0.977	0.977	0.977	35.84	35.83	35.83
Lena	**0.984**	0.981	0.981	**40.52**	38.81	38.72
Peppers	0.956	0.955	0.955	33.56	33.47	33.45
F16	0.944	0.943	0.943	33.25	33.18	33.17
Boat	0.979	0.979	0.979	36.24	36.17	36.16
Average	**0.972**	**0.971**	**0.971**	**36.62**	**36.29**	**36.27**

V. COCLUSIONS

A novel QR code guided DWT-SVD steganographic technique for consumer and business applications is presented. We altered the 2nd SV of DWT blocks' values in such a way that it does not change the stego-image quality and it does not violate the SVs ($\sigma_1, \sigma_2, \ldots, \sigma_r$) order. Also, we have introduced QR-code as payload to guide the alteration of 2nd SV. We exploited the QR code self-correcting capability of up to 7% to reach 100% message recovery. Our simulation results for the SIPI (137 images) have similar SSIM and PSNR as those in Table 1. Our message can be hidden in the images in more than 85% of the cases, and once it is embedded, it is 100% recoverable.

REFERENCES

[1] B. Li, J. He, J. Huang, Y. Q. Shi, 'A Survey on Image Steganography and Steganalysis', Journal of Information Hiding and Multimedia Signal Processing, vol. 2, no. 2, 2011, pp 142-172.

[2] A. Cheddad, J. Condell, K. Curran, P. M. Kevitt, 'Digital Image Steganography: Survey and Analysis of current methods', Signal Processing, Elsevier, 90(2010), pp 727-752

[3] K. L. Chung, C H Shen, L. C. Chang, "A novel SVD and VQ-based image hiding scheme", Pattern Recognition Letters, vol. 22, 2001, pp 1051-1058.

[4] P. Y. Chen, and H. J. Lin, "A DWT based approach for image steganography", International Journal of Applied Science and Engineering, vol. 4, no. 3, 2006, pp. 275-290.

[5] L. Driskell , "Wavelet based steganography", Cryptologia, Taylor & Francis, 28:2, 2010, pp. 157-174.

[6] M. W. Islam, S. alZahir, "A robust color image watermarking scheme", in IASTED Intl. Conf. on Visualization, Imaging, and Image processing, VIIP 2012, Banff, Canada, July 3-5, 2012.

[7] http://www.guillermito2.net/stegano/tools/index.html

Live Video Streaming with Adaptive Pre-Processing by Using Scalable Video Coding

Dan Grois[1], *Senior Member, IEEE,* Ofer Hadar[1], *Senior Member, IEEE,*
Rony Ohayon[2] and Noam Amram[3]

Abstract — *In this work, a novel live scalable video streaming scheme, with adaptive pre-processing (pre-filtering), is presented. The scheme employs Scalable Video Coding (SVC), which is an extension of the H.264/AVC. An adaptive pre-filter is provided for each SVC layer, while each pre-filter's parameters, such as the standard deviation and kernel matrix size, are dynamically adjusted according to the varying network conditions, thereby enabling to continuously obtain an optimal visual presentation quality at the decoder end. The performance of the presented live scalable video streaming scheme is evaluated and tested in detail, thereby demonstrating significant improvements of more than 2dB.*

Index Terms — Scalable Video Coding (SVC), live/real-time video streaming, adaptive pre-processing/pre-filtering, high-quality visual presentation.

I. Introduction

Much of the attention in the field of video adaptation is currently directed to the Scalable Video Coding [1], which is an extension of the H.264/AVC standard [2]. A major requirement for the Scalable Video Coding is to enable encoding of a high-quality video bitstream that contains one or more subset bitstreams, each of which can be transmitted and decoded to provide video services with lower temporal/spatial resolutions, or reduced fidelity, while retaining the reconstruction quality that is highly relative to the rate of the subset bitstreams [1]. As a result, the SVC enables to provide important functionalities, such as the variable spatial formats and power adaptation [1]. These functionalities lead to dramatic enhancements of video streaming applications. In addition, in order to support the SVC streaming [3], a new payload format for the Real-time Transport Protocol (RTP) was recently specified by the IETF [4]. However, the SVC video stream quality is usually significantly reduced under varying network conditions [1].

The most recent works [5] in the SVC field enable to improve the Region-of-Interest (ROI) video presentation quality by providing adaptive pre-filtering schemes. However, these works do not present any solution with regard to the live (real-time) SVC streaming, and particularly, they do not consider any change in the varying network conditions.

In the next section, a detailed description of the proposed scheme is presented, followed by experimental results and conclusions.

II. Proposed Live SVC Streaming Scheme

In the proposed live SVC adaptive video streaming scheme, which is schematically illustrated in *Fig. 1*, the adaptive pre-filtering is performed for each SVC layer according to the varying network conditions, such as the varying network load that indicates a change in the available bandwidth. It is noted that for simplicity, *Fig. 1* presents a spatial SVC scheme, in which only two layers (the Base-Layer and one Enhancement Layer) are encoded and then transmitted over the RTP protocol.

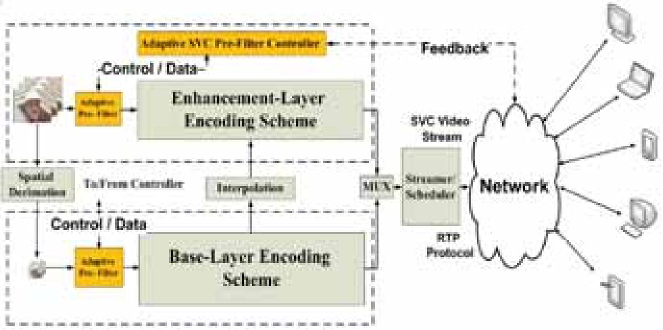


Fig. 1. Proposed live scalable video streaming scheme with adaptive pre-filtering under varying network conditions.

As seen in *Fig. 1*, an adaptive pre-filter is provided for each SVC layer for enabling to optimally vary each filter's parameters, such as the standard deviation and kernel matrix size, by means of the Adaptive SVC Pre-Filter Controller (that can be provided, for example, within a Video Controller of [6]), which in turn continuously receive a feedback from the network with regard to the varying network conditions, such as the Round-Trip-Time (RTT) that is presented in *Fig. 2* below (or as described in [7]). The filters that are used in the proposed scheme are Gaussian filters due to their relatively low computational complexity (when comparing, for example, to Wiener or Wavelet filters) in term of the CPU processing time, i.e. CPU clocks [5].

Also, it should be noted that by using the proposed adaptive pre-filtering scheme, better video quality can be obtained at the decoder side, when comparing for example to state-of-the-art techniques of the adaptive quantization parameter change [2].

III. Experimental Results

The test conditions for evaluating the presented live SVC adaptive video streaming scheme are as follows: spatial resolution of the Base-Layer (*Layer 0*) is QCIF, and of the Enhancement Layer (*Layer 1*) is CIF; frame rate is 30 fp/sec for each layer; a number of coded frames is 300. The tests were carried out on computers with Intel Core i3 CPU, 2.3 GHz, 4GB RAM with Windows® 7 OS, Service Pack 3.

The standard deviation (STD) and kernel matrix size values of each Gaussian pre-filter are adaptively updated according to the continuously measured RTT and PSNR values, as presented for example, in *Figs. 2* and *3* below.

This work was partially supported by the NetHD consortium, MAGNET Program of the Israeli Chief Scientist, Israeli Ministry of Trade and Industry under Grant 87190911 (http://www.nethd.org.il).

Also, this work was partially supported by the MEDIEVAL project (Multimedia Transport for Mobile Video Applications) of the 7th Framework Program of the European Commission (Ref.: 258053).

[1] Dr. D. Grois and Dr. O. Hadar are with the Communication Systems Engineering Department, Ben-Gurion University of the Negev, Israel (correspondence e-mail: grois@cse.bgu.ac.il, to Dan Grois). [2] Rony Ohayon is with the Engineering Faculty, Bar-Ilan University, Israel, and [3] Noam Amram is with LiveU® Ltd., Israel.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

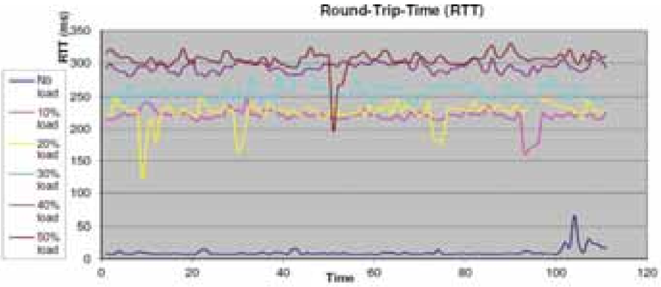
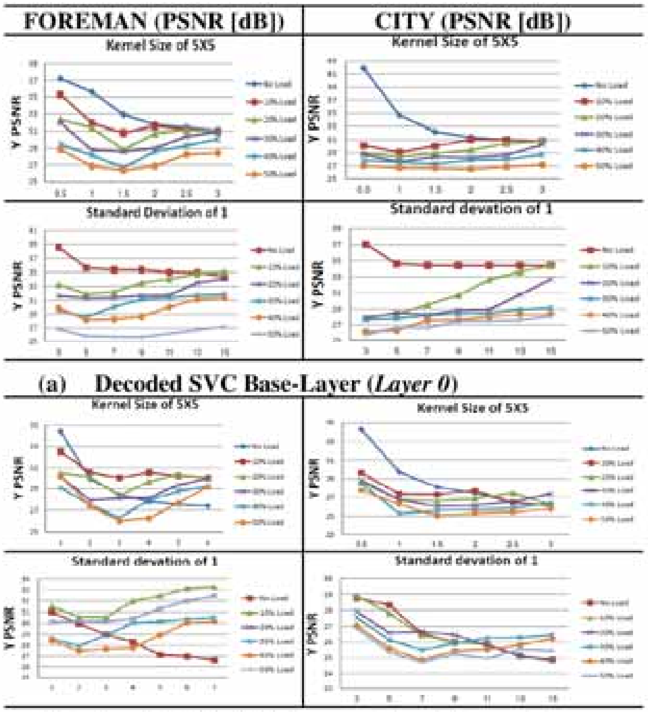


Fig. 2. Measured Round-Trip-Time as a function of the varying network conditions, i.e. the increase/decrease of the network load (i.e. the increase/decrease of the available bandwidth).



(a) Decoded SVC Base-Layer (*Layer 0*)

(b) Decoded Two SVC Layers (*Layer 0+Layer 1*)

Fig. 3. PSNR values of the *FOREMAN* and *CITY* video sequences, transmitted over the RTP protocol: a) SVC Base-Layer; and b) Two SVC Layers, at the decoder end, as a function of the STD and kernel matrix size of each Gaussian pre-filter. For simplicity, various STD values were evaluated, while the kernel matrix size was fixed to 5x5, and on the other hand, various kernel matrix sizes were evaluated, while the STD value was fixed to "*1*".

As seen from *Fig. 3*, by using the proposed live SVC adaptive video streaming scheme, better PSNR values can be achieved under every network load conditions, while adaptively varying each filter's standard deviation and/or kernel matrix size values. For this, a detailed *look-up table* with a plurality of *threshold RTT values* is predefined (not shown due to the paper size limit), thereby enabling to update each filter parameters in real-time, during the video streaming (e.g., enabling to increase the standard deviation value of the *Layer 1* pre-filter by "*0.5*" if the RTT value reaches *100 ms*, and increase by "*1*", if the RTT value reaches *200 ms*, which relates to the *10%* network load conditions).

TABLE I presents an improvement in the PSNR values under the varying network load conditions. As seen from *TABLE I*, when there is no network load, the pre-filtering with any STD or kernel matrix size parameters decrease the video quality. On the other hand, when there is a network load of *10%* or more (which means that the available network bandwidth becomes limited), the video quality can be improved *up to about 2dB* by setting appropriate filter parameters. As a result, by predefining a detailed *look-up table*, the corresponding STD and kernel matrix size values are selected in real-time, during the live video streaming over the RTP protocol (the complexity and delay of accessing the predefined *look-up table* is negligible).

TABLE I
QUALITY CHANGE WHEN USING THE PROPOSED LIVE SCALABLE VIDEO STREAMING SCHEME UNDER THE VARYING NETWORK LOAD; *FOREMAN*, TWO-LAYER SVC (QCIF+CIF), 300 FRAMES.

STD and Kernel Matrix Size	No Load [dB]	10% Load [dB]	20% Load [dB]	30% Load [dB]	40% Load [dB]
0.5 STD	-0.41	**0.13**	-0.37	**0.22**	**0.50**
1 STD	-1.02	-0.44	0.22	-0.91	-0.68
1.5 STD	-1.85	-0.41	-1.02	-0.41	-1.44
2 STD	-2.20	**0.59**	**0.73**	**0.08**	**0.40**
2.5 STD	-2.43	**0.49**	**1.55**	**1.44**	**0.85**
3.0 STD	-2.83	**1.92**	**1.41**	**2.21**	**1.96**
3x3	-0.31	-0.13	-0.43	-0.72	-0.24
5x5	-0.64	-0.44	**0.22**	-0.91	-0.69
7x7	-1.03	-0.36	**0.32**	-0.08	-0.49
9x9	-1.43	**0.66**	**0.52**	**0.99**	-0.36
11x11	-1.95	**1.02**	**1.19**	**1.12**	**0.53**
13x13	-1.98	**1.66**	**1.70**	**1.36**	**1.65**
15x15	-2.16	**1.79**	**2.07**	**1.48**	**1.77**

IV. CONCLUSION

In this work, a novel live SVC video streaming scheme is presented by employing an adaptive pre-processing for each SVC layer in order to improve the SVC video presentation quality under the varying network conditions (such as the varying network load that indicates a change in the available network bandwidth). The performance of the presented scheme was evaluated and tested in detail, thereby introducing significant improvements of more than 2dB.

ACKNOWLEDGMENT

We gratefully thank Tal Ben-Avraham, Itay Bochner, Daniel Kaufman and Guy Shalom for their assistance in testing and evaluation.

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[2] D. Grois, E. Kaminsky, and O. Hadar, "Optimization methods for H.264/AVC video coding," in "The Handbook of MPEG Applications: Standards in Practice", (eds M. C. Angelides and H. Agius), John Wiley & Sons, Ltd, Chichester, UK, 2011.

[3] T. Schierl, C. Hellge, S. Mirta, K. Gruneberg, and T. Wiegand, "Using H.264/AVC-based Scalable Video Coding (SVC) for real time streaming in wireless IP networks," *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on* , pp.3455-3458, 27-30 May 2007.

[4] S. Wenger, Y.-K. Wang, T. Schierl, and A. Eleftheriadis, "RTP payload format for Scalable Video Coding," Internet Draft: draft-ietf-avt-rtp-svc-27.txt, Feb. 2011.

[5] D. Grois, Dan, and O. Hadar, "Efficient adaptive bit-rate control for Scalable Video Coding by using computational complexity-rate-distortion analysis," *Broadband Multimedia Systems and Broadcasting (BMSB), 2011 IEEE International Symposium on* , vol., no., pp.1-6, 8-10 Jun. 2011.

[6] MEDIEVAL Project, Deliverable D2.2., "Final Specifications for video service control," [Online: http://ict-medieval.eu/], Jun. 2011.

[7] MEDIEVAL Project, Deliverable D1.1., "Preliminary architecture design," [Online: http://ict-medieval.eu/resources/Medieval_D1.1_final$5B1$5D.pdf], Jun. 2011.

Exploring Visual Temporal Masking for Video Compression

Velibor ADZIC, Hari KALVA, and Borko FURHT

***Abstract*—In this paper we present work on exploiting visual temporal masking phenomenon applied to video compression. Our results show that it is possible to reduce bitrate of the compressed video sequence without affecting subjective quality and quality of experience (perceptually lossless). The principles we present here are applicable to all modern hybrid coding systems and can be implemented seamlessly with video delivery platforms. Results show up to 6% of additional savings when implemented on top of the state of the art encoder.**

I. INTRODUCTION

Modern video compression algorithms rely in some part on characteristics of human visual system (HVS). However, there are many findings in psycho-visual studies that haven't been explored in the context of video compression applications. One such finding is the phenomenon of temporal visual masking. Visual masking in temporal and spatial domain has been discovered by psychologists more than a century ago [1], [2]. It occurs when the visibility of target stimulus is reduced by the presence of mask stimulus. This paper focuses on temporal masking – particularly backward masking. It is manifested at abrupt scene changes, when new scene masks certain amount of frames from the previous scene. A number of frames that precede a scene change are essentially erased form higher levels of processing in HVS. Subject is unable to consciously perceive these frames. Although scientific community doesn't have clear explanation for this phenomenon, one of the promising explanations for backward masking could be the variation in the latency of the neural signals in the visual system as a function of their intensity [3]. Detailed overview of models and findings in visual backward masking can be found in [4]. Our goal is to explore visual masking phenomena and its potential benefits for video compression.

II. RELATED WORK

Although significant amount of research related to visual masking and signal processing has been done over past years, it is mostly focused on spatial masking for image compression [5], [6]. As far as temporal masking is concerned seminal paper by Girod [7] explores forward masking - showing that there is some form of masking effect immediately after scene change. Tam et al. [8] investigated the visibility of MPEG-2 coding artifacts after a scene cut and found significant visual masking effects only in the first subsequent frame. Carney et al. [9] investigated levels of sensitivity of HVS to blur in the first 100-200 milliseconds (ms) after scene cut. However, in our initial tests we couldn't confirm any usefulness of forward masking, since subjects began to notice distortions even in the first frame after scene cut. Hence, we focused on backward masking trying to achieve better results.

For the reasons that are not apparent, much less work has been done towards application of backward masking to video coding. One reason could be that backward masking requires buffering to detect scene changes and hence not suitable for realtime broadcast applications. Majority of video services over the web today are on-demand services delivering pre-coded video and are best suited to exploit backward masking. Pastrana-Vidal et al. [10] studied the presence of backward and forward temporal masking based on visibility threshold experiments using video material in common intermediate format (CIF) resolution (352x288 pixels). They simulated a single burst of dropped frames near a scene change, for different impairment durations from 0 to 200 ms. The transitory reduction of the HVS sensibility was reported to be significant in the first 160ms for forward masking and up to 200ms for backward masking. Study by Huynh-Thu and Ghanbari [11] also showed that backward masking is more significant than forward masking. They used burst of frozen frames as stimulus and scene cut as mask. These papers, however, do not explore the impact of masking on video bitrate and bandwidth reduction. Our work is the first in which possibilities of bitrate savings based on backward masking are explored.

III. EXPERIMENTS AND RESULTS

Our experiments were aimed at discovering how bitrate can be saved by introducing distortions or impairments in the frames just before scene change. We tested both frame freezing and our proposal (more aggressive quantization). In order to confirm our hypothesis we conducted experiments with sequences obtained using process flow shown in Figure 1. The process is similar to traditional two-pass coding. However, algorithm can be applied immediately without the need for first pass and parsing if we know positions of the scene changes. Such list of I-frames can be supplied as metadata with original sequence. We used "x264" open source encoder to produce H.264 bitstream sequence. Our source dataset contained 20 video sequences with standard definition

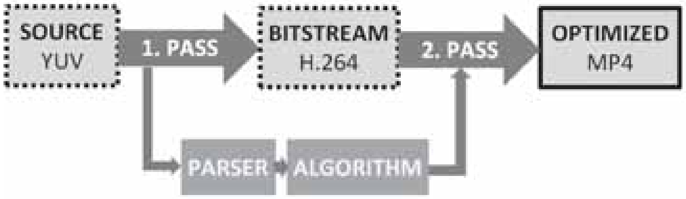


Fig. 1. Flow of the experimental setup.

TABLE I

BITRATE SAVINGS FOR DIFFERENT METHODS COMPARED TO BASELINE

CBR	**SAVINGS**		
(bitrate)	Freezing, N=3	Freezing, N=5	TMFQ, M=10
900kbps	1.54%	3.48%	5.66%
1300kbps	1.53%	3.46%	5.82%
1900kbps	1.54%	3.45%	5.58%

resolution (SD, 720x480p) obtained from DVD sources. Videos are 30 second long clips from popular feature and animated movies and music videos – in general, content that is very popular and generates most of the traffic on the Internet. All videos were presented at 25 frames per second (fps) on the 20 inch monitors, in a setting that complies with ITU-R recommendation BT.500-11. Subjects were 5 students with normal or corrected to normal vision.

Freezing was implemented by repeating last selected frame until the scene change; i.e., a set of N frames immediately before a scene change are replaced by an identical frame. Temporally masked frame quantization (TMFQ) was implemented by raising quantizing parameter (QP) for target window of M frames immediately before a scene change. Last two frames were quantized with QP = 51 (maximum allowed in H.264). For the rest of preceding frames (M-2) we implemented sigmoid-like ramp that gracefully lowered the QP increase.

The first set of experiments showed that freezing can be applied with limited success for frames in the range of 100 – 200 ms before scene change. At least one subject noticed freezing impairments for N ≥ 3. For more than 5 frames impairments were noticeable in 100% of cases. Freezing was reported as being most annoying for the frames at the end of a scene that contain high motion activity.

For second set of experiments we targeted perceptually lossless optimization using TMFQ. We hypothesized that high quantization is not going to impair the whole motion flow, and hence will have higher threshold of noticeability. Subject reports were analyzed in order to find the limit at which there are 0% of reported distortions (perceptually lossless). We were able to achieve this for M = 10 frames before scene cut, using the ramp described earlier. TMFQ allowed for additional distortions in more frames than freezing. Our hypothesis for better results with TMFQ was confirmed. Achieved savings are presented in Table 1 (Freezing with N = 3 and 5 and TMFQ with M = 10). Savings are calculated compared to constant bitrate H.264 coding (CBR). We benchmarked CBR as baseline because it is used in platforms such as adaptive streaming which are reported to contribute the most to video traffic on the Internet.

TMFQ can be implemented together with other techniques, such as optimized adaptive streaming [12] to introduce improved bitrate savings. These savings (~6%) are not negligible given that recent study estimated that the sum of all forms of video will exceed 86% of global consumer traffic by the year 2016 [13].

IV. CONCLUSIONS

Using cues from HVS studies can be very beneficial for modern video compression optimization. Once we reach limits of traditional hybrid coding the most promising path of improvement is further exploration of psycho-visual and perceptual studies. Our experiments have demonstrated that visual temporal masking can be used to achieve savings in bitrate.

Our algorithm can be used in conjunction with all modern video coding standards and on top of popular platforms for the delivery of on demand content (such as adaptive streaming over HTTP). Furthermore, all principles explored in this paper are expected to work on top of future standards (i.e. High Efficiency Video Coding - HEVC).

Algorithm can be implemented for live video streaming in the scenarios which allow short delay. The only information that is needed in advance is the position of scene change.

Implementation of psycho-visual algorithms such as the one that we introduced here can have significant impact on bandwidth usage optimization, given the trend of fast growing video content related traffic on the Internet.

REFERENCES

[1] C.S. Sherrington, “On the reciprocal action in the retina as studied by means of some rotating discs,” *J. Physiology* 21, 1897, p. 33–54.

[2] W. McDougall, “The sensations excited by a single momentary stimulation of the eye,” *Brit J. Psychol* 1, 1904, p. 78–113.

[3] A.J. Ahumada Jr., B.L. Beard and R. Eriksson, “Spatio-temporal discrimination model predicts temporal masking function,” *Proc. SPIE Human Vision and Electronic Imaging*, vol. 3299, 1998, pp. 120–127.

[4] B.G. Breitmeyer and H. Ogmen, “Recent models and findings in visual backward masking: A comparison, review, and update,” *Percept Psychophys* 62, 2000, pp. 1572–1595.

[5] A.N. Netravali and B. Prasada, “Adaptive quantization of picture signals using spatial masking,” *Proceedings of the IEEE* , vol.65, no.4, pp. 536- 548, April 1977.

[6] M. Naccari and F. Pereira, “Comparing spatial masking modelling in just noticeable distortion controlled H.264/AVC video coding,” *11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, 2010 , vol., no., pp.1-4.

[7] B. Girod, “The information theoretical significance of spatial and temporal masking in video signals,” *Proc. SPIE Human Vision, Visual Processing and Digital Display*, vol. 1077, 1989, pp. 178–187.

[8] W.J. Tam, L.B. Stelmach, L. Wang, D. Lauzon and P. Gray, “Visual masking at video scene cuts,” *Proc. SPIE Human Vision, Visual Processing and Digital Display*, vol. 2411, 1995, pp. 111–119.

[9] Q. Hu, S.A. Klein and T. Carney, “Masking of high-spatial-frequency information after a scene cut,” *Society for Informational Display 93 Digest*. n. 24, 1993, p. 521-523.

[10] R.R. Pastrana-Vidal, J.-C. Gicquel, C. Colomes and H. Cherifi, “Temporal Masking Effect on Dropped Frames at Video Scene Cuts,” *Proc. SPIE Human Vision and Electronic Imaging IX*, vol. 5292, 2004, pp. 194-201.

[11] Q. Huynh-Thu and M. Ghanbari, “Asymmetrical temporal masking near video scene change,” *ICIP 2008. 15th IEEE International Conference on Image Processing*, vol., no., pp.2568-2571.

[12] Adzic, V.; Kalva, H.; Furht, B.; , "Optimizing video encoding for adaptive streaming over HTTP," *Transactions on Consumer Electronics, IEEE* , vol.58, no.2, pp.397-403, May 2012.

[13] Cisco, “Visual Networking Index Services Adoption (VNI SA) Forecast, 2011-2016,” *Whitepaper*, 30 May 2012.

Context Adaptive Block Scan for Video Coding

Hyun-Soo Kang[1], Si-Woong Lee[2], and Yun-Ho Ko[3]

Abstract--**A new coefficient scanning method using context adaptive block scan (CABS) is presented. A number of additional scan patterns for entropy coding are employed in addition to the conventional zigzag scan. The scan pattern is adaptively selected per 4X4 block basis. In order to avoid any increment of overhead bits, the scan pattern for each block is determined in the context adaptive manner.**

I. INTRODUCTION

Scanning of a two dimensional array $C_{m,n}=\{c(i,j):1\leq i\leq m, 1\leq j\leq n\}$ is an invertible mapping function from $C_{m,n}$ to a one-dimensional array $\{c(k):k=1,\ldots,mn\}$. By the large number of possible scan patterns, the scan methodology has been intensively studied for image and video encryption [1][2].

For coding applications, a scan process is also requisite to rearrange a two-dimensional array of transform coefficients into a one-dimensional sequence to make the run-length symbols for the VLC. In H.264 video coding, from the coefficients arranged into a one-dimensional sequence, those symbols including TotalCoeff (number of non-zero coefficients), TrailingOnes, trailing_ones_sign_flag, levels, total_zeros, and run_before are determined, and then encoded according to the corresponding VLC tables when the entropy coding mode is CAVLC [3][4]. Except TotalCoeff, other symbols are strongly dependent on scan order, thus different scan patterns result in different compression ratio. Among many possible scan patterns, the effective one for bit saving is that which strings the significant non-zero coefficients as continuously as possible prior to the zero coefficients. This implies that the scan pattern should be adaptively decided according to the distribution pattern of non-zero coefficients in the block. Nonetheless, just zigzag scan is used in the H.264 video coding, which limits coding performance.

This paper proposes an adaptive coefficient scanning method where a set of scan patterns is employed for the CAVLC of the H.264 video coding. It is important to mention that the proposed method uses context adaptive approach in determining the scan pattern for each block, meaning that no additional overhead bit is needed to be transmitted.

II. PROPOSED ALGORITHM

A. Scan patterns

Fig.1 shows the five scan patterns used in the proposed method. They are empirically selected based on the relative frequency with which each pattern is selected as the optimal scan pattern in our experiments using diverse test sequences. They are named zigzag scan, anti-zigzag scan, field scan, H scan, and V scan, respectively.

The anti-zigzag scan is diagonally symmetric to zigzag scan. The field scan which is used for coding of field macroblocks in H.264 [5] is adapted as one of the candidate scan patterns, since it is efficient for the blocks where the significant coefficients are mostly distributed in the vertical direction. To cope with the residual blocks from the horizontal and vertical directional prediction in the intra coding, H and V scans are also included. It is important to note that more patterns can be employed if necessary at no additional cost, since the proposed method needs no overhead bits for scan patterns.

B. Context adaptive scan pattern decision

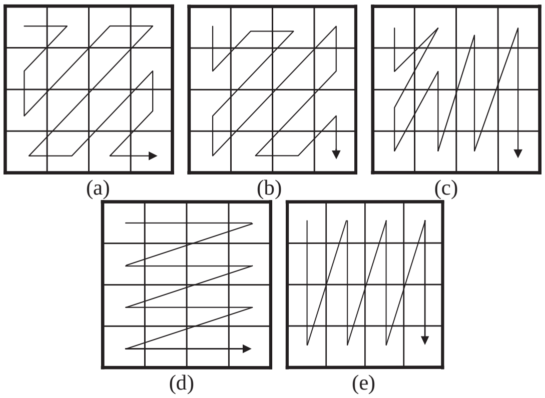


Fig. 1. Scan patterns used in CABS (a) zigzag scan, (b) anti-zigzag scan, (c) field scan (d) H scan, (e) V scan

The optimal scan pattern for a given block can be determined by applying each pattern to it and selecting the one with the minimum number of generated bits. However, this kind of solution requires overhead bits, since the information signaling the scan pattern should be transmitted to the decoder. Assuming 4 scan patterns and fixed length coding, 2 bits for each block is necessary, which results in a severe increment of the overhead bits and thus offsets the gain obtained by the adaptive coefficient scanning.

In order to overcome this problem, the context adaptive approach which utilizes the information of already coded neighboring blocks, i.e., the upper block (*U*) and the left block (*L*), in the scan-pattern-decision process is proposed in this paper. In the respect of optimality, the scan pattern determined by the proposed CABS method is suboptimal, since the selected pattern is optimal not to the current block but to the reference neighboring block. Nonetheless, if there is sufficient correlation in the optimal scan patterns between the current block and the neighboring reference block, which is a general

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012R1A1A4A01014227).

[1] College of ECE, Chungbuk National University, Korea. [2] Dept. of ICE, Hanbat National University, Korea. [3] Corresponding author, College of Eng, Chungnam National University, Korea.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

case, the high coding gain can be achieved with the merit of no additional overhead bits. The overall algorithm is as follows:

① For each 4X4 block, compute the absolute differences of the number of non-zero coefficients between the current block (*X*) and the neighboring blocks by

$$D(L) = \left| TotalCoeff\,(X) - TotalCoeff\,(L) \right|$$

$$D(U) = \left| TotalCoeff\,(X) - TotalCoeff\,(U) \right|$$

② If *D*(*L*)≤*D*(*U*), the left block is marked as the reference block, and vice versa. By comparing the number of nonzero coefficients, more correlated one with the current block can be selected as the reference.

③ Apply 5 scan patterns in Fig. 1 into the reference block, and find the one having the minimum number of generated bits.

④ Scan the transform coefficients of the current block with the scan pattern found in step 3.

Since the scan pattern is determined using the already coded reference block, the decoder can identify it without any overhead bit. Note that the proposed adaptive scan method does not apply to the 4X4 DC block for 16X16 intra prediction for which only zigzag scan is used.

III. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, we implemented the CABS within the reference software JM10.1, and compared the results with that of the original JM10.1. In the experiment, four test sequences of Coastguard, Carphone, Container and Table with QCIF format were used. The encoding parameters are as follows: number of frames-100, frame rate-30Hz, rate control off; RD optimization on; entropy coding mode-CAVLC; intra slice coding only.

In order to prove the close correlation in the optimal scan directions between the current block and the reference one, we measured the matched ratio that is computed by dividing the number of the matched blocks where the optimal scan pattern for the current block is identical with that of the reference block with the total number of coded blocks. The averaged value over four different QP values of 22, 27, 32 and 37 are shown in Table 1, which is in the range of 48 to 69%. Considering the five scan directions, this result proves high spatial correlation in the optimal scan patterns as expected.

To evaluate the coding efficiency of the proposed method, we measured the PSNR and bitrates of the two methods. Table 2 shows the performance comparison between the proposed method and the conventional JM10.1. The results are summarized as follows:

i) In terms of PSNR, both methods show almost the same quality. Small discrepancy in PSNR is due to the coding mode change in some blocks. By altering the scan pattern, the variation in rate has an effect on the RD optimized mode decision process, which results in the modification of block modes. But as shown in Table 2, it is negligible.

ii) In terms of bitrates, the proposed method gives superior results over the JM10.1 for all sequences. The percentage of bit saving is in the range of 0.6 to 2.3% as shown in the last column of Table 2. This amount of additional bit saving in total bitrates can be regarded as a significant result, since it is achieved by just incorporating the simple scan pattern adaptation process.

IV. CONCLUSIONS

This paper presented an adaptive block scanning method. To prevent any increment of overhead bits, the context adaptive approach is employed in scan pattern decision process. Experimental results demonstrate that the proposed method has a better performance in bit reduction than the conventional zigzag scan. Since no overhead bit is required, additional bit reduction can be expected by employing more efficient scan patterns which is remained as a further work.

REFERENCES

[1] S.S.Maniccam and N.G.Bourbakis, “Image and video encryption using SCAN patterns,” *Pattern Recognition*, vol. 37, no. 4, 2004, pp.725-737

[2] N. Bourbakis and C. Alexopoulos, “Picture data encryption using scan patterns,” *Pattern Recognition*, vol. 25, no. 6, 1992, pp.567-581

[3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, 2000, pp. 560-576.

[4] ITU-T Recommendation H.264|ISO/IEC 14496-10 AVC, “Draft text of final draft standard for advanced video coding,” 2003

[5] B. Jeon, W. Choi, and J.H.Park, “Alternate scan for interlaced video coding,” JVT-D037, Geneva, 2002

TABLE I MATCHED BLOCK RATIO

Sequence	Matched block ratio(averaged)
Coastguard	47.70 %
Carphone	68.92 %
Table	62.10 %
Container	56.96 %

TABLE II PERFORMANCE COMPARISON

	QP	PSNR		Total Bits		Bit saving (%)
		JM	CABS	JM	CABS	
Coast guard	22	40.52	40.54	5,512,784	5,384,424	2.33
	27	36.14	36.14	3,482,232	3,407,016	2.16
	32	32.42	32.42	2,033,528	1,992,464	2.02
	37	29.34	29.34	1,145,184	1,128,336	1.47
Carp hone	22	42.78	42.78	3,378,904	3,342,016	1.09
	27	38.99	38.99	2,186,600	2,163,144	1.07
	32	35.44	35.46	1,391,456	1,379,312	0.87
	37	32.13	32.13	900,152	892,552	0.84
Table	22	41.03	41.03	4,700,672	4,650,056	1.08
	27	37.02	37.02	2,830,008	2,790,936	1.38
	32	33.89	33.90	1,648,136	1,620,704	1.66
	37	31.25	31.26	1,008,168	991,352	1.67
Cont ainer	22	41.72	41.72	4,281,472	4,221,272	1.41
	27	37.91	37.91	2,763,768	2,728,184	1.29
	32	34.39	34.39	1,746,480	1,731,416	0.86
	37	31.00	31.00	1,084,840	1,078,864	0.55

Subjective Quality Assessment of Object-Based Video Material

Juergen Wuenschmann and Albrecht Rothermel
Universität Ulm
Institute of Microelectronics, Albert-Einstein-Allee 43, 89081 Ulm, Germany
Email: {juergen.wuenschmann, albrecht.rothermel}@uni-ulm.de

***Abstract*—A subjective quality test for the comparison of object-based (modeled, animated) and pixel-based video material has been carried out and the results are summarized in this paper. Tests showed that well known objective quality measures produce arguable results in this test constellation and therefore subjective quality assessment is the only way to reliably judge these test cases. The results show, that observers are very confident with the judgment of video material and can very well distinguish between different quality levels. For 10 out of 12 test cases, the observers could not distinguish the object-based from the original video material.**

I. Introduction

For animated content, the object-based storage of video can be a great opportunity to save storage space as well as to preserve a high video quality. Since the object-based storage and compression is a relatively new topic in video signal processing, the possibilities are diverse and the potentials not really exhausted but the contrary is true. To analyse the behavior of the object-based video storage, different factors which influence the file size have been isolated and the results have been presented in [1]. The codec that was used to compress the video representation is based on the MP4.25 object-based compression standard [2]. As the data-rate saving results have been promising, video quality evaluation was done. First, different objective quality measures were tested, but the results were not significant [3]. To prove the good quality impression of the object-based video material, a subjective quality test has been conducted. The paper is organized as follows: In Section II the test material and the compression used are explained. An introduction to subjective quality assessment is given in Section III and the results of the test are presented in Section IV. The paper is concluded in Section V.

II. Testmaterial

Selected test cases from the test scenarios described in [1] and [4] have been taken to carry out a subjective quality assessment. The test is composed of five different scenarios, which are growing numbers of objects without (1) and with textures (2), rising animation speed (3), variable scene length (4) and growing level of detail (5). For more settings see Table I. Included in the test are two different iterations of each of the six test scenarios, to cover a broad range of the test scenarios. For the scenario 'level of detail' three different compression levels have been produced for one iteration.
The test material is based on scenes represented in Collada [5].

TABLE I
Settings of the different tests.

Test	Settings	Quality Levels	File Size Tolerance
1.1	1600 Cubes	Standard	$< 3\%$
1.2	6400 Cubes	Standard	$< 3\%$
2.1	1600 Cubes	Standard	$< 3\%$
2.2	6400 Cubes	Standard	$< 3\%$
3.1	0005 Animation Speed	Standard	$< 3\%$
3.2	0050 Animation Speed	Standard	$< 3\%$
4.1	0151 Frames	Standard	$< 3\%$
4.2	0401 Frames	Standard	$< 3\%$
5.1	0008 Vertices	Standard	3129%
5.2	393218 Vertices	Low, Mid, High	$< 3\%$

First, the Collada files have been compressed using our working and improved version of the object-based MPEG 4 Part 25 Codec (MP4.25 [4]). The settings for the standard setup as well as the three quality levels can be seen in Table II. As comparison, pixel based versions of every clip have been rendered and encoded using h.264. The data rate has been set, that the object based and pixel based files have the same file size. While the pixel-based files are always a little bit bigger, file size tolerance lies below 3% except for the low level of detail test, 5.1, (see Table I). For that test a compression as high as with the object-based codec was not possible with h.264. Therefore the pixel-based version has been compressed to produce a file size as low as possible. Each test case consists of either an object-based or h.264 compressed video and its according reference, which is the rendered version of the Collada file.

TABLE II
Quality settings for the object-based video codec.

Parameter [bits] / Quality Setting	Standard	Low	Mid	High
Coordinate	10	15	10	5
Normal	9	15	9	5
Texture Coordinate	7	12	7	5
FAMC Coordinate	16	16	10	5
FAMC Normal	13	13	10	5
Key	20	20	15	10
Keyvalue	20	20	15	10
Keyvalue 360	4	4	4	4

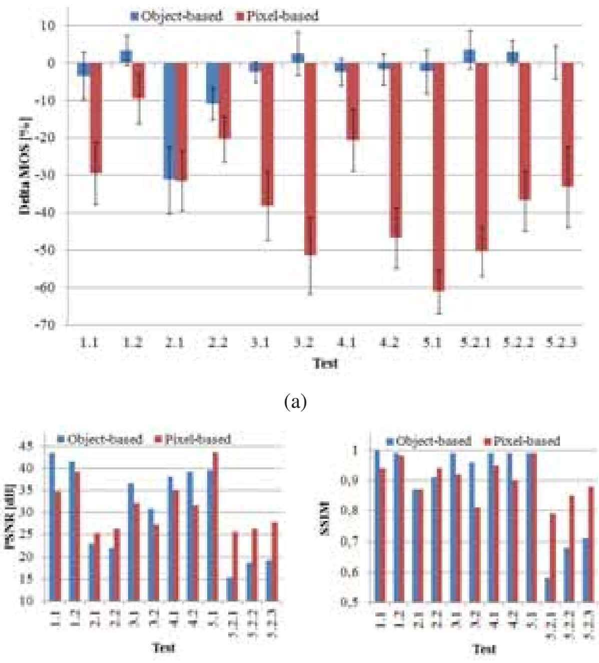
978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. Subjective Quality Assessment

The subjective quality test has been conducted following the recommendations of ITU-R BT.500-11 using the double stimulus continuous quality-scale (DSCQS) method [6]. A total of 24 tests have been shown to 24 expert and non-expert viewers of different age and gender. Every test contained two versions of a scene. One of them was always the unimpaired original and one was encoded either using our version of the object-based video codec MP4.25 or h.264. The order of the presentations was altered between test persons to even out e.g. tiring effects. Afterwards the results from each observer have been screened for outliers, but no observer had to be rejected.

IV. Results

In Figure 1(a) the results of the subjective quality test are shown. The measure used here is the relative mean opinion score (ΔMOS). Since every test case consists of the reference and an encoded version, the ΔMOS directly shows the relative quality and therefore the scores for object-based and pixel-based videos are directly comparable. The quality for the object-based test clips ranges between -31.3% and 3.6%, which is remarkable, because the observers judged the quality of four of the clips under test as better than the reference. This is, of course, virtually impossible, but together with the plotted standard deviation it proofs that no difference between reference can be observed for most of the test cases. The lowest quality scores for the object-based files were given for the test cases which use textures. The reason for this is, that the texture mapping gets distorted in the en-/decoding process and therefore the textures are shifted in the clips. For the pixel-based encoded clips the quality rating lies between $-61.1\%$and -9.5%. For every test the object-based clips have been judged better than their pixel-based counterparts. Even though the pixel-based version of test clip 5.1 is 31.29 times bigger than the object-based one, the observers rated its quality with the lowest value. The three different quality levels of test 5.2 can be found in the observer´s judgements for the pixel-based versions of the clips, whereas the three quality levels in object-based versions again could not be differed from the original and are judged equally.



(a)

(b)

(c)

Fig. 1. Test results of the subjective quality test (a) and the objective measures PSNR (b) and SSIM (c).

In Figure 1(b),(c) the peak signal to noise ration (PSNR) and structural similarity SSIM values of the test clips are plotted. It can be observed, that the tendency for some of the test cases is the same as in the subjective quality test. For PSNR the tendency was correct for 5 of 12 test cases, while, in principle, SSIM judged 8 of 12 correctly. Especially for test 5 the results are very contrary to those of the subjective test. Also the high quality difference identified by the observers between the object-based and the pixel-based version of the tests (except test 2) is not observed in the PSNR and SSIM results. Another problem of the objective quality measures used here, can be identified in test 5.1. This test contains an artificial rotating object and is dominated by a uniform background. In the pixel-based version the object is highly distorted, which is proofed by the low ΔMOS score, but the objective measures both mainly judge the non distorted uniform background and therefore attest the pixel-based version high video quality. These results confirm the observations made in [3], that there currently is no objective video quality measure available suited to reliably judge object-based video material.

V. Conclusion

The assumption, that the well known objective video quality measures are not suitable to judge object-based video material was confirmed. While PSNR and SSIM can be used to get a first impression of the quality differences of object-based video files with e.g. different data-rate, the comparison between pixel-based and object-based video material is unreliable. The subjective quality test revealed, that the object-based videos can not be distinguished from the original ones in 10 of 12 test cases. On average, the according pixel-based videos with the same data-rate are judged 35.7% worse than the original.

References

[1] J. Wuenschmann, C. Feller, and A. Rothermel, "File size comparison of modeled and pixel based video in five scenarios," in *Proc. International Conferences on Advances in Multimedia*, May 2012.

[2] B Jovanova, M Preda, and F Preteux, "Mpeg-4 part 25: A graphics compression framework for xml-based scene graph formats," *Signal Processing: Image Communication*, vol. 24, pp. 101–114, 2009.

[3] J. Wuenschmann and A. Rothermel, "On quality of object-based video material," in *Proc. IEEE Int Consumer Electronics - Berlin (ICCE-Berlin) Conf*, 2012.

[4] J. Wuenschmann, T. Roll, C. Feller, and A. Rothermel, "Analysis and improvements to the object based video encoder mpeg 4 part 25," in *Proc. IEEE Int Consumer Electronics - Berlin (ICCE-Berlin) Conf*, 2011, pp. 115–119.

[5] "Collada – digital asset schema release 1.5.0," .

[6] "Recommendation itu-r bt.500-11 - methodology for the subjective assessment of the quality of television pictures," 2002.

PCA based Shape Recognition for Capacitive Touch Display

I. Guarneri[1], A. Capra[1], A. Castorina, S. Battiato[2] and G. M. Farinella[2]
[1]AST - Computer Vision, STMicroelectronics Catania, Italy
[2]University of Catania, Viale A. Doria, Catania, Italy

Abstract— **With the growing diffusion of touch screen based consumer devices, the development of algorithms able to discriminate among different shapes obtained by touching the display becomes very important. For instance, the detection and recognition of fingers represent fundamental information in many touch based user applications such as image zooming, swipe, figure rotation, gaming, finger painting, etc. These algorithms are also extremely useful to recognize accidental touches in order to avoid involuntary activation of the device functionalities. Moreover, the recognition engine should be able to identify shapes acquired in critical conditions (e.g. slightly wet fingers). In this paper we present both hardware and software components to recognize different classes of shapes: single finger, double fingers and palm.**

I. INTRODUCTION

Projected capacitive touch displays [1][2] have become more and more popular and have been adopted by a wide range of consumer electronic devices. This technology is based on measuring the capacitance between two electrodes which are usually arranged in two layers as rows and columns. This kind of display can be divided into two different technologies:
mutual-capacitance and self-capacitance. In case of mutual-sensing, a finger touching the surface of the display changes the capacitance of electrodes and the signal variation is detected at the intersection of each row and column producing unique touch-coordinate pairs. In case of self-sensing, the capacitance variation is measured independently on each electrode on both X and Y panel directions, hence there is no way to recover unambiguously the touching position on the sensing surface. This leads to ghosting problems in the shape tracking step. Furthermore self-sensing technology is affected by lower SNR. Due to their desiderable properties, mutual-sensing touch displays are preferred in most of the current devices. An in-depth description of the projected capacitive touch screen can be found in [1][2].

Several techniques have been proposed to discriminate among different shapes acquired by touch displays. In [3] the authors exploit the integral image obtained by summing the signals data along rows and columns. The obtained *x* and *y* curves are segmented into peaks and valleys and the finger identification is obtained applying a threshold on peaks values. Despite the method is very simple it cannot guarantee robustness when shapes are obtained in critical conditions, such as in the case of slightly wet fingers. In [4] two dimensional patches are considered and several features (e.g., patch energy, patch radius) are computed after a preliminary high pass filtering stage. The features are then used together with simple discrimination rules for final shape classification.

In this paper we propose a method to discriminate among three different classes: single finger, double fingers and palm. The method is based on aggregating detected shapes, which are then represented through PCA decomposition [5] and classified through a discriminative decision tree [6]. To properly perform tests we use a projected capacitive touch display in order to acquire samples and hence test the proposed method on a real dataset of touches.

II. HARDWARE PROTOTYPE

The hardware equipment used in our test environment is composed by a 7-inch mutual capacitive touch panel with a resolution of 16×27 electrode nodes operating at 100 frames/sec. The board is equipped with an STMT05 microcontroller [7] that hosts a 32-bit RISC processor, the STM32 [8]. This allows for flexible customer code implementation of proprietary touch applications. Indeed, built-in associated processor memory (ROM, RAM) is optimized to run the desired fixed program codes in the ROM and to maintain the data, event stack and system variables in a RAM. An additional patch RAM can be used for implementing customized codes or algorithms. The board is connected to the PC through a USB 2.0. The prototype used in our experiment is shown in Fig. 1.

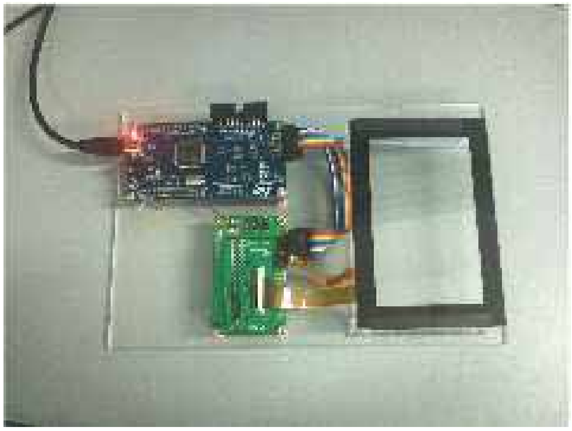


Figure 1 - The hardware prototype.

III. PROCESSING PIPELINE

In our work we focused on the recognition of the following commonly classes: Finger, Double Finger (e.g., two adjacent fingers) and Palm.

The samples have been acquired in order to guarantee high variability and slightly wet conditions (in our experiments a small quantity of water was spilled on the display). Figure 2 shows examples of the 3D capacitance shapes related to the considered classes.

A noise removal filter is applied to each acquired capacitance image (a 16×27 matrix containing raw data) to deal with noisy spikes. The filtered frames are then analyzed and single cells are aggregated into a patch using a watershed like algorithm [10].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

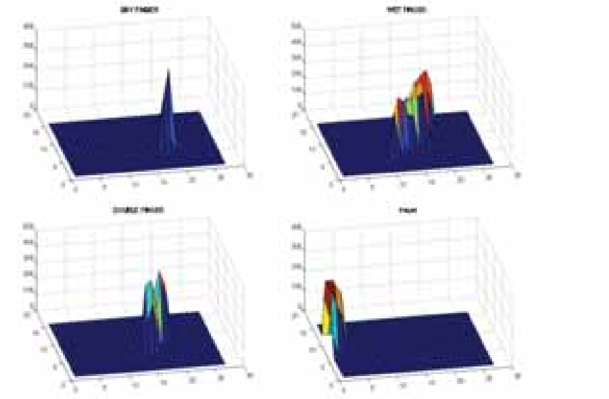


Figure 2 - 3D capacitance shapes. Top left: Dry Finger. Top right: Wet Finger. Bottom left: Double Finger. Bottom Right: Palm.

The patches aggregation step is especially useful in case of multi-touch samples. In order to guarantee the invariance to rotation and magnitude oscillation of the input data, the filtered and aggregated blobs are aligned and the 3D shapes are normalized. For each patch, a dominant orientation and the centroid coordinates are computed as follows:

$$P_M = (x_M, y_M), P_m = (x_m, y_m) : \|(P_M - P_m)\| = \max\{\|P' - P''\|\}, P', P'' \in \Omega$$

$$\theta = arctg\left(\frac{x_M - y_m}{x_M - x_m}\right)$$

$$Cx = \frac{\sum_{(x,y)\in\Omega} x \times f(x,y)}{\sum_{(x,y)\in\Omega} f(x,y)} \qquad Cy = \frac{\sum_{(x,y)\in\Omega} y \times f(x,y)}{\sum_{(x,y)\in\Omega} f(x,y)} \tag{1}$$

where Ω is the set of all patch points and *f* is the capacitance image in input.

Once *θ* and (*Cx,Cy*) have been derived the shape is rotated, centered and normalized as follows:

$$\begin{gathered}\Omega_{ROT} = \{(x\cos\theta - y\sin\theta - C_x, x\sin\theta + y\cos\theta - C_y) : (x,y)\in\Omega\} \\ f_{NORM}(x',y') = \frac{f(x',y')}{\sum_{(x',y')\in\Omega_{ROT}} f(x',y')}\end{gathered} \tag{2}$$

After collecting rotated and normalized capacitance patches we proceed similarly to [9] [11]. The PCA bases are computed on a dataset and only the basis corresponding to the 97% of total energy are retained. Each normalized capacitance image is hence projected into the subspace induced from the selected PCA basis obtaining an *N*-dimensional feature vector ***v****=(v₁,v₂,…v_N)*. Each feature vector is then classified accordingly to the rules obtained by training a decision tree on a set of labeled samples.

IV. Experiments and results

The experiments have been performed on a dataset composed of 1800 samples belonging to the three considered classes: Finger, Double Finger and Palm. The data have been acquired with the prototype presented in Section II. To properly collect the data and to guarantee samples variability, six subjects (3 male and 3 female) have been involved. Both dry and wet fingers have been included into the dataset.

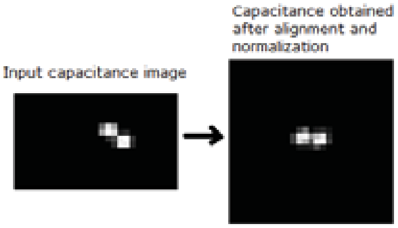


Figure 3 – Left: input capacitance image. Right: Capacitance image obtained through alignment and normalization.

The experimental results are obtained through *k*-fold cross validation procedure. At each run the set corresponding to one subject has been used as test set and the other 5 sets have been employed to train the decision tree. The subspace induced by the PCA decomposition is obtained from the training set at each run. Before training and testing, the overall data are projected in the feature space taking into account the selected PCA basis. The obtained results for each run are reported in Table 1.

TABLE 1: Experimental results

Fold	TP Finger	TP Double Finger	TP Palm	FP Finger	FP Double Finger	FP Palm
1	0.80	0.63	0.84	0.26	0.04	0.10
2	0.90	1.00	0.66	0.21	0.06	0.01
3	0.90	0.93	0.92	0.08	0.04	0.01
4	0.94	0.98	1.00	0.01	0.03	0.00
5	0.81	0.65	0.95	0.19	0.02	0.09
6	0.96	0.92	0.95	0.07	0.01	0.02
Avg	**0.89**	**0.85**	**0.89**	**0.14**	**0.03**	**0.04**

TP= True Positives, FP= False Positives, Avg= Average

V. Conclusions

In this paper both hardware and software components to discriminate among three different classes of shapes obtained from capacitive display devices have been presented. The proposed technique exploits a feature space obtained through PCA after the alignment and normalization of the capacitance images. A decision tree classifier is employed for final classification. The preliminary results on a real dataset show the effectiveness of the proposed method.

References

[1] Tim Wang & Tim Blankenship, "Projected-Capacitive Touch System from the controller Point of View," Society for Information Display, pp. 8-11, 2011.

[2] Geoff Walker, "Fundamentals of Touch Technologies and Applications," Society for Information Display, 2011.

[3] Jingkai Zhang, Yan Guo, Lianghua Mo, "Multi-touch detection method for capacitive touch screens," US 2011/0221701 A1.

[4] Wayne Carl Westerman, "Multi-Touch Input Discrimination," US Patent 0158185, 2008.

[5] Abdi. H., & Williams, L.J. ,"Principal component analysis," Wiley Interdisciplinary Reviews: Computational Statistics, pp. 433–459, 2010.

[6] Quinlan, J. R. "C4.5 Programs for Machine Learning," Morgan Kaufmann Publishers, 1993.

[7] http://www.st.com/internet/analog/product/252166.jsp

[8] http://www.st.com/internet/mcu/class/1734.jsp

[9] M.A. Turk, A.P. Pentland, "Face Recognition Using Eigenfaces," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1991.

[10] R.C. Gonzalez, R.E. Woods, "Digital Image Processing," third edition, Prentice Hall, 2008.

[11] G. Azzaro, M. Caccamo, J.D. Ferguson, S. Battiato, G.M. Farinella, G.C. Guarnera, G. Puglisi, R. Petriglieri, G. Licitra, "Objective Estimation of Body Condition Score by Modeling Cow Body Shape from Digital Images," Journal Dairy Science, Vol. 94, N. 4, pp. 2126-2137, 2011.

Distributed Architecture for Efficient Indoor Localization and Orientation

Félix J. Villanueva, Julio Dondo Gazzano, David Villa, David Vallejo, Cesar Mora,
Carlos Gonzalez Morcillo, Juan Carlos López
School of Computer Science
University of Castilla-La Mancha, Ciudad Real, Spain
Email: felix.villanueva@uclm.es

***Abstract*—In last decade, indoor GPS-like navigation has devoted attention from research community since is a key point for advanced services. Assistance, direct marketing, localization, etc. are examples of services that an appropiate navigation infrastructure could enable. With the evolution of cameras integrated in consumer electronics devices (mobile phones, notebooks, etc.) a new method, based on video streamming analisys gotted from mobile device videocamera, can provide us with the two necessary features for a sucessfull navigation experience, localization and orientation. In this paper, we present an architecture for videocamera streamming analisys joining existing algorithms for providing, in a robust way, to the consumer electronic device with its orientation and localization.**

I. Introduction

Indoor localization of users enable a great variety of services in order to help people to find the way to a meeting, to find a specific shop, to find the office for an administrative task, to know position of other users, etc. There are a great variety of methods and technologies which have been used for indoor localization, RFID tags, Wifi, Bluetooth, etc. The wireless technologies use or RSSI methods, with low accuracy and without orientation information or Time of Arrival and/or Angle of Arrival methods with special and costly hardware.

We think that a promising approach is the analysis of video streaming captured by the consumer device owned by the user. This analysis try to recognize interesting points in the video streaming and infers the position of the device who is sending the video. Obviously the environment has to be previously parametrized in order to lookup for patterns (some works use visual tags).

The solution we propose in this paper is based on the grid computing paradigm, with the purpose of integrating heterogeneous processing devices (partially reconfigurable or not) under the umbrella of the distributed object paradigm to make efficient implementation of Indoor localization and orientation services.

We model a multi-user architecture for include in user devices a generic application which send video streaming to the GPU/FPGA Grid getting its position and orientation. The application running in the user's mobile device is generic and can be used for position and navigation in any building.

We have a set of GPU/FPGA (dimensionated according to the environment and number of posible users) in which we previously store a set of interesting points according to the tracking algorithms. The user takes, for example, its notebook starting to send the streaming of the camera of its device to the GPU/FPGA grid. According to the processing of the streaming video, the GPU/FPGA grid compares with previously store information. Finally, the GPU/FPGA grid sends the calculated position and orientation to the device.

This works is embedded in a global architecture [1] which provides support to handicapped people. The current architecture is an autonomus location provider of that global information system.

II. High performing processing architecture

Image processing is normally decomposed in a processing pipeline where a set of different operations are applied over a stream of data with the possibility of having multiple inputs and outputs. FPGAs are specially well suited for this kind of process acceleration, in special when the type and size of data is not the standard in general purpose computing. On the other side parallelism can be exploited to the last consequence, not only for fine grain repetitive operations applied to a pixel neighborhood, but also for the global concurrent processing of several images at a time.

In our approach heterogeneous resources composed of GPU and FPGAs are used to accelerate part of the tracking algorithms of the indoor localization process. The main target of the proposed architecture is to provide a fast, energy efficient and escalable product involving different technologies. FPGAs offer the possibility of implementing tracking algorithms in two different fashions. One enterely software, using embedded processor such as Power PC, microblaze or Cortex 9 ARM; or a mixed approach accelerating some critical part or time demmanding part of the tracking algorithm. In our case we use the second approach. In this approach biggest time consuming parts of the different algorithms used for tracking methods are accelerated through the design of specific hardware modules, and implemented using the partial reconfiguration capability of the FPGAs. Each hardware module is implemented in one reconfigurable area (RA) (see Fig 1).

The design is formed by two main parts: an static one that have the control of the reconfiguration process, is composed of a reconfiguration controller plus a the reconfiguration engine,

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

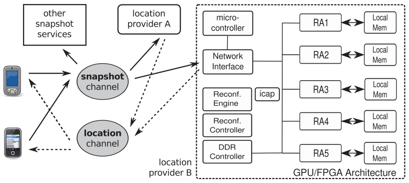


Fig. 1. Architecture overview

both implemented completely in hardware. In this way we obtain very short partial reconfiguration time. Besides, a network interface is provided to have access to the outside of the FPGA. This architecture allows to instantiate or replicate modules according to application demand when computational needs require, reducing power consumption. ARToolkit and PTAMM tracking methods were accelerated in our design. In the former the Binarization, Labeling and Pattern extraction subalgorithms were accellerated in hardware modules. In the latter the Features from Accelerated Segment Test (FAST) corner detection algorithm [2] was implemented using pattern compression. The hardware module designed to accelerate binarization process is formed by a component that perform the sum of the red, green, and blue values of each pixel, a component that perform the comparison with the threshold value, and a memory controller that takes data from DDR3 memory. This memory is used as a buffer for images to binarize. The image result is also stored in this memory. To accelerate Labeling sub algorithm a hardware module was developed using several blockrams in order to save data to be processed. The information in these blockram includes a row of binarized pixels to be processed taken from binarized images from DDR3 memory, the history buffer, the auxiliary values such as area, position, and the list of connected components. The processor sends information to the hardware labeling module in several format depending on the capacity of the blockram. All these modules can be replicated in different dinamically reconfigurable areas depending on the amount of information to be processed.

III. Communication infrastructure

The communication model is based on message events. The video sources (on smartphones) provide captured images tagged with timestamp and unique node identifiers. These *snapthot messages* are sent to a well-know specific event channel called *location-snapshot*. Any entity in the system is able to subscribe to it and receive these events.

Usually, the subscribers of this channel are *location providers* in turn. Location providers sent location and orientation information of especific physical entities in the environment. It may be different kind of location providers depending on their input information, but all of them send exactly messages with same format.

As snapshot messages, location events are sent to event channels too. This way the information can reach different consumers for different purposes. In summary, mobile nodes are snapshot publishers and location subscribers, but not limited to their own *queries*. That event-based model has many advantages respect to a RMI synchronous invocation model as, for example, a simpler programming models for providers, non blocking invocations, stateless location providers, decoupling publishers and subscribers, etc.

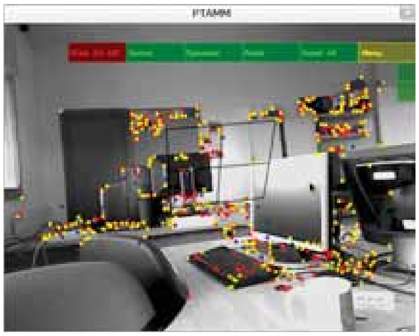


Fig. 2. PTAM interesting points detected

Snapshot messages are composed by three parameters: an image of 640x480 pixels and 24 bits color depth, a globally unique identifier for the device capturing the image and an time-stamp. Note that it is not required an accurate global clock reference due to this field is mainly relevant for the issuer device.

Location algorithms inputs are single images in all cases. Mobile nodes must produce images in the appropriate rate to ensure correct location feedback. Around 3-5 snapshots/location events by second must be enough for pedestrians. Note that same snapshot may be processed by several algorithms at different computing nodes. Depending on algorithm complexity, node performance and network load the location result may reach to mobile nodes in an arbitrary order, however this point should not be a problem at all.

Location messages are based on the MLP standard. It uses MLP::Position that includes an identifier for the provider, an identifier for the information source (the smartphone in this case), the timestamp and a shape due to just a point is not realistic.

IV. Conclusion

Our GPU/FPGA based architecture enable an efficient multi-user location and orientation position provider. In figure 2 we can see points detected using PTAM algorithm with our architecture. Next steps are stress test in order to model number of users vs necessary resources.

Acknowledgment: This research was supported by the Spanish Ministry of Economy under the project DREAMS (TEC2011-28666-C04-03) and by the Catedra Indra UCLM.

References

[1] Villanueva, F.J.; Martinez, M.A.; Villa, D.; Gonzalez, C.; Lopez, J.C.; , *Elcano: Multimodal indoor navigation infrastructure for disabled people* Consumer Electronics (ICCE), IEEE International Conf. on, pp.611-612, 2011

[2] Rosten, Edward and Drummond, Tom, *Machine learning for high-speed corner detection*, Proceedings of the 9th European conference on Computer Vision, ECCV'06, pp 430-443, 2006.

Touch Pointer: Rethink Point-and-Click for Accurate Indirect Touch Interactions on Small Touchscreens

Taekyoung Kwon, *Member*, IEEE, Sarang Na, *Non-member*, IEEE, and Sooyeon Shin, *Student member*, IEEE

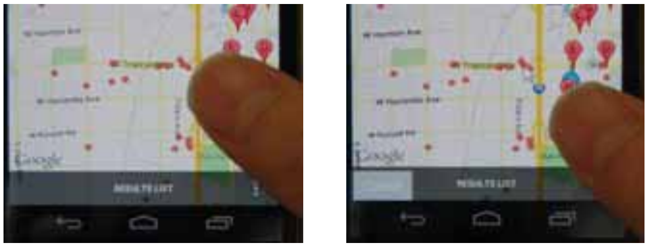
Abstract — *As a video resolution of handheld devices, such as a smartphone, increases higher, it becomes harder for a user to touch a tiny item accurately on a small touchscreen. The touch pointer implemented in dual layers may resolve this problem and render more functions to the touch action*[1].

Index Terms — Smartphone, touchscreen, and touch pointer.

I. Introduction

Smartphones are changing the way of life and everyday interaction between consumers and electronic computing devices. It is now commonplace to see users who make an entry with their fingers on a small touchscreen of handheld electronic devices. In particular, smartphones are used for accessing various kinds of Internet services and applications, which require a number of direct touch interactions on the small touchscreen. As the video resolution increases greatly higher, e.g., 4.8" 1280x720 pixels (306 ppi) and 3.5" 640x960 pixels (326 ppi), there should be more chances to display tinier items in a close location on the small touchscreen. Though the high resolution makes them visible, it gets harder for a user to touch or tap accurately one of those items with a finger [1]. From the famous equation of *Fitts's law*, we could observe a speed-accuracy trade off associated with pointing, whereby targets that are smaller and/or further away require more time to acquire. The tiny items close to each other may need to be zoomed in for accurate discrimination and click actions by a user, resulting in much more time to be taken. Otherwise, erroneous touches may require additional time to fix them. Double-click and/or right-click actions, for instance, to make a copy of text, would become more inconvenient. Thus, we are motivated to resolve this touch interaction or fat thumb problem on small touchscreen devices, such as a smartphone.

In this paper, we develop a novel touch interaction method called the touch pointer (TPointer, shortly) in software, under the paradigm of *point-and-click* rather than direct *tap-to-click* actions. Although the contemporary touchscreen expects the direct tap-to-click actions, we rethink the indirect point-and-click actions through TPointer, which incorporates *offset-free dual-layer* implementation for enhancing accuracy as well as avoiding visual occlusion, as illustrated in Fig. 1, and also for rendering more versatile functions to the click actions.



(a) (b)

Fig. 1. Direct vs. indirect touch interactions. (a) It is difficult to tap a tiny target item directly. (b) It is easier to point it and then tap a larger space of touchscreen via TPointer. An offset-free cursor is pointing the target.

II. Related Work

The offset cursor was studied by Potter et al. under the so-called *take-off* paradigm [3], but it yielded a difficulty and/or impossibility in accessing borders and took much time to find the offset precisely. There have been attempts to improve it but problems still remain. ThumbSpace [2] results in visual occlusion while Shift [5] loses accessibility to vertical borders. TapTap [4], similarly to the pinch-to-zoom action, disrupts an original scene, while MagStick [4] requires semantic pointing not to be general, and more training for its unusual control.

III. Touch Pointer

Touch interactions are classified into tapping and dragging conducted by a finger on a flat touchscreen, whereas multiple touch interactions involve simultaneous and/or consecutive actions of them. For an accurate click action on a minute item represented on the small touchscreen, we put a transparent layer of point-and-click action over the present application. TPointer should allow a user first to point the target item by dragging the layer and tap easily any place on the touchscreen rather than to tap the target item directly. TPointer should also allow the user to render particular functions to the click action, for instance, making a copy of text by one click or zooming a screen by one finger. There are at least five execution phases for TPointer; activation, setup, use, change, and deactivation.

A. Activation and Deactivation

Since TPointer is a temporary gadget to be used on various applications, it must be activated and deactivated by a user when necessary. For *activation*, the user may double-tap a (particular or any blank place of) touchscreen. A cursor then

[1] This work was supported in part by the IT R&D program of MKE/KEIT [10039180, Intuitive, convenient and secure HCI-based usable security technologies for mobile authentication and security enhancement in mobile computing environments]. This work was also supported in part by the National Research Foundation of Korea Grant funded by the Korean Government (MEST) (NRF-2012-R1A1B3-000965).

T. Kwon is with the Dept. of Computer Engineering, Sejong University, Seoul, 143-747, Korea (e-mail: tkwon@sejong.edu).

S. Na is with the Dept. of Computer Engineering, Sejong University, Seoul, 143-747, Korea (e-mail: tot@sju.ac.kr).

S. Shin is with the Dept. of Computer Engineering, Sejong University, Seoul, 143-747, Korea (e-mail: shinsy80@sju.ac.kr).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

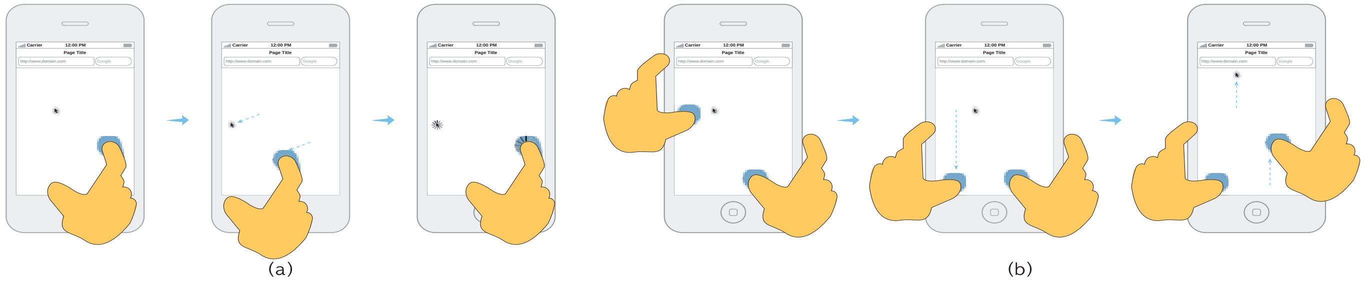
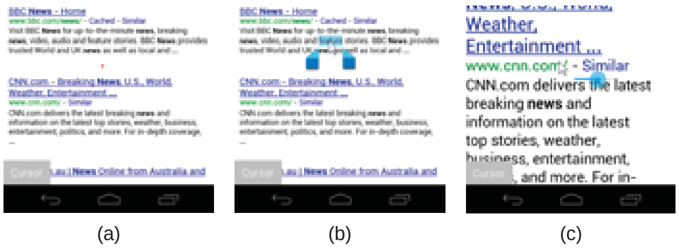


Fig. 2. TPointer in use. (a) When TPointer is activated, a user may touch any place of touchscreen for dragging and pointing of the offset-free cursor and tap any place for a click. (b) If there are dual touches, the second one (a left thumb in this example) works on an application layer while the first one (a right thumb) works on a TPointer layer. The left thumb scrolls the application whereas the right thumb navigates the cursor by dragging.

appears on the touchscreen. More precisely, a virtual layer representing the cursor is covering the current application temporarily. Thus, a single touch event will be treated as a touch event on the virtual layer as illustrated in Fig. 2-(a). There could be dual touch events, far enough not to be interpreted as a right-click. As illustrated in Fig. 2-(b), in this case, the second touch will be treated as a touch event on the application layer, not on the virtual layer of TPointer. So, the user is able to touch directly the original application as well as navigate the cursor on it while activated. For *deactivation*, the user may double-tap again a (particular or any blank place of) touchscreen. We could make the cursor blink for a small amount time before disappear, so that the user can revoke deactivation by holding the screen when the cursor is blinking.

B. *Setup*

It is possible to set up distinct shapes of cursors and render various functions to them. In the *setup* mode, a user may customize those cursors with preferred functions, such as click a minute item, copy text, and zoom in/out, as snapshotted in Fig. 3. The user may swipe the touchscreen from right to left ends to enter the setup mode and reversely to return.



BBC News - Home
www.bbc.com/news/ - Cached - Similar
Visit BBC News for up-to-the-minute news, breaking news, video, audio and feature stories. BBC News provides trusted World and UK news as well as local and ...

CNN.com - Breaking News, U.S., World, Weather, Entertainment ...
www.cnn.com/ - Similar
CNN.com delivers the latest breaking news and information on the latest top stories, weather, business, entertainment, politics, and more. For in-depth coverage, ...

Cursor .au | News Online from Australia and

(a)

BBC News - Home
www.bbc.com/news/ - Cached - Similar
Visit BBC News for up-to-the-minute news, breaking news, video, audio and feature stories. BBC News provides trusted World and UK news as well as local and ...

CNN.com - Breaking News, U.S., World, Weather, Entertainment ...
www.cnn.com/ - Similar
CNN.com delivers the latest breaking news and information on the latest top stories, weather, business, entertainment, politics, and more. For in-depth coverage, ...

Cursor .au | News Online from Australia and

(b)

Weather, Entertainment ...
www.cnn.com/ - Similar
CNN.com delivers the latest breaking news and information on the latest top stories, weather, business, entertainment, politics, and more. For in-

Cursor

(c)

Fig. 3. Functional cursors in use. (a) A tiny red cursor for a minute item. (b) Text copy. (c) Zoom-in and zoom-out. SeekBar is used for adjustment.

C. *Use and Change*

When TPointer is activated, a user can navigate the offset-free cursor by dragging, i.e., similarly to a touchpad, in the *use* mode. Fig. 2-(a) illustrates an example. A single touch event at any place of touchscreen will be treated as a touch on the layer of TPointer. The user can point a tiny item on the touchscreen by navigating the cursor and then tap any place of touchscreen for a click. It is straightforward for TPointer to point-and-click the item indirectly and avoid visual occlusion. TPointer could allow not only the single left-click action but also a right-click, e.g., replaced with a close dual touch, or double-click action, e.g., replaced with a touch-and-hold action, on the same layer. As we described above, more functions can be rendered for various purposes. For the *change* of functions, the user may swipe the touchscreen from left to right ends, so that a distinct cursor replaces the current cursor and vice versa.

IV. Implementation and Experiment

We implemented a prototype system on the Google Android platform and conducted a user study. In the pilot study, we presented the Google map sample (Fig. 1) to 8 participants for a specific destination and gained impressive results: TPointer recorded no error, but direct tap-to-click actions 51.25% errors without zooming actions. In more controlled experiments, we presented ten tiny color spots as illustrated in Fig. 4 and asked 15 participants to touch them in sequence. Impressively, 42.33% of trials were only succeeded in direct tap-to-click actions, but 99.67%, that is, 2.35 times more, in indirect point-and-click actions through TPointer ($t(14)=-10.051$, $p<0.001$). The cost was 1.85 times more action time ($t(14)=-15.720$, $p<0.001$) but justifiable. We will deal with more details of implementation and user experiments in the full paper.

Fig. 4. Click experiments. (a) Tiny color spots. (b) TPointer on the spots.

V. Conclusion

TPointer enables more accurate click actions and renders more functions on a small touchscreen. It is a wise choice to incorporate TPointer into systems of smartphone-like devices.

References

[1] V. Balakrishnan, T. Park, and C. Meet, “A study of the effect of thumb sizes on mobile phone texting satisfaction,” *Journal of Usability Studies*, vol. 3, pp. 118-128, May 2008.

[2] A. Karlson and B. Bederson, "ThumbSpace: Generalized One-Handed Input for Touchscreen-Based Mobile Devices," In Proc. of *Interact'07*, pp. 324-338, 2007.

[3] R. Potter, L. Weldon and B. Shneiderman, "Improving the Accuracy of Touchscreens: An Experimental Evaluation of Three Strategies," In Proc. of ACM *CHI'88*, pp. 27-32, 1988.

[4] A. Roudaut, S. Huot, and E. Lecolinet, "TapTap and MagStick: Improving One-Handed Target Acquisition on Simple Touch-screens," In Proc. of ACM *AVI'08*, 2008.

[5] D. Vogel and P. Baudisch, “Shift: A Technique for Operating Pen-Based Interfaces Using Touch,” In Proc. of ACM *CHI’07*, pp. 657-666, 2007.

Grip-Ball: A Spherical Multi-Touch Interface for Interacting with Virtual Worlds

Seungju Han and Joonah Park
Advanced Media Lab., Samsung Advanced Institute of Technology, Samsung Electronics, KOREA

Abstract— **Graspable user interface (UI) is a useful method for interacting with virtual world. It allows users to hold and manipulate a device intuitively for manipulating virtual 3D objects. In this paper, we propose Grip-Ball as graspable UI, which is a spherical shaped device based upon capacitive multi-touch sensing. User's grip pattern can be recognized by measuring grip touch on the surface of the device. In the study, we define 5 hand-grip patterns of the Grip-Ball according to user's purpose (platform, power-grip and precise-grip) and number of hand (one-handed grip, two-handed grip). An accuracy of up to 98% could be achieved for hand grip recognition.**

I. INTRODUCTION

A concept of graspable user interfaces (UI) is a natural, intuitive way to manipulate 3D objects in a virtual environment [1, 2]. For instance, one can pick up a virtual object, move it around, and deform the object by squeezing it, with the ease of manipulating a real object. Therefore, instead of controlling through arbitrary key-mappings, the graspable UI can leverage user's intuitions about how devices should be used for certain functions.

Several prototypes of the graspable UI have been recently presented. One of them, the Tango, is a hand-sized spherical interface with a 3-axis accelerometer and an 8x32 capacitive sensing grid housed in a compressible dielectric material [3]. Since it is actually attempting to infer general hand poses and uses single-handed model, it requires additional constraints such as single hand use. Also, iGrip presented as a mobile user interface for handheld devices (Box-type) by recognizing hand grip pattern [4, 5].

In this paper, we propose Grip-Ball, a spherical shaped device based upon capacitive multi-touch sensing. As shown in Fig. 1, the Grip-Ball senses grip touch on its surface and recognizes five hand grip patterns; Platform, one-handed power grip, one-handed precise grip, two-handed power grip, two-handed precise grip. It enables intuitive accessing and manipulation of 3D virtual objects.

II. SYSTEM ARCHITECTURE OF GRIP-BALL

The Grip-Ball detects the touch on its spherical surface by capacitive sensing technique. Major components of capacitive touch sensing systems include a sensing mean, a measurement circuit, and a signal processing algorithm. The sensing mean comprises a set of electrodes and a panel on which the electrodes are arranged. The measurement circuit converts measured capacitance into voltage or current values. The signal processing algorithm maps these values to position, area, etc. to be used by the user interface code - determining a single touch area may require reading measurements from multiple electrodes depending on the sensing architecture.

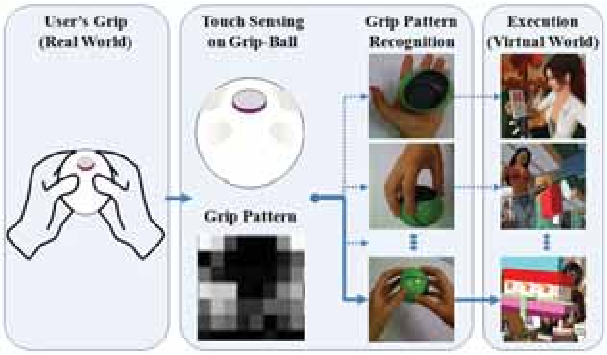
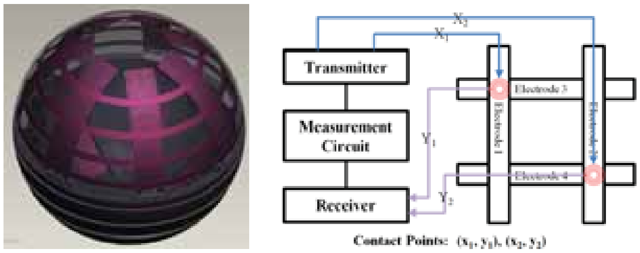


Fig. 1. Grip-Ball System enables intuitive accessing and manipulation of 3D virtual object. It senses the hand grip by detecting the touched portion of the user's hand and launches relevant application.



(a) A Spherical Multi-touch Pad (b) A Coupled Matrix Array

Fig. 2. System Architecture of Grip-Ball.

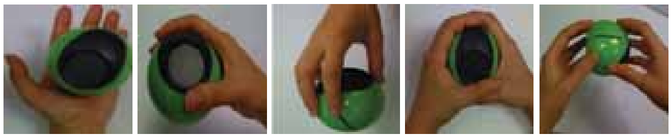
Fig. 2 shows the spherical multi-touch system. In order to handle multiple contacts, we use the simplest method, which is assigning each electrode to each channel of the device. The architecture has the coupled matrix array with 64 channels to build a multi-touch sensing system with 8 by 8 resolutions. In the coupled matrix array architecture, the sensor detects the capacitance between electrodes in the same column and those in the same row. It is similar to the matrix scan technique used for keyboards.

III. HAND GRIP PATTERN OF GRIP-BALL

Hand grip classification is made according to Cutkosky's grasp taxonomy [6]. It separates grasps according to their purpose (power-grip, precise-grip) and the object size, and allows distinguishing most of the grasps required for manufacturing tasks. In Fig. 3, we defined five hand grip patterns for the Grip-Ball as following Cutkosky grasp taxonomy; Platform, one-handed power grip, one-handed precise grip, two-handed power grip, two-handed precise grip.

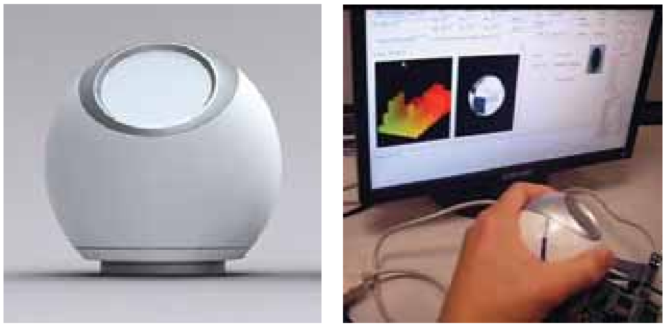
Grip-recognition is difficult to evaluate because of the lack of any ground truth. Even in situations where a grasp is

978-1-4673-1363-6/13/$31.00 ©2013 IEEE



(a) (b) (c) (d) (e)

Fig. 3. Hand Grip Patterns for Grip-Ball ((a) Platform, (b) One-handed Spherical Power Grip, (c) One-handed Spherical Precision Grip, (d) Two-handed Spherical Power Grip, (e) Two-handed Spherical Precision Grip)



(a) Proto-Type (b) Demonstration

Fig. 4. Grip-Ball System

universally recognized, the exact grip will vary from person to person based on hand size. The study on the classification accuracy of our system is conducted using the grip-pattern data collected from 5 users. 100 grip-pattern data for each grip type was used for the test. The users were guided to take hold of the device according to examples shown in Fig. 3.

IV. CLASSIFICATION RESULTS

For this study, we experimented with three classifiers for the grip-pattern recognition; a minimum distance classifier (MDC) [7], a naive Bayes classifier (NBC) [8], and a support vector machine (SVM) classifier [9].

The improvement in our method comes from the signal processing algorithm for stabilizing and denoising the sensor measurements, rather than relying on the naive signal processing provided by the sensor chip itself. Note that our signal processing algorithm uses raw sensor measurements, rather than binary measurements.

As shown in Fig. 5, the accuracy of each classifier is evaluated based on n-fold cross validation test. The three classifiers (MDC, BNC, and SVM) are built on the MATLAB (MathWorks, Inc.) environment for training. For MDC, NBC, and SVM classifier, we obtain 98.2 %, 98.8 %, and 99.4 % of best success rate on the 4-fold cross validation test, respectively. We found that three classifiers were able to correctly identify user's grips with over 98% accuracy. This led us to conclude that the sensor design was adequate for grip-recognition.

Tables I shows the confusion matrix of MDC, which is the worst case on the 4-fold cross validation test. We observe that one-handed power grip (class 2) is comparably the lowest class on the MDC classifier. Also, it is the most confusing class compared with class 3, 4, and 5.

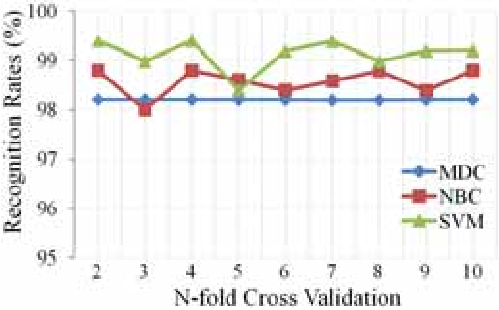


Fig. 5. N-fold Cross Validation Test to compare performance of MDC, NBC and SVM.

TABLE I
CONFUSION MATRIX OF MDC (%)

Class	1	2	3	4	5
1	1.00	0	0.01	0	0
2	0	0.94	0.01	0	0
3	0	0.02	0.98	0	0.01
4	0	0.02	0	1.00	0
5	0	0.02	0	0	0.99

Class 1: Platform, 2: One-handed power grip, 3: One-handed precision grip, 4: Two-handed power grip, 5: Two-handed precision grip

V. CONCLUSIONS

In this paper, we propose Grip-Ball as a graspable UI, which is a spherical shaped device based upon capacitive multi-touch sensing. It allows users to hold and manipulate naturally and intuitively virtual 3D objects. From the experiment, we found three classifiers (MDC, NBC, and SVM) could correctly identify user's grip with over 90% accuracy. Also, we conclude that the sensor design of Grip-Ball was adequate for grip-recognition.

REFERENCES

[1] G. W. Fitzmaurice and W. Buxton, "An Empirical Evaluation of Graspable User Interfaces: Towards Specialized, Space-Multiplexed Input," *CHI 1997: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems.* ACM Press, pp.43-50, 1997.

[2] B. Taylor and V. M. Bove, "Graspables: Grasp-Recognition as a User Interface," *CHI 2009: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems.* ACM Press, pp.917-925, 2009.

[3] D. K. Pai, E. W. VanDerLoo, S. Sadhukhan, and P. G. Kry, "The Tango: A Tangible Tangoreceptive Whole-Hand Human Interface," In Proc. *of World Haptics.* pp.141-147, 2005.

[4] W. Chang, K.E. Kim, H. Lee, J.K. Cho, B.S. Soh, J.H. Shim, G. Yang, S. Cho, and J. Park, "Recognition of Grip-Patterns by Using Capacitive Touch Sensors," *IEEE ISIE 2006,* pp.2936-2941, 2006.

[5] K.E. Kim, W. Chang, S. Cho, J. Shim, H. Lee, J. Park, Y. Lee, and S. Kim, "Hand Grip Pattern Recognition for Mobile User Interfaces," In Proc. *of AAAI 2006.* pp.1789-1794, 2006.

[6] M. R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *IEEE Trans. On Robotics and Automation,* vol. 5, no. 3, pp. 269-279, 1989.

[7] T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, and A. Linney, "Classification of audio signals using statistical features on time and wavelet transform domains," In Proc. *of ICASSP '98,* vol. 6, pp. 3621-3624, 1998.

[8] I. Rish, "An empirical study of the naïve Bayes classifier," In Proc. *of IJCAI '01,* pp. 41-46, 2001.

[9] B. Schlkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning).* 1st ed. The MIT Press, 2001.

Design and Implementation of a New Thin Cost Effective AC Hum Based Touch Sensing Keyboard

H. M. Elfekey, *student member, IEEE,* and H. A. Bastawrous, *Member, IEEE*

***Abstract*--Although touch pads and screens have been widely researched recently, complexity and high cost have been major issues in their implementation. In this paper, we present a new touch-sensing technique for the implementation of a simple and cost-effective touch keyboard based on the AC hum phenomenon which exists in the vicinity of any AC source. The proposed technique has the advantage of being customizable according to the demands of the user. Moreover, it can be integrated with other applications as it can be made flexible, transparent, in any size and in any color.**

I. INTRODUCTION

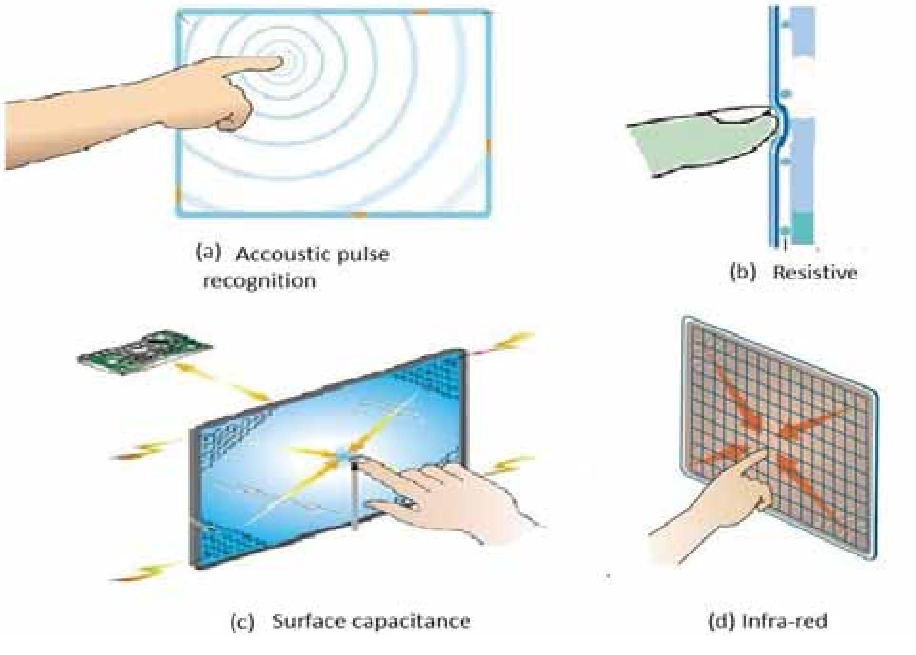
The human to machine interface is a vital part in the design of any new device. This interface can be designed to accept human speech, motion or just a touch, as human input. Touch-sensing techniques have been widely used nowadays in various applications, for example game consoles, all-in-one computers, tablet computers, and smart phones. Each technique depends on a set up that is interrupted when the surface is touched.

In order to build a touch keyboard, one may use any of the available touch-sensing techniques. One example is the acoustic pulse recognition sensing technique which uses sound waves, as shown in Fig. 1 (a). The glass panel shown is made of very pure glass such that when a user touches a certain location, a sound with certain frequency is generated in this panel. The frequency generated depends on the location of the touch. Then, transducers are placed around the edges of the screen to detect this touch. The controller then compares the data received using a simple look up table and determines the location of the touch [1].

Another well-known touch-sensing technique is the resistive one that uses 3 main layers. As shown in Fig. 1 (b), the first layer contains conductive coating on the bottom. The second layer is made of insulating spacers while the third has conductive coating on the top. When the device is touched, the upper layer bends a little at the location of the touch and the upper and lower layers are connected. Stripped electrodes are scattered evenly on the first and third layers so that when the two layers are connected at a certain location, current passes through the corresponding electrodes and the controller connected to these electrodes can determine the location [2].

Also, of the popular techniques is the surface capacitance technique which is used to implement touch screens. As shown in Fig. 1 (c), the screen consists of a uniform conductive coating on the corners of a glass panel causing a uniform electric field inside that panel. When the user touches the surface, four currents are drawn from the four corners because of the touch action. The amount of current from each corner depends on the distance of the touch position to that corner. By measuring and comparing the four currents, the position can be determined [2].

The Infrared technique shown in Fig. 1 (d) depends only on the usage of IR waves. In this technique, two arrays of IR



(a) Accoustic pulse recognition

(b) Resistive

(c) Surface capacitance

(d) Infra-red

Fig. 1. Examples of available commercial touch-sensing technologies.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

sources are placed on the x-axis and y-axis as well as opposite two arrays of IR receivers. When the user touches the surface, these rays are cut and so the detectors receive no signals. By knowing which signals were lost, the location of the touch is determined [2].

All the above mentioned touch-sensing technologies have either expensive hardware compared to normal push buttons like the infra-red or the acoustic pulse recognition techniques, or complicated algorithms as in capacitive touch technique for instance, or the need of stylus to enhance the sensitivity of the resistive touch panel. In this paper, a new touch sensing keyboard is designed based on the AC hum phenomenon which will allow a relatively simple and cost-effective implementation that directly senses the human finger touch. The proposed technique benefits from the unwanted presence of AC hum noise to achieve high touching sensitivity.

II. AC HUM PHENOMENON

AC hum can be defined as the noise that is generated by any AC current nearby. It represents a problem which is normally associated with audio systems used in personal computers, broadcasts, commercial sound or music [3]. The origin of this phenomenon comes from Faraday's general law [4]; stating that any AC current in a wire produces an alternating magnetic field around that wire according to the following equation,

$$\oint \vec{E} \cdot \vec{ds} = \frac{d\phi_B}{dt}, \tag{1}$$

where E and Φ_B are the electric and magnetic fields, respectively. This alternating magnetic field induces voltage on any conductive materials nearby adding noise.

III. OPERATION PRINCIPLE

As human body contains minerals and salts, it behaves as a conductive material and may be modeled as an antenna [5]. Therefore, AC hum-based magnetic fields induce an alternating voltage on the human body as shown in Fig.2. The internal impedance of the equivalent antenna is the human body's impedance which is in the scale of hundreds kΩs [6]. This will lead to the fact that an amplifier is required in order for the controller to read such weak AC signal. When the human body connects the keyboard surface to the earth, it works as an antenna closing the whole circuit and producing a touch event.

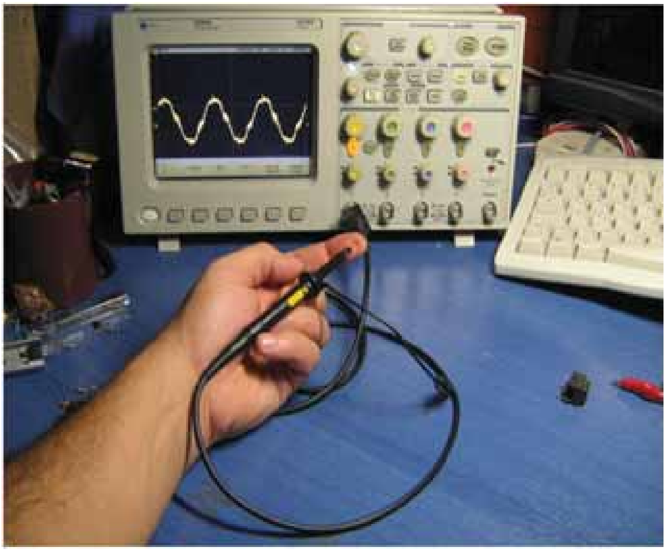


Fig. 2. Voltage associated with the human body due to AC hum.

IV. PROPOSED KEYBOARD STRUCTURE

The keyboard is divided into 3 building blocks as shown in Fig.3. The first block is the touch panel which functions as the human interface that contains all the touch buttons, taking the input from the user as a weak AC signal to the amplifier. As mentioned in the previous section, the human body may be modeled as an antenna producing a weak AC voltage signal, and hence, it is crucial to have an amplifier in order for the controller to detect a non-zero input signal.

The last part in this setup is the controller. The required controller should support USB protocols and have enough pins

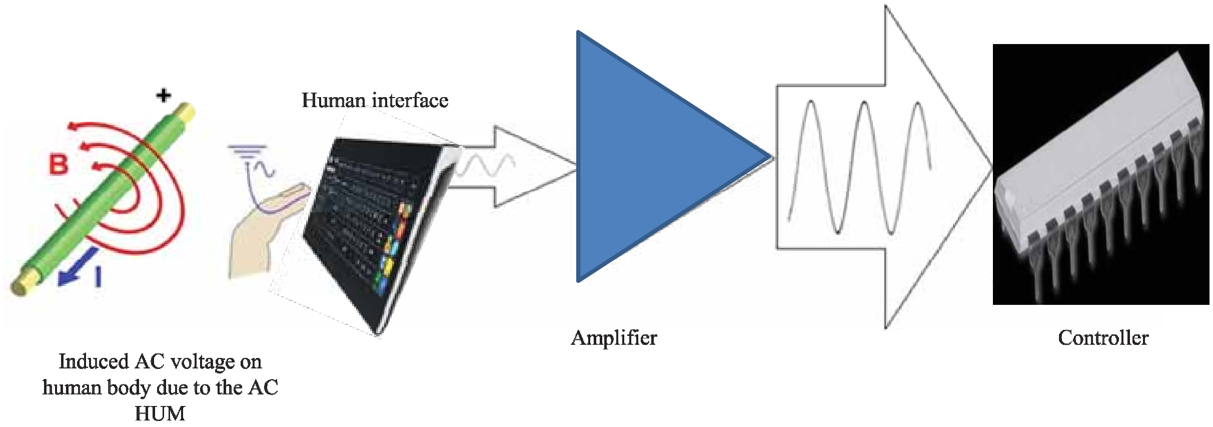


Fig. 3. Proposed keyboard's block diagram.

due to the fact that several tens of pins are needed here. In the following sub-sections, each block of the proposed keyboard (i.e., interface, amplifier, and controller) will be explained in details.

A. Human Interface

The human interface is the proposed touch panel consisting of 21 bare wires and insulators which are implemented in four layers, as shown in Fig.4. The bottom transparent layer is a normal plastic sheet referred to as the substrate. The next layer is made of 6 horizontal wires which will be referred to as rows. The third layer contains 15 vertical insulators while the last is composed of vertical wires on the insulators referred to as the "columns". Within this setup, the first row and column are at the upper right corner and all the columns and rows are arranged to form a matrix of coordinates.

This setup can support up to 90 buttons in total where each touch button is an open circuit between a row and a column until the user touches it. Then, AC signal is transmitted through this row and this column closing the circuit at their intersection. Hence, a button can't be represented by more than one row or one column. Moreover, it can't be represented by a single column only or a single row only.

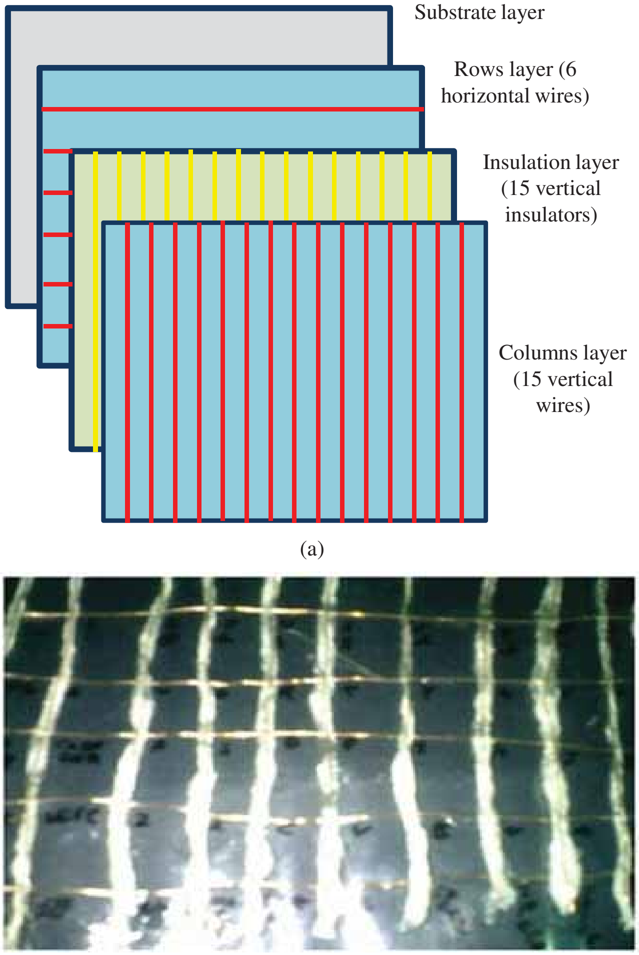

Substrate layer
Rows layer (6 horizontal wires)
Insulation layer (15 vertical insulators)
Columns layer (15 vertical wires)
(a)
(b)

Fig. 4. Proposed keyboard layers' structure (a) Block diagram showing the design of interface layers, (b) Hardware implementation of the interface.

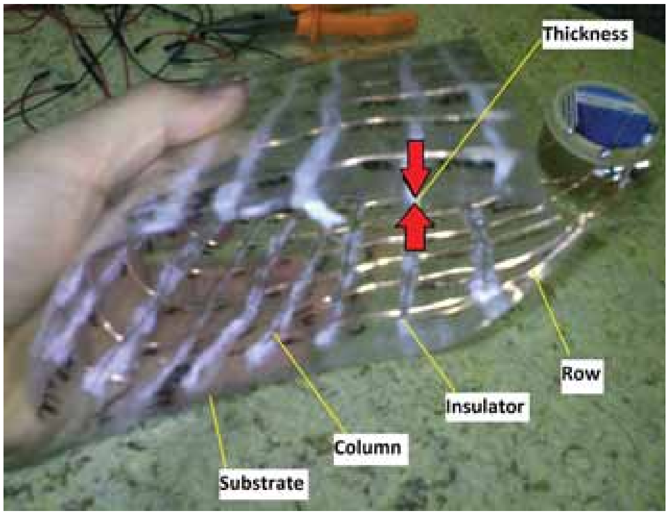


Fig. 5. Interface of the implemented touch-sensing keyboard.

The proposed interface design has many desirable features. First of all, it can be bent as shown in Fig. 5, and so, flexible and easy to store. Second, water doesn't destroy it and so easy to wash. Third, it can be customized to be smaller or bigger according to application. Finally, it can be easily attached to any surface.

B. The Amplifier

The amplifier block is composed of 21 Darlington amplifier cell for the 21 inputs (15 columns and 6 rows). The Darlington pair consists of two cascaded transistors. As shown in Fig. 6, the emitter of the first BJT is connected to the base of the second BJT. The amplification of a single BJT "β" is typically around 100 times the base current in the active region [7]. Therefore, the overall amplification is β^2 or 10000 times the base current of the first BJT which is the AC current associated with the human body [8]. Each amplifier cell is made of a Darlington Pair, two resistors, and a LED as show in Fig. 6. The LED is for indication purposes while the 50Ω

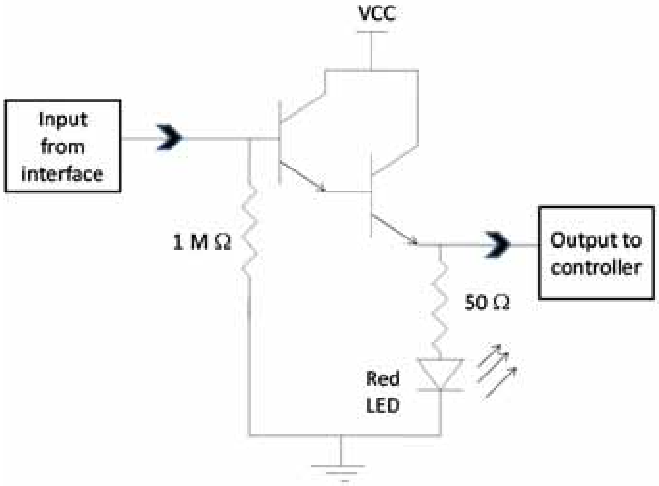


Fig. 6. Single Darlington pair amplifier cell.

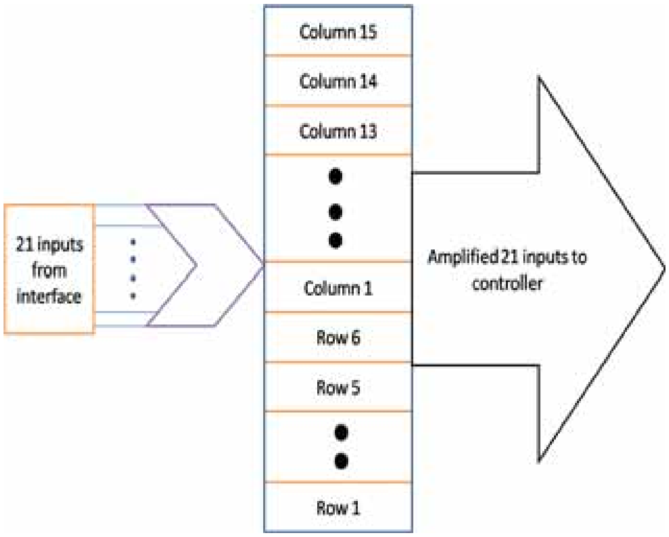


Fig. 7. Block diagram of the amplifier cells' configuration.

series resistor is just for protection. The 50Ω resistor and the LED are placed in parallel to the input of the controller. The 1MΩ is a pull down resistor to nullify the effect of parasitic capacitance in the circuit.

C. The controller

The controller should receive an analogue input. However, the outputs of the amplifier section need to be connected to digital pins for convenience, as some controllers may have fewer analogue pins compared to digital ones. If they were to be processed as analogue signals, more than one controller would be needed. But fortunately, in our setup, these signals can be connected to digital pins, because only a non-zero voltage is needed to recognize a touch event.

Now the controller has to distinguish between a square wave and no signal. Reading a zero at the input is not enough to judge no touching event. Therefore a timer is essential to keep track of how long the controller is reading a zero. The run time for judging that there's no touching event is 30ms since any AC power line is 50/60 Hz. 30 ms is a period and a half assuming 50 Hz. If logic '1' is detected again before the timer reaches 30 ms, then the algorithm continues assuming square wave input. If logic '0' is detected after the time limit, then the algorithm continues taking into account no touching event occurred.

According to which pins carry the square wave, the controller can determine the position of the touch. The controller simply compares the input to a simple look up table similar to that of acoustic pulse recognition technique [1]. Therefore, the algorithm is very simple and the input is read as normal digital signal, which are significant advantages for this keyboard implementation.

V. Conclusion and Discussion

In this paper, a new touch-sensing operation principle based on the AC hum voltage associated with the human body was utilized to build a touch-sensing keyboard. The implemented keyboard was constructed with simple components (e.g., basic Darlington cell amplifier and typical controller) and hence it did not require complicated or expensive hardware. Moreover, the keyboard has very desirable features as it is more resistant to breaking and withstands dust and harsh surroundings better than the conventional keyboards.

We are now working on several future upgrades for the implemented keyboard. First upgrade is by powering up the input signal using an antenna that gives high power within a small range, and so, increasing the touching sensitivity of the keyboard. A second upgrade to the interface would be using a flexible PCB to reduce the overall thickness. Actually, it is also possible to add a very luxurious upgrade by using transparent, conductive and flexible polyester like those used in the resistive touch screens although this will increase the cost significantly.

A Third possible upgrade would be the design of the buttons themselves, for instance, using a circular or a rectangular shape. In addition, the distance between the buttons has to be considered for optimum comfort while using the keyboard.

Finally, future enhancement for the amplifier is to use the Sziklai pair [9] instead of the Darlington pair. The main advantage of this upgrade is that the required activation voltage will be reduced from 1.4V to 0.7V and hence improving the touching sensitivity further.

References

[1] *Acoustic Pulse Recognition*. Elo Touch Solutions Inc., California, CA, 2006.

[2] T. Hoye and J. Kozak, "Touch screens: a pressing technology", presented at University of Pittsburgh. Conf. Tenth annual freshman, Pittsburg, PA, April 2010.

[3] *AC Power Systems.* Arrakis Systems inc., Loveland, CO, 2007.

[4] J. Jewett and R. Serway, "Faraday's law", in *Physics for scientists and engineers with modern physics*, 8th ed. Belmont: Marry Flinch, 2010.

[5] G. cohn, D. Morris, S. N. Patel, D. S. Tan, "Humantenna: Using the Body as an Antenna forReal-Time Whole-Body Interaction". Presented at ACM CHI 2012, Austin, TX, 2012.

[6] P. Sutherland, D. Dorr and K. Gomatom, "Human Current Sensitivities and Resistance Values in the Presence of Electrically Energized Objects," IEEE Industrial and Commercial Power Systems Technical Conference, pp. 159–167, 2005 IEEE, ISBN 0-7803-9021-0.

[7] A. R. Hambley, "Bipolar Junction Transistor", in *Electrical Engineering Principals and Application*, 5th ed. Boston: Prentice Hall, 2011.

[8] A. S. Sedra and K. C. Smith, "Single-Stage Integrated-Circuit Amplifiers", in *Microelectronic circuits*, 5th ed. New York: Oxford Univ. press, 2004.

[9] B. Pandey, S. Srivastava, S. N. Tiwari, J. Singh and S. N. Shukla, "Qualitative analysis of small-signal modified Sziklai pair amplifier" *Journal of Pure & Applied Physics,* vol. 50, April, pp.272-276, 2012.

A Touch Based Affective User Interface for Smartphone

Mira Kim, Hyun-Jun Kim, Sun-Jae Lee, Young Sang Choi
Intelligent Computing Lab, Future IT Research Center
Samsung Advanced Institute of Technology, Samsung Electronics Co., Ltd.

***Abstract*—Nowadays, people interact with smartphones constantly throughout their daily lives so that smartphones become the most ideal device for recognizing users' physiological and emotional states. In this paper, we present a novel approach for providing users with personalized interface by inferring human emotions based on the touch behavior of a smartphone. Recognizing human emotions is a challenging task and we propose to use touch related data to infer user's emotional states. We also propose a personalized user interface which applies the observed emotional variance to emotion-aware interaction between users and the devices.**

I. Introduction

Smartphones equipped with various sensors are proliferating worldwide and are turning into sensing systems benefitting users with various types of data collected from the internal sensors. The motivation of the study is to provide smartphone users experience emotional familiarity and personalized recommendations on contents and services. Although various of smartphones exist in the current market, they use a few choices of operating systems and therefore their user interfaces are nearly uniform and do not change dynamically. The look-and-feel of a smartphone screen including icons of installed applicatoins has a potential to improve user experiences if it dynamically adapts to users' emotional changes such highes and lows. There exists previous researches to recognize human emotions with smartphones[1][2] and with touch pattern[3]. However, most studies did not utilize the recognized emotion to improve user interfaces by dynamic adaptation. In a previous study, we developed a framework for detecting human emotions based on the touch interface of a smartphone[4]. In this paper, we utilize the touch based emotion recognition framework to create a system for dynamically altering and readjusting the user interface of smartphone users based on the users' emotion change. To demonstrate the effect of our approach, we implemented user interface adaptions such as transforming the color, size, border and position of application icons based on the inferred users' emotional state to assist users with their state of mind.

II. Touch based Emotion Recognition

Many studies helped to find the universal correlation between human emotions and the facial expression[5]. In addition to recognize human emotion based on the facial expressions, other bio-signals such as skin conductance and heart rate variability can be used to detect human emotions.

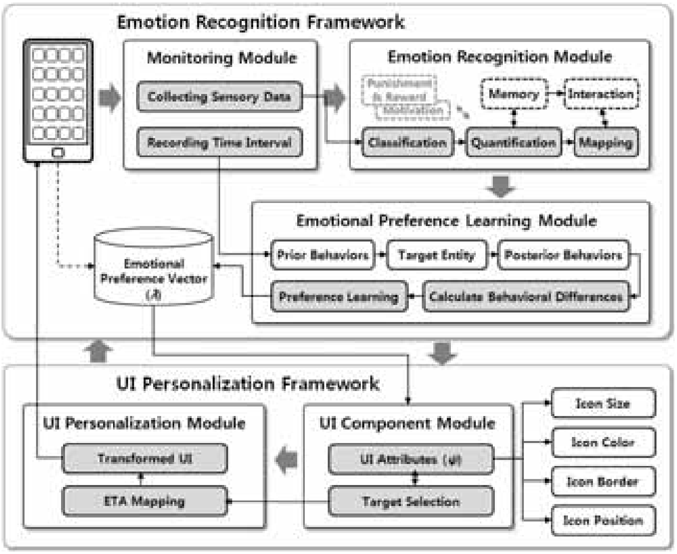


Fig. 1. The control flow of the emotion-aware process

Compared to these modalities of emotion recognition, our touch base recognition relies only on user's natural interaction with mobile devices not requiring taking videos or wearing obtrusive sensors. Our touch based recognition framework comprises of an emotional recognition process and emotional preference learning algorithm which learns the difference between two quantificated emotional states; the prior and the posterior emotional states for each interaction entities in a mobile device such as specific media contents, application software, or communication contact. Once the user's emotion has been determined, we apply the appropriate modifications to each entities. The framework utilized the touch sensor and the accelerometer sensor of a smartphone. We utilized 11 features such as duration, distance, speed, and acceleration. Fig. 1 shows the overall control flow of the system. The monitoring module collects device's sensory data with its timestamp. The recognition module infers the emotional states and the emotional preference learning module builds an emotional preference matrix. Utilizing our emotion based service framework, we obtained 0.907 of f1-score by using C4.5 decision tree[4].

III. System Design

Based on the emotional preferences and states detected, we devise a mapping function for dynamically modifying

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

the user interface in various aspects such as icon color, icon size, icon border and rearrangement. Users can also manage the configuration to control their preferences such as favorite colors and favorite mobile application genres based on defined emotional states. To enable on demand modification of mapping emotion to user interface change, we define an *Emotion-to-Action* profile. Each item in a profile is a tuple of emotion, action, and confidence. The affective user interface calculates similarity between detected emotional states, invokes the necessary actions based on the emotions, and transforms the user interface. With our trained data, emotional scores for each application are stored in the emotional preference vector(λ). The preference vector holds emotions and the score of emotions detected for all of user's applications. When a touch occurs, the recognition module creates an emotion vector of *n* emotions defined in the system which, for instance can be shown as $F = \langle 0.7, 0, 0, 0, 0.1, 0, 0.2 \rangle$. This means detected emotion was happiness with 70 percent probability, surprise with 10 percent and anger with 20 percent. The next part of the affective user interface is visualization rules according to the emotions. User preferences specified by users are taken into account as well. That is, transformation value(T) can be defined as $\{\overline{\delta}_{size}, \overline{\delta}_{border}, \overline{\delta}_{color}, \overline{\delta}_{position}\}$ and UI attributes(ψ) define the amount of transformation needed for each emotion. For example, $\overline{\delta}_{size}$ defined as $\langle 0.5, 0.1, -0.2, 0.2, -0.1, 0.4, 0 \rangle$ means when happiness is detected, we enlarge the width and height of icons by 0.5. Therefore, after emotion recognition, we efficiently compare the values in similarity between F and the preference matrix to validate if the conditions are met for transformation.

Algorithm 1 Affective UI Transformation Algorithm

```
$input \leftarrow F$                                  ▷ Detected Emotions
procedure ETA($F$)
    $A[][] \leftarrow \psi$                            ▷ UI Attributes
    $M[][] \leftarrow \lambda$                         ▷ Preference Matrix
    while $i \leq F.length$ do
        while $j \leq M.length$ do
            if $(M[i][j] = F[i])$ then
                $\gamma \leftarrow similarity(M, F)$
                if $\gamma \geq \theta$ then
                    $Transform(M[i][j], A[i][j])$
                end if
            end if
        end while
    end while
end procedure

procedure TRANSFORM($M, A$)
    $MaxThresh = \alpha/2$                 ▷ $\alpha$=Maximum screen size
    while $i \leq A.length$ do
        if $A \leq MaxThresh$ then
            $A' \leftarrow ln(A \times (i + 2))$
            $redraw(M, A)$
        end if
    end while
end procedure
```

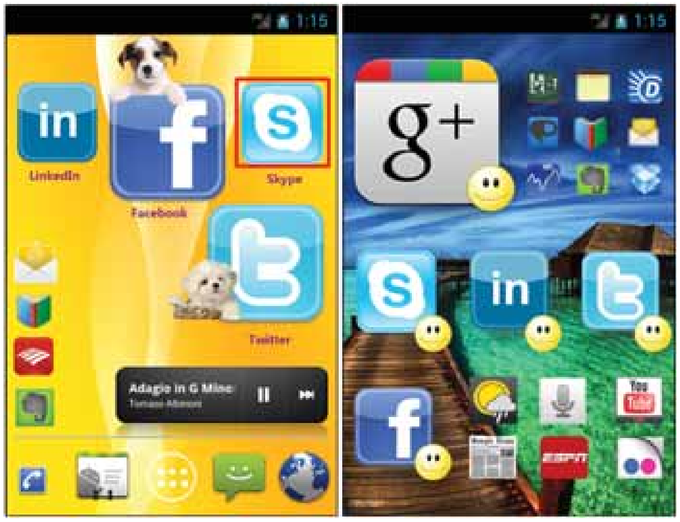


Fig. 2. Examples of The transformed User Interface

Lastly, we display the transformed user interface. The user can manually override the defined actions by setting users' preference later. We provide a predefined action for each emotion and allow changing actions at runtime to make users conveniently initialize the system and enhance it with further interaction.

Fig. 2 illustrates the transformed user interface according to the sensed emotion and user preferences. As shown, the icons for social networking service applications are enlarged and applications for finance and task-oriented applications are shrinked. Furthermore, icons are rearranged and the color schemes are modified according to the user's preference and detected emotions.

IV. Conclusion and Future Works

We created a novel framework for providing smartphone users with a personalized user interface based on the current emotion and established emotional preference. We believe that emotion based personalized smartphone interface will improve user satisfaction by providing differentiating user experience. In this paper, we used predefined actions for given emotions, but there is a need for further research in recognizing the best case scenarios of UI transformations for detected human emotions.

References

[1] R. LiKamWa, Y. Liu, N.D. Lane and L. Zhong, "Can Your Smartphone Infer Your Mood?", Workshop on PhoneSense, 2011.

[2] K. Church, E. Hoggan and N. Oliver, "A Study of Mobile Mood Awareness and Communication through MobiMood", Nordic Conference on Human-Computer Interaction, 2010.

[3] Y. Kim, S. Koo, J. Lim and D. Kwon, "Computational Models of Emotion", A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing, 2010.

[4] H. Kim and Y. Choi, "Exploring emotional Preference for Smartphone Applications", Consumer Communications and Networking Conference, 2012.

[5] Z. Zhihong, P. Maja, P.I. Glenn and S.S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions", IEEE Transactions on Pattern Recognition and Machine Intellgence, Vol.31, No.1, 2009.

HW/SW Architecture for Speech Recognition Acceleration

Richard Fastow, Stephan Rosner, Venkat Natarajan, Qamrul Hasan, Jens Olson,
Markus Unseld, Feng Liu, Handoko Chendra, Ojas Bapat, Chen Liu

Spansion, Inc.
915 DeGuigne Drive, MS 177, P.O. Box 3453
Sunnyvale, CA, U.S.A. 94088
Phone: (408) 616-8167, email: richard.fastow@spansion.com

Abstract

This paper describes a new hardware architecture for fast Acoustic Model scoring in embedded speech recognition systems by integrating an 8-way data-path with a NOR Flash array. This architecture localizes computation of critical algorithms and reduces decode latency and CPU load by 50% for large acoustic models, resulting in lower word error rates for natural language speech recognition.

Introduction

Systems for natural language speech recognition typically utilize three main processing stages (Fig 1) [1]. After the incoming utterance is sampled and digitized in the DSP stage (Phase 1), the generated feature vector enters the Acoustic Modeling stage (Phase 2), where it is compared to a list of senones in the library. Each comparison results in a senone score which is then input to the Language Modeling stage (Phase 3). The time spent in the Acoustic Modeling stage typically represents 30%-70% of the total decoding time, depending on the speech task. Larger Acoustic Models, composed of either more senones or more Gaussians, take longer to process and generally result in improved accuracy (Fig 2) [2].

In this work, a new hardware architecture to accelerate the Acoustic Modeling stage of Speech Recognition on a single chip is proposed and tested on an FPGA. The Acoustic Coprocessor (ACP) chip is designed to replace the software subroutine that calculates senone scores in most Hidden Markov Model based Speech Recognition systems. It is evaluated on both an open source decoder (Sphinx 3) as well as on a commercial decoder (VoCon 3200). The ACP is shown to score Acoustic Models 8 times faster than an ARM CORTEX A8 processor on a SABRE platform, resulting in a 50% reduction in the decoding latency and enabling use of larger Acoustic Models while maintaining real time performance in embedded systems.

Chip Architecture

The key idea is to implement the Acoustic Modeling as a data path integrated inside a memory holding the Acoustic Models. This allows parallelization of the computation and communication and almost completely eliminates the external memory bandwidth requirements for the Acoustic Modeling stage. The ACP is comprised of 256 Mb of MirrorBit NOR Flash memory which store the Acoustic Models, along with on-chip logic to process the data and compute the senone scores (Fig 1). The flash array reads 768 bits of data every 80 ns, resulting in a dedicated on-chip data processing throughput of 1.2GB/s. Eight Arithmetic Logic Units process the data from the flash array, calculating the log-likelihood of the 39 dimensional incoming feature vector with the Gaussians in the Acoustic Model. The highly pipelined and parallelized data path results in an effective processing speed of 2.0 GOPS. After the individual Gaussian scores are calculated they are weighted and summed in the log domain resulting in the senone score. A block diagram of the ACP is shown in figure 3 illustrating the main blocks and the interface between the memory and logic sections.

Experiment and Results

The raw Acoustic Model scoring speed was characterized on typical embedded processors (ARM11, Cortex A8/A9) and for comparison on a high-end desktop processor (Intel i7). We measured how many Gaussians are computed in a given time period by continuously looping through the senone scoring routine. Figure 4 shows that the ACP outperforms ARM11 by 30x, Cortex A9 by 8x, and achieves almost 90% of the i7-class performance on this benchmark. In order to measure the effect of accelerating the Acoustic Modeling calculation on the overall speech recognition performance, the complete stack was implemented. As host systems, the first configuration used an ARM11-based iMX31 under Linux and the second configuration used a Cortex A8-based iMX53 under QnX. The host executes the speech recognition stack. Calls from the speech stack to the Acoustic Modeling are routed via an SIO-SPI bus to the Acoustic Coprocessor emulated on the Stratix3 FPGA (Fig 5). The logic portion of the coprocessor is synthesized into the Stratix3. The Flash array is modeled with a dedicated 64bit DRAM on the emulation platform. The internal ACP bandwidth of 1.2GB/s can be achieved if the DRAM delivers the same 768bits in the same 80ns access time. However, meeting real-time constraints with DRAM is not trivial because of the timing-variances of refresh cycles and performance-driven open-page policies. For the experiments we dedicated a 64bit DRAM port operating at 333MHz. It delivered an average latency of 110ns due to controller and bus latencies as well as delays for clock domain synchronization. This represents a

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

27% longer latency than the ACP and therefore is a conservative model for the proposed chip architecture. Figure 6 shows a latency improvement of 50% on the VoCon 3200 v4 Full Acoustic Models. For communication between host and FPGA via SPI bus a bandwidth of 3MB/s was measured. These results demonstrate that accelerating the Acoustic Model scoring in a speech recognition system with the proposed ACP architecture reduces latency by 50% and reduces the bandwidth for operand fetches by a factor 400 from 1.2GB/s to 3MB/s. This is a key prerequisite for real time speech processing in embedded systems.

Acknowledgments

We gratefully acknowledge the help and guidance of Bart Baeyens and Johan Vantieghem from Nuance Inc., and Prof. Paul Franzon from NCSU.

References

[1] D.Chandra, U.Pazhayaveetil, P.Franzon, “Architecture for Low Power Large Vocabulary Speech Recognition,” IEEE International SOC Conference, pp. 25-28, 2006.

[2] K.Vertanen, “Baseline WSJ Acoustic Models for HTK and Sphinx: Training Recipes and Recognition Experiments,” Technical Report, Cavendish Laboratory, 2006.

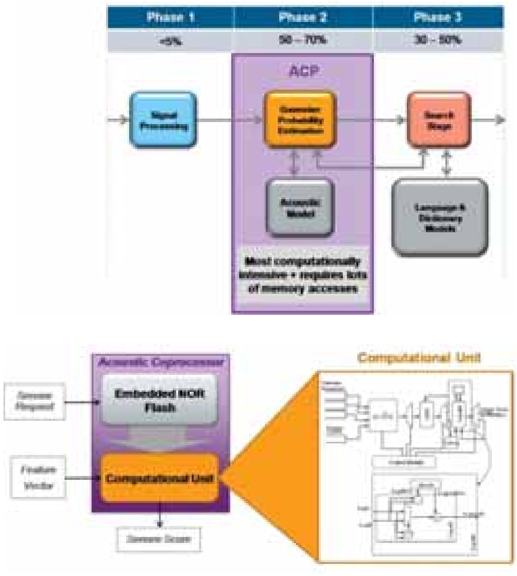


Figure 1. Top – Speech Recognition decoding flow used in Sphinx 3. Bottom – HW architecture used to accelerate the Acoustic Modeling.

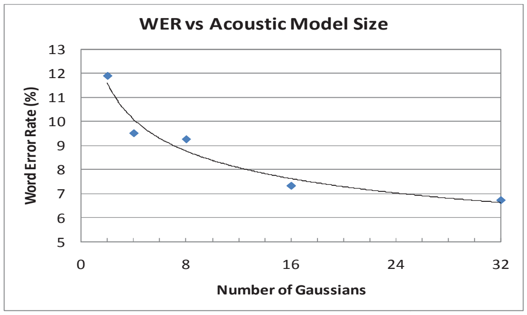


Figure 2. Effect of Acoustic Model size (number of Gaussians) on WER. Data collected using Sphinx 3, reported in reference [2].

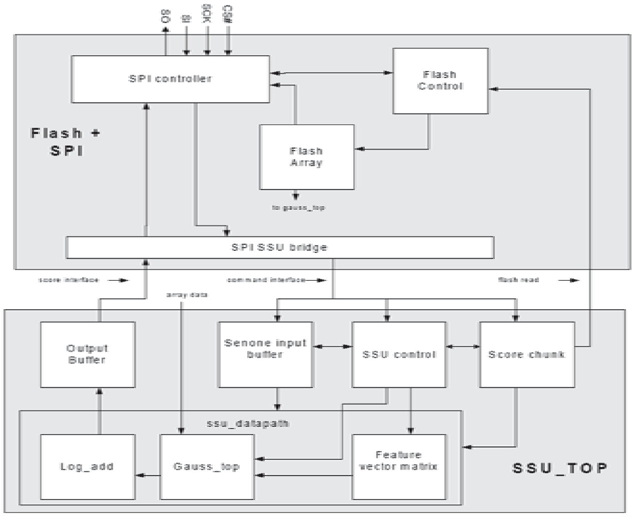


Figure 3. Block diagram of the ACP chip.

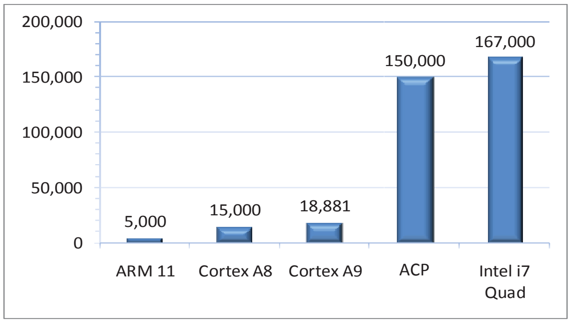


Figure 4. System benchmarking of Acoustic Modeling performance. The y-axis is the number of Gaussians calculated in a given period.

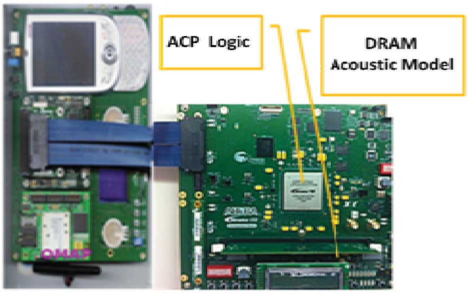


Figure 5. Stratix3 FPGA ACP emulator (right) connected to an ARM11-based iMX31 system using Linux

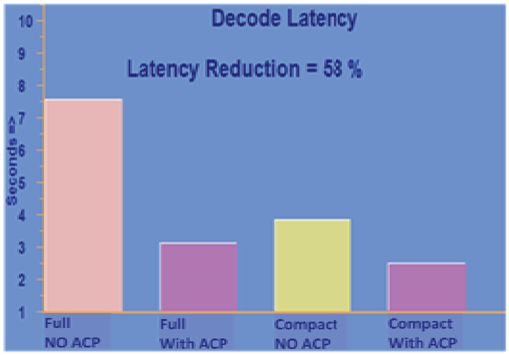


Figure 6. Total decode latency with and without the ACP. Compact and Full refer to the size of the Acoustic Models.

A Delay-based Transceiver with Over-current Protection for ECU Nodes in Automobile FlexRay Systems[1]

Chih-Lin Chen, *Student Member, IEEE*, Zong-You Hou, Sheng-Chih Lin, and Chua-Chin Wang[2],
Senior Member, IEEE

***Abstract* –This work presents a FlexRay Transceiver (FRT) used in an in-vehicle network compliant with the latest FlexRay physical layer standards. The proposed FRT utilizes a delay-based mechanism to reduce glitches. Besides, an Over-current Protection (OCP) circuit is included to avoid short-circuit hazard. The proposed is implemented on silicon using a typical 0.18 μm CMOS process. The total core area is 0.774 × 0.565 mm^2 and the power consumption is 158.4 mW at 10 Mbps data rate.**

Key word: FlexRay, transceiver, Over-current Protection.

I. INTRODUCTION

FlexRay V3.0.1 is the latest communication protocol [1] proposed by several automobile power houses, including BMW, Daimler-Chrysler, General Motors, Freescale, Philips, Bosch, Volkswagen, etc., in 2010. Recently, many researches regarding the FlexRay transceiver have been publicized [2]-[3]. According to FlexRay specification, a robust FlexRay transceiver is required to include a Bus Failure Detector (BFD) and short-circuit protection. A BFD is in charge of detecting whether buses are shorted to VDD or GND, which usually utilizes a voltage comparison method to distinguish bus status. However, BFD could generate a wrong warning signal if a glitch or large over-shoot voltage caused by the power MOS or switches' transitions accidently appear on bus [2]. Therefore, we propose a delay-based method in FlexRay transceivers to prevent such a situation. Besides, according to FlexRay specifications, the maximum output current limit of the transceivers is 60 mA no matter the buses are shorted to VDD, GND, or other buses. Thus, an Over-current Protection is needed in the transceivers. After the bus current is scaled down by a current mirror, we utilize a Current Bias Circuit, which generates a 60 μA current, and a current comparator to realize the required Over-current Protection on silicon.

II. DELAY-BASED TRANSCEIVER WITH OVER-CURRENT PROTECTION FOR FLEXRAY STSTEMS

Fig. 1 shows the proposed FlexRay Transceiver (FRT) design, including a Transmitter (Tx), a Receiver (Rx), Bus Failure Detector, a Current Bias Circuit (CBC), and a Voltage Bias Circuit (VBC). Tx is in charge of transmitting data on bus (BP/BM), where an Over-current Protection (OCP) is needed to avoid short-circuit hazard. Rx and Bus Failure Detector (BFD) are used to receive data and monitor bus status, respectively. Whenever BFD detects any short-circuit alarm on bus, BFD sends a warning to upper level as well as the controller (not shown). Notably, VBC and CBC generate a reference voltage, Bias_2V=2 V, and a reference current, Bias_60μA=60 μA, to Tx and Over-current Protection, respectively.

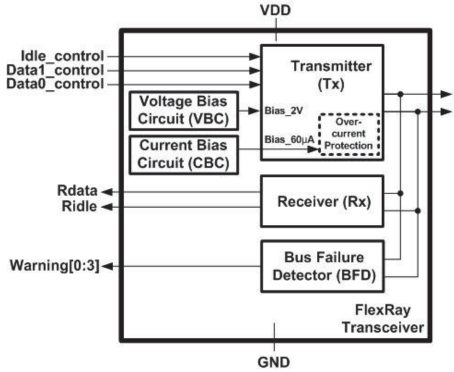


Fig. 1. Explosive view of FlexRay Transceiver (FRT).

The proposed delay-based Transmitter and Over-current Protection are shown in Fig. 2. The transitions of power PMOS transistors (e.g., LH0~LH3) and power NMOS transistors (e.g., LL0~LL3) can not be synchronized perfectly. A large current will be generated on BP and BM if they are accidently turned on at the same time. That is, a "glitch" is generated. A delay switch design (i.e., RH, RL, LH, and LL) are added to reduce the glitch. Besides, the power PMOS and NMOS transistors are equally divided into many transistors, i.e., M102 ~ M117. Because the size of each transistor is small and the transition time of gate drive thereof also is equally delayed, no large current will be resulted in BP and BM. Take M110 ~ M113 as an example. Buffer_N1, Buffer_N2, and Buffer_N3 provide different delays to drive the gate of M111, M112, and M113, respectively. The power transistors, M110~M113, are turned on step by step. An illustration in Fig. 3 shows the timing waveform of Buff_N1~ Buff_N3, and the without (prior work) and with (this work) glitch reduction. The glitch will be apparently eliminated by properly delaying the transition time of each power MOS.

As mentioned, the absolute maximum output current limit is 60 mA when the bus is shorted. Fig. 2 shows Over-current Protection (OCP) circuit to detect such a current limit. We use M119 as a current-mirror to scale down M118's tail current (≤60 mA) to 1/1000, which is supposed to be 60 μA at most. A Current Bias Circuit without BJTs based on [4] generating a reference 60 μA current (Bias_60μA) is included in FRT. When the $I_{tail,M119}$ is equal to or over Bias_60μA, OCP sends a signal, "Over-Current" to turn off Tx.

III. IMPLEMENTATION AND MEASUREMENT RESULT

The proposed FRT, including Tx/Rx, Bus Failure Detector, Voltage Bias Circuit, and Current Bias Circuit, is carried out

[1] This investigation is supported by National Science Council, Taiwan, under grant NSC99-2221-E-110-082-MY3.

[2] C.-C. Wang, C.-L. Chen, Z.-Y. Hou, and S.-C. Lin are with Department of Electrical Engineering, National Sun Yat-Sen University, 80424, Taiwan. (email: ccwang@ee.nsysu.edu.tw)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

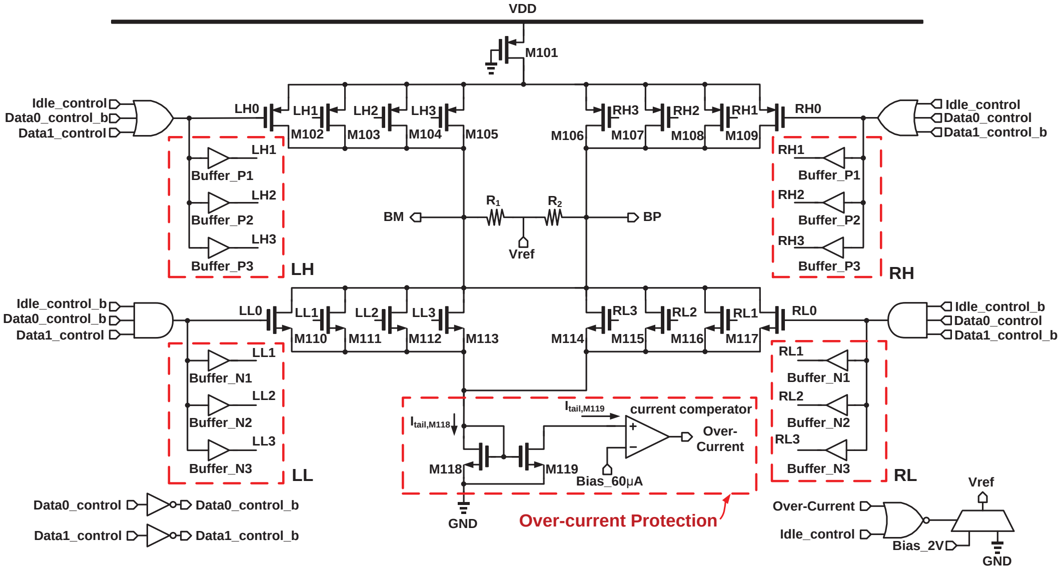


Fig. 2. The schematic of delay-based transceiver with over-current protection.

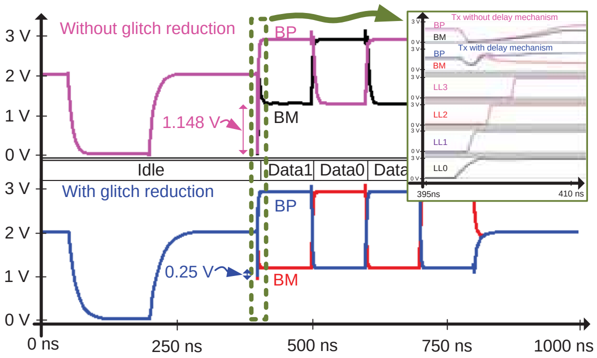


Fig. 3. The waveforms of Buffer_N1~ Buffer_N3 and glitch reduction (without and with).

using a typical 0.18 μm CMOS process. Fig. 4 shows the measurement set-up and readings, including all instruments, die photo, and measured waveforms. The total core area on silicon is 0.774 x 0.565 mm^2. The function generator, AFG 3252, generates a pair of Data0_control and Data1_control to FRT. The oscilloscope shows the BP and BM signals decoded by DPO 4AUTOMAX Automotive, which match the decoded Rdata by Rx. Therefore, the functionality of our proposed design is proved. Table I shows the comparison between the required FlexRay V3.0.1 specification and the measurement results of our FRT to prove that all of required transceiver specifications are met. Table II shows the performance comparison of the proposed FRT with our previous FlexRay transceiver design [2].

TABLE I
SPECIFICATION OF THE PROPOSED FRT

FLEXRAY TX SPECIFICATIONS V3.0.1		**MEAS. RESULTS**
Absolute value of uBus while sending	600 ~ 2000 mV	1520 mV
Absolute value of uBus while Idle (mV)	0 ~ 30 mV	13.1 mV
Transmitter delay negative edge (ns)	< 75 ns	5.18 ns
Transmitter delay positive edge (ns)	< 75 ns	4.35 ns
Transmitter delay Mismatch (ns)	< 4 ns	0.73 ns
Receiver delay, negative edge (ns)	75 ns	6.95 ns
Receiver delay, positive edge (ns)	75 ns	6.30 ns
Receiver delay mismatch (ns)	5 ns	0.65 ns
Throughput (Mbps)	10 Mbps	10 Mbps

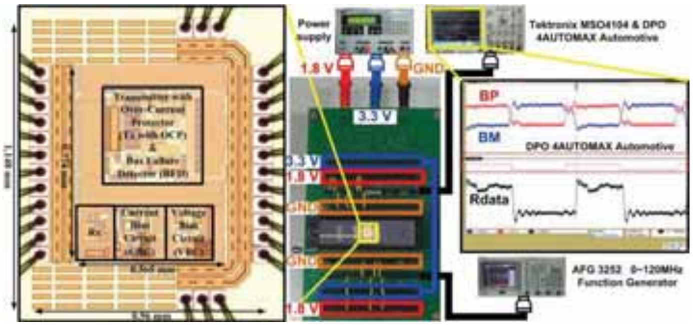

Power supply
Tektronix MSO4104 & DPO 4AUTOMAX Automotive
1.8 V
GND
3.3 V
Transmitter with Over-Current Protector (Tx with OCP) & Bus Failure Detector (BFD)
Rx
Current Bias Circuit (CBC)
Voltage Bias Circuit (VBC)
0.565 mm
0.774 mm
BP
BM
DPO 4AUTOMAX Automotive
Rdata
AFG 3252 0~120MHz Function Generator

Fig. 4. The measurement set-up, die photo, and measured waveform of the proposed system.

TABLE II
COMPARISON TABLE OF THE PROPOSED FRT AND PRIOR WORK

	This work	[2]
Technology	Typical 0.18 μm CMOS process	Typical 0.18 μm CMOS process
FlexRay version	V3.0.1	V2.1B
Core area	0.43731 mm^2	0.117 mm^2
Data rate	10 Mbps	10 Mbps
Input voltage	1.8 V DC & 3.3 V DC	1.8 V DC & 3.3 V DC
Bus Failure Detection (BFD)	Yes	No
Over-current Protection (OCP)	Yes	No
Glitch reduction	Yes	No
Power consumption (each FRT)	158.4 mW	43.01 mW

REFERENCES

[1] FlexRay Communication System Electrical Physical Layer Specification V3.0.1 (http://www.flexray.com), 2010.

[2] C.-C. Wang, C.-L. Chen, J.-J. Li, and G.-N. Sung, "A low power wake up detector for ECU nodes in an automobile FlexRay system," in *Proc. 2011 IEEE ICCE*, pp. 519-520, Jan. 2011.

[3] S.-H. Zheng, Z.-M. Lin, D.-C. Liaw, "Transceiver design for the bus driver of the FlexRay communication system," in *Proc. of SICE Annual Conf.*, pp. 3625-3630, Aug. 2010.

[4] S. Sengupta, K. Saurabh, and P.E. Allen, "A process ,voltage ,and temperature compensated CMOS constant current reference," in *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS)*, vol. 1, pp. I-325-I-328, Sep. 2004.

Reduction of the Amount of Probe -Data in Telematics Services

Yuta Nakase[1)], Taro Hiei[1)], Masashi Saito[2)], Kambe Hidetoshi[3)] and Ryozo Kiyohara[1)], *Member, IEEE*
1) Kanagawa Institute of Technology 2) Mitsubishi Electric Corporation 3) Morpho Inc.

Abstract-- **Telematics services are predicted to spread to many more consumers with the increased availability of free telematics smartphone applications. In these telematics services, servers and consumer devices such as smartphones and car navigation terminals have to communicate large amounts of various types of data by uplink and downlink. Finding methods to decrease the data size for communication is one of the most significant issues in this field because of the cost for consumers and service providers. In this paper, we propose a new data compression method for telematics, which deletes redundant data and applies an appropriate compression method for a situation. Moreover, we evaluated our method with the probe data defined by ISO22837 and report positive results.**

I. INTRODUCTION

Recently, many telematics services have become available. Vehicle information devices can show the driver a suitable route with minimum time or with no traffic jams.

These services get traffic information from sensors on the road, historical traffic information, and from smartphones (See Fig. 1). In-vehicle information devices such as car navigation systems and maps are used by many drivers who have smartphones [1]. In addition, their costs include only network traffic. Moreover, the number of vehicle information devices that can be connected with smartphones is also increasing. Therefore, large amounts of probe data are being transferred from cars to telematics service providers (TSP). However, increasing data traffic needs to be avoided because of the communication costs for both users and TSPs.

In this paper, we propose a method to reduce the size of data based on ISO22837 [2], which defines probe data format. We further evaluate our method and show our encouraging results.

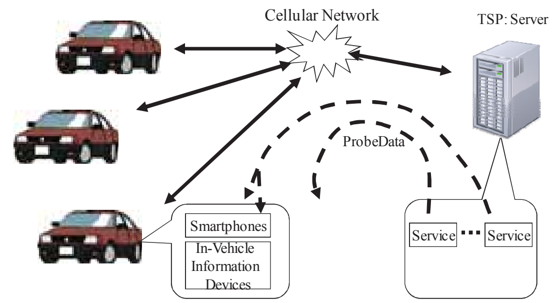


Fig. 1 Probe cars and telematics service

II. PROBE DATA

There are many telematics services, such as those that show the estimated travel time, traffic jams, incident information, and route guidance. The TSP collects large amounts of probe data in real-time, such as current position, current speed, breaking information, status of wipers, and temperature information. These data are defined and formatted in ISO22837.

However, personal data are not included in ISO22837. Therefore, the actual size of data transmitted is larger than this. Recently, many drivers have started using smartphones that can also execute car navigation applications and gather probe data.

As a result, the amount of probe data in telematics services is very large, increasing the cost for drivers and TSPs. To reduce the amount of data, we consider two approaches. One is to decrease the amount of data using a data compression method. Another is to decrease the number of communications. In this paper, we propose the application of a data compression method to probe data.

III. RELATED WORKS

There has been some research into reducing the size of probe data for telematics services. Research such as hung et al.[3] proposed a method that reduces the size of probe data based on the hypothesis that the telematics server might be able to calculate or guess a large amount of probe data if the speed and direction of the vehicles have not changed. However, the most important information is the probe data from areas where there are several vehicles. In these cases, it is difficult to guess or calculate much of the probe data.

Other research such as Adachi et al.[4] proposed a technique for solving the problem of the size of data for Dedicated Short-Range Communication (DSRC) networks. In these cases, vehicles can communicate only within a small area where can connect to the stations by DSRC. There are a lot of these areas on the road. However, it is not enough areas covered by DSRC and cellular networks are different from this type of network.

IV. COMPRESSION OF PROBE DATA

A. ISO 22837

ISO22837 defines a data structure that consists of “core data elements,” “normative probe data elements,” and values. ISO22837 also defines packages, each of which represents a part of the normative probe data.

As shown in Fig. 2, the core data elements are time stamp, latitude, longitude, and altitude. In addition, the format provides for many other fields, as listed in Table 1.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

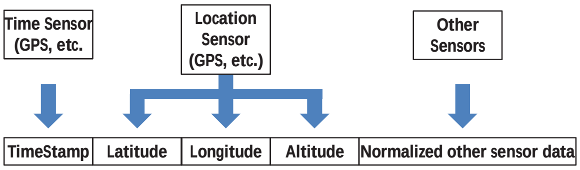


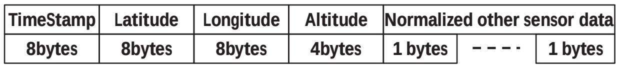
Fig. 2 Core data elements

TABLE I
EXAMPLES OF PROBE DATA FIELDS IN ISO22837

Name	Data type
Temperature	Integer[-49….50]
Wiper-Status	Integer[0…3]
Rainfall Intensity	Integer[0…999]
Exterior Lights	For each light [0…1]
Velocity	Integer[0…100]
Obstacle-Detected	0 or 1
Obstacle-Distance	Integer[0…999]
Obstacle-Direction	Integer[-90…90]
Antilock Brake	0 or 1
Traction Control	0 or 1
Vehicle Stability Control	0 or 1

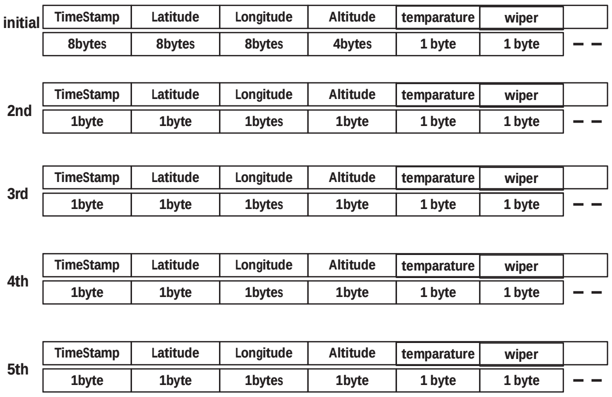
B. *Proposed Method*

There are many fields in the probe data format. However, the values of these fields change in small increments. Therefore, the values of these fields should be represented by the difference from their previous values. If each field value is represented by the difference, the total size of the data is expected to be small. Fig. 3 shows an example of the proposed format.



TimeStamp	Latitude	Longitude	Altitude	Normalized other sensor data		
8bytes	8bytes	8bytes	4bytes	1 bytes	- - - -	1 bytes

Fig. 3 Data size of core data elements



initial	TimeStamp	Latitude	Longitude	Altitude	temparature	wiper	
	8bytes	8bytes	8bytes	4bytes	1 byte	1 byte	– –
2nd	TimeStamp	Latitude	Longitude	Altitude	temparature	wiper	
	1byte	1byte	1bytes	1byte	1 byte	1 byte	– –
3rd	TimeStamp	Latitude	Longitude	Altitude	temparature	wiper	
	1byte	1byte	1bytes	1byte	1 byte	1 byte	– –
4th	TimeStamp	Latitude	Longitude	Altitude	temparature	wiper	
	1byte	1byte	1bytes	1byte	1 byte	1 byte	– –
5th	TimeStamp	Latitude	Longitude	Altitude	temparature	wiper	
	1byte	1byte	1bytes	1byte	1 byte	1 byte	– –

Fig.4. Proposed data format with difference in values

Vehicles are only able to move a limited distance in a fixed period. Therefore, the difference of positions can be represented by very small values that depend on speed and sensing intervals. If sensing intervals are less than a minute and the latitude and longitude are represented by a 64-bit integer, the differences of the latitude and longitude can be represented by less than an 8-bit integer. The proposed data format is shown in Fig. 4. These data are transferred to the TSP at fixed intervals that are larger than the sensing intervals. Moreover, various fields are required for various services. Therefore, data that depend on subscribed services are transferred only as needed.

V. EVALUATION

We evaluated the proposed method in many aspects. In Fig. 5, we show one of the results of our simulation. In the graph, A represents a non-compression method, while B shows the performance when all data are transferred with the proposed compression method. C shows performance when only the required data are transferred by the proposed compression method. The vertical axis represents the total data size and the horizontal axis represents the intervals over which the data were transferred. The sensing interval is fixed at 1 min.

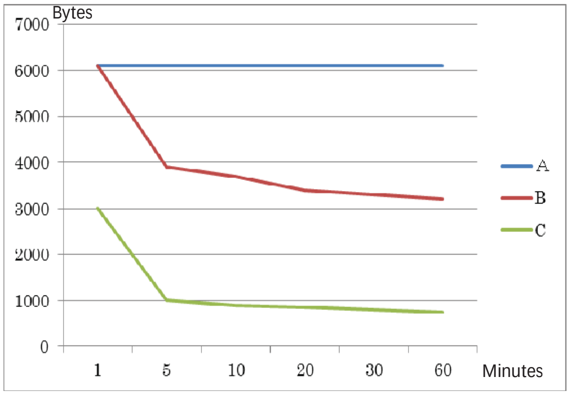


Fig.5 Result of data size

VI. CONCLUSION

We proposed a method for reducing the amount of probe data in telematics services. We showed an example of the results of our method through a simulation. The result proved that our method can effectively reduce data size.

REFERENCE

[1] M. Maekawa,T. Fujita ,A. Satou, and S. Kimura," Usage of M2M Service Platform in ITS," NEC Technical Journal, Vol.6. No.4, pp.43-47

[2] ISO 22837, "Vehicle probe data for wide area communications," International Standard, 2009

[3] T. Hung, H. Ikeda,K. Kuribayashi, and Nikolaos Vogiatzis, "Reducing the Network Load in CREPEnvironment," Journal of Information Processing, Vol.19, pp.12-24(2011)

[4] S. Adachi, R Ikeda, H. nishii, et al, "Compression Method for Probe Data," Proc. Of the 11^{th} World Congress on ITS (2004)

Traffic Information System: A Lightweight Geocast-based Piggybacking Strategy for Cooperative Awareness in VANET

Rasheed Hussain*, Junggab Son*, Hasoo Eun*, Sangjin Kim**, and Heekuck Oh*
*Hanyang University, ** Korea University of Technology and Education, South Korea

***Abstract*— As a cornerstone of VANET (Vehicular Ad Hoc Networks) system, vehicles need to have awareness information about other vehicles in the vicinity for safe driving. In this paper, we propose a lightweight geocast-based piggybacking mechanism for cooperative awareness applications in VANET. We leverage only one-hop beacons to construct both local and extended traffic views for drivers. Since multi-hop communication does not scale well, consequently our proposed scheme only exploits single-hop beacons to provide the same comparable functionality as multi-hop communication would do. The goal is achieved at the expense of a small piece of additional information (traffic classes) in the scheduled beacons. Geocast-based piggybacking has a twofold advantage over traditional flooding-based approaches; it serves as virtual multi-hop communication paradigm and it avoids broadcast storm problem.**

I. INTRODUCTION

Broadcasting is the major building block for Cooperative Awareness Applications (CAA) in VANET. This is realized by broadcasting so-called single-hop beacon messages containing the current position, speed, and other information about the vehicle. The frequency of beacons is still controversial among research community [1],[2]. High frequency of beacons affects the network quality in dense traffic regimes where the probability that two vehicles transmit the message at the same time increases. This situation causes collisions and network congestion. Traditional message dissemination schemes are based loosely on flooding where every vehicle retransmits the message with probability 1 (1-persistence approach) in order to get higher penetration rate. But flooding causes broadcast storm problem [3].

To countermeasure the aforesaid issue, different forwarding mechanisms have been proposed [3],[4] employing multi-hop communication. Tonguz et al. [4] proposed a scheme called DV-CAST to remedy the broadcast storm problem in VANET by considering dense, sparse, and disconnected traffic regimes. Their work focuses on the safety messages dissemination and includes multi-hop broadcasting to cover the maximum AoI (Area of Interest). Piggybacking is another solution where the data to be forwarded is encapsulated in normal messages. In [5], the authors put light on the impact of 'piggybacking on beacons' by setting grounds for network parameters. Mittag et al. [6] scrutinized the effect of multi-hop beaconing in VANET and compared the load on the wireless channel when the beacons are transmitted over single-hop and multi-hop.

Our work focuses on CAA and we exploit only regularly sent single-hop beacons to provide local and extended traffic views to the drivers. Our scheme leverages geocast-based piggybacking mechanism which is different from traditional piggybacking. In traditional piggybacking, data from received beacons is encapsulated in outgoing scheduled beacons and transmitted. While in our scheme, the vehicle classifies the beacons data to certain classes, and then sends only fine-grained information about the traffic in its beacons thereby extending the traffic view of the drivers from local to extended view. The degree of extension depends on the nature of application. To the best of our knowledge, our scheme is the first approach to extend the traffic view without employing complex and inefficient multi-hop rebroadcasting. Geocast-based piggybacking serves as virtual multi-hop beaconing without introducing any extra processing or computationally heavy communication overhead to the network.

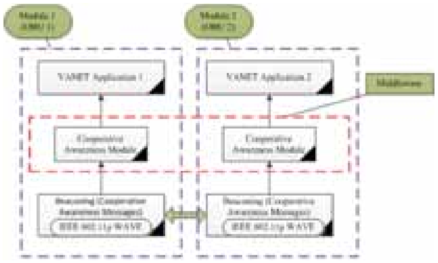

Mobile 1 (OBU 1)
Mobile 2 (OBU 2)
VANET Application 1
VANET Application 2
Middleware
Cooperative Awareness Module
Cooperative Awareness Module
Beaconing (Cooperative Awareness Messages)
IEEE 802.11p WAVE
Beaconing (Cooperative Awareness Messages)
IEEE 802.11p WAVE

Fig.1. Functional layers of VANET architecture

II. PROPOSED GEOCAST-BASED PIGGYBACKING

A. Baseline

Fig.1. illustrates the layered structure of our proposed scheme. Cooperative Awareness module serves as middleware between VANET application and IEEE 802.11p (WAVE) beaconing module. It is worth noting that beacon messages are used for variety of other purposes which are out of the scope of this paper. CAA usually follows push-based strategy for information dissemination and in our proposed scheme we only take beacon messages into account.

According to DSRC (WAVE) standard, every vehicle broadcasts beacons with a predefined frequency. Note that the optimal beaconing frequency is still an ongoing research topic. For ease of understanding, we assume static beaconing frequency for now. In addition, urban and highway scenarios exhibits different topology dynamics. For topology, we assume a straight and a curved road with vague and/or no LoS (Line of Sight). Vehicles tend to construct a local view of the traffic ahead through the received one-hop beacons.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education, Science and Technology (No. 2012-R1A2A2A01046986).

This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education, Science and Technology (MEST) (No. 2012-R1A1A2009152).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Local View and Extended View

As outlined in previous subsection, vehicles construct local view ($local_x$) after analyzing single-hop beacons. To increase the dissemination boundaries, a virtual multi-hop communication paradigm is developed by using only single-hop beacons without disturbing their normal flow. We do that by defining traffic classes and construct a knowledge base into the vehicles. The abstract format of beacon message is given below.

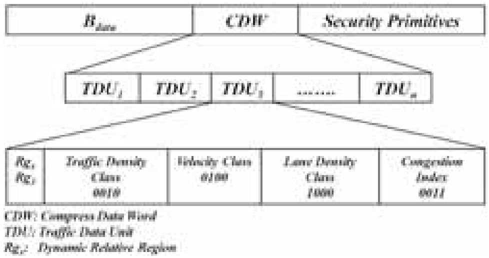


Fig. 2. Proposed beacon format

Where B_{data} is the beacon data containing current location, speed, lane information, and heading etc. of the vehicle. CDW is the combination of TDUs which represent the local views of vehicles ahead. As illustrated in Fig. 2, every TDU exhibits a certain region to which that TDU corresponds (Rg_x) and gives traffic information about that region. The region is dynamically calculated by geocast function $f_g(\ )$. For instance, $TDU_3 = \{Rg_3,0010,0100,1000,0011\}$ translates TDU_3 (Fig. 2) into bits which shows *region 3* in front of Vehicle *A* (Fig. 3), *Traffic Density Class (0010), Velocity Class (0100), Lane Density Class (1000),* and *Congestion Index (100)* respectively. The values of the aforementioned classes are stored in the knowledge base of each vehicle beforehand and it takes only few bits to represent each value. How local view is extended by these beacons, is outlined in Fig. 3.

Fig. 3 shows the local and extended views constructed by vehicle *A* with only single-hop information in hand by using the following formula.

$Extended_A=\{local_A,f_{1g}(local_A),f_{2g}(f_{1g}(local_A)),f_{3g}(f_{2g}(f_{1g}(local_A)))$
$,\ldots.., f_{ng}(f_{(n-1)g}(..f_{1g}(local_A)))\}$,

where $f_{1g}(local_A)=local_B$, $f_{2g}(local_B) = local_C$, and so on.

Hence $Extended_A=\{local_A, local_B,local_C,local_D,local_E,local_F\}$

$f_g(\ )$ selects the neighbor based on the distance from the immediate beacon source. The farther the selected vehicle is, the farther is region that is included in CWD and hence the longer is the extended view. More precisely, vehicle *A* collects information about the local views of its multi-hop neighbors *C*, *D*, *E*, and *F* through its one-hop neighbor *B*. It must be noted that the number of TDUs in a CDW must be a tradeoff between the level of extension of the view and the channel quality.

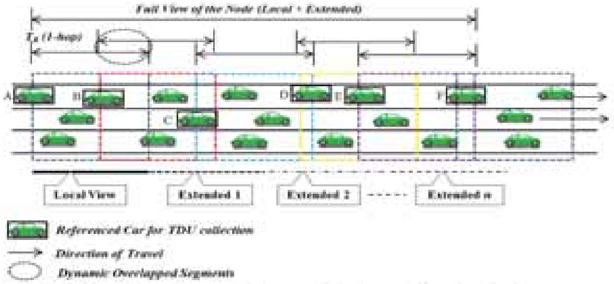

Full View of the Node (Local + Extended)
T_R (1-hop)
A
B
C
D
E
F
Local View
Extended 1
Extended 2
Extended n
Referenced Car for TDU collection
Direction of Travel
Dynamic Overlapped Segments

Fig.3. Local and extended traffic view with single-hop

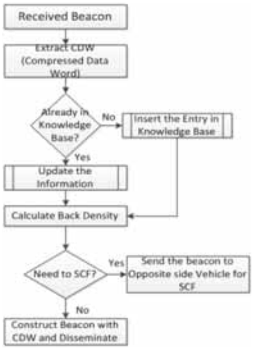

Received Beacon
Extract CDW (Compressed Data Word)
Already in Knowledge Base?
No
Insert the Entry in Knowledge Base
Yes
Update the Information
Calculate Back Density
Need to SCF?
Yes
Send the beacon to Opposite side Vehicle for SCF
No
Construct Beacon with CDW and Disseminate

Fig.4. Sequence of operations from receiver's point of view

In case of no LoS, for instance a curved road or mountain areas, we leverage RSUs (Road-Side Units). In such areas, RSUs will disseminate the traffic classes in their beacons to the nearby vehicles to construct their views. In case of sparse traffic regime behind the immediate source of the beacon, the beacon must be sent to the opposite side vehicles too, in order to carry the beacon in opposite direction until it finds a vehicle to hand over the beacon to, which is known as Store-Carry-Forward(SCF). The overall process of the proposed scheme is depicted in the Fig. 4. The process shows the flow of operations after the beacon is received from a particular source. Depending upon the traffic density behind the receiver of the beacon, it is decided whether to send the beacon to opposite direction for SCF or not.

III. Conclusions

We propose a lightweight geocast-based piggybacking mechanism for cooperative awareness applications in VANET. Our proposed scheme leverages only beacons to construct short range local view and long range extended view for VANET applications. Geocast-based piggybacking works as virtual mutli-hop beaconing mechanism. Our proposed scheme avoids broadcast storm problem. The same level of extended view is provided with only single-hop beacons as it would be done with communication inefficient multi-hop beaconing.

References

[1] R. K. Schmidt, T. Leinmuller, E. Schoch, F. Kargl, and G. Schafer, “Exploration of Adaptive Beaconing for Efficient Intervehicle Safety Communication,” *IEEE Network,* Vol. 24, No. 1, pp. 14-19, 2010.

[2] C. Sommer, O. K. Tonguz, F. Dressler, “Traffic Information Systems: Efficient Message Dissemination via Adaptive Beaconing,” IEEE Communication Maazine, Vol. 49, No. 5, pp. 173-179, 2011.

[3] N. Wisitpongphan, et al., “Broadcast Storm Mitigation Techniques in Vehicular Ad Hoc Networks,” IEEE Wireless Communications, Vol. 14, No. 6, pp. 84-94, 2007.

[4] O. K. Tonguz, et al., “DV-CAST: A Distributed Vehicular Broadcast Protocol for Vehicular Ad Hoc Networks,” *IEEE Wireless Communications*, Vol. 17, No. 2, pp. 47-57, 2010.

[5] W. K. Wolterink, G. J. Heijenk, and G. Karagiannis, “Information Dissemination in VANETs by Piggybacking on Beacons- An Analysis of the Impact of Network Parameters,” *Proceedings of IEEE Vehicular Networking Conference (VNC)*, pp. 94-101, 2011.

[6] J. Mittag, et al., “A Comparison of Single- and multi-hop beaconing in VANETs,” *Proceedings of ACM VANET’09*, pp. 69-78, 2009.

A Color Scenario of Eco & Healthy Driving for the RGB LED *Based Interface Display of a Climate Control Device*

Hyeon-Jeong SUK, *Member, IEEE*
KAIST, Daejeon, Republic of Korea

Abstract-- **The study demonstrates a process of synergizing both exploratory and confirmatory research approaches to design the color for a luminescent surface facilitated by RGB LEDs. Focusing on the relationship between color and in-door climate of automobiles, the study consists of three parts: In Part I, a workshop of ten designers was executed in which ideas were exploited to find in-car scenarios. The scenarios were evaluated based on the criteria of interesting, informative, and inspiring aspects to conclusively derive the scenario labeled "Eco & Healthy Driving"; In Part II, a user test was carried out to investigate the relationship between the attributes of luminescent color—hue, brightness, and purity- and an indoor climate condition. In the user test (n= 36), subjects were instructed to match a luminescent color to a given in-car climate condition. The user test results revealed that hue category of luminescent surface is related to temperature while brightness of luminescent color is correlated with blow level; Lastly, in Part III, by employing the results of user test, a guideline for implementing the new design scenario, "Eco & Healthy Driving" was projected for further development and application.**

I. INTRODUCTION

Researches on in-car interface take into account the ergonomic factors related to the drivers' recognition and reaction times. It is generally accepted that, unlike home appliances, the degree of complication of information displayed through the in-car interface exceeds the recognition load manageable by drivers and has a direct relationship with safety [1]. Lighting and color can be used to generate new methods of intuitively displaying information to drivers. [2] proposed that for cars that experience frequent emergency situations, using LED light on in-car interfaces can be an efficient method to induce coping. Moreover, [3] suggested that when LED is used in various different in-car interface displays, drivers should be provided with the potential to customize a range of functions from a standard list so as to best suit their environment. This provides supporting evidence for the possibility to efficiently provide users with intuitive information by implementing LED based indoor lightings and interface design. However, there is a need to further pursue research that determines how accurate the intuitive information provided through LED interface displays are communicated to the users. Therefore, by establishing a distinct relationship between certain color identities and methods of intuitive information expression, valuable knowledge for in-car interface design can be provided. This study focuses on designing the background color of interface display on in-car climate control devices (CCD here in and after) to ultimately increase user satisfaction and to increase the competitiveness of interface displays as a product of car design.

II. LUMINESCENT COLOR SURFACE AS PRODUCT COLOR

The color of products does not only concern light reflected off of object surfaces, but also highlights direct light, especially when the direct light is observed by users. As such, the display color of interface design should be recognized as an element of product design. Accordingly, using symbolic and emotional features of color to design product interfaces is a general process followed by designers. By using the meaning of color derived from previous research, it is possible to interpret the meaning of colorful light created by RGB LEDs. [4] asserted that current status information about a product can be intuitively recognized by observing the color properties of the products' status lights. The study showed that contents of information that users intuitively associate with the color of RGB LED status lights are similar to those associated with actual object color. Based on the results, perceiving the color of luminescent surface can be seen as a phenomenon comparable to perceiving the color of object.

III. OBJECTIVE

The study is comprised of three parts and three goals were set respectively: First, Part I explores the intuitive information expressed in interface displays of CCDs to set a color design strategy; Part II conducts a user test to identify the scenario of color contents for surface lights that would provide reliability at a satisfactory level for the implementation of color design strategy. Lastly, by using the empirical results of the previous goals, Part III involves arranging a proposal on a complete surface light design that can be built into in-car CCD interface displays.

IV. PART I: FINDING THE SCENARIO OF COLOR OF INTERFACE DISPLAY

A. Objective

A color implemented in interface displays should serve not only functional needs but also emotional needs. Part I attempts to exploit and find a scenario of color presentation on the interface display by conducting a workshop aimed at pinpointing a color that can most intuitively be applied to a certain scenario.

This study has been supported by the following programs: 1) "Convergent Design University Program" (H1101-12-1001); 2) "User-Centered LED Lighting System Development Enabling Daylight Spectrum for Educational/Residential Facilities" (IT R&D program of KEIT). Both funded by Ministry of Knowledge Economy.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Method

To generate color scenarios for the display of CCD, firstly, various instances detectable by CCD from inside and outside the car were collected: e.g. value of solar radiation, in-car temperature, discharge temperature, blow level, and outside temperature. The ergonomics standard data proposes the desirable temperature ranges for users depending on the type of building or space: For an indoor context with the occurrence of cognitive activities (e.g. office, auditorium, classroom, etc.) 24.5 °C ± 1.0°C is recommended for summer and 22.0°C ± 1.0°C for winter [5]. Based on the market research of domestic automobile industry in Korea, 23 °C is adopted by all Korean car manufacturers as the target in-car temperature for pleasant driving. In addition, the temperature range that a CCD can vary independently from the outside temperature is between 4 °C and 45 °C. Secondly, color attributes of interface display emitted by RGB LED were identified through the three characteristics, dominant wavelength (nm), luminance (cd/m2 or nit), and purity (%). Therefore, technically, the color scenarios can be seen as a combination between temperature-related measurements and color attributes.

During the workshop participants were instructed to investigate the potential of color presentation for the different scenarios as shown in Table 1 below. Five males and five females, for a total of 10 graduate students majoring in design were recruited to take part in the workshop. (M of age = 24.50, SD = 1.75).

C. Results

Participants were first asked to generate various matching combinations of input measurements and output colors, and then to extract five combinations that exhibited an intuitive correlation (Table 1). Each scenario was then evaluated from the user's perspective. The evaluation criteria were weighed from the users' initial thoughts on 1) "how interesting is it?" referring to whether the color emitted by the interface display helps derive an intuitive interest to its users; 2) "how informational is it?" referring to whether a scenario can actually deliver a sense of comfort in a real driving situation; 3) "how inspiring is it?" asking whether a scenario has the potential for prolonged satisfaction as oppose to temporary satisfaction that is acquired during the first encounter with the product. Scenario E was concluded by the 10 designer participants as the scenario that best satisfied all three criteria points.

Scenario E does not only take into account the excessive fuel used during winter, but also implicates health factors. Thus, Scenario E was distinguished from the other scenarios with the label "Eco & Healthy Driving" [2].

TABLE I
THE FIVE SCENARIOS GENERATED IN PART I

Scenario	The measurements that climate control device(CCD) reads	Color presentation
A	Value of solar radiation	Dynamically animated pure colors when starting a car in daytime/ Dynamically animated impure colors when starting a car at night
B	In-car temperature	Cool colors below 23 °C, Warm colors above 23 °C
C	Discharge temperature and Blow level	Change of dominant wavelength for discharge temperature, Increase of intensity[a] in proportion to blow level
D	Temperature difference between inside and outside of a car, Blow level	Increase of purity in proportion to increase of temperature difference, Increase of intensity in proportion to blow level
E	In-car temperature: above, below, or around 23 °C, Blow level	Cool colors below 23 °C, Warm colors above 23 °C, Another color around 23 °C, Increase of intensity in proportion to blow level

[a] Refers to the combination of lumination and purity, similar to the concept of how luminescent light from object color is expressed as a combination of value and chroma.

V. PART II: USER TEST TO MATCH DISPLAY COLOR AND IN-CAR CLIMATE CONDITION

A. Objective

A User Test was carried out to investigate the relationship between the attributes of color displayed in the interface of CCD and the perceived quality of the in-car climate condition. As participants were searching for the appropriate display color while directly experiencing the in-car climate, the effects that the display colors would have in actual circumstances and similar environment were also tested.

B. Method

Sixteen male and 20 female graduate students of different majors were recruited for a total of 36 participants. The average age of the subject population was 22.33 with a standard deviation of 2.95 years.

1) Collection of 45 color stimuli

In order to create a collection of display colors for the user test evaluation process, the aspects of hue category, luminance and purity were considered as follows. With reference to CIE 1976 Chromaticity Diagram, ten variations of hue categories — red, orange, yellow, yellowish green, green, greenish blue, blue, bluish purple, purple, and purplish red – were selected. The shaded region in Fig. 1 represents the color gamut emitted by the RGB LED of the CCD. Luminance is related to the intensity of a lighting surface and is measured by the unit of

[2] The concept of Scenario E is concerned with how the messages "it is hot" and "it is cold" can be transmitted to users, depending on whether the in-car temperature is higher or lower than the overall domain temperature of 23 °C, a temperature that Koreans find most pleasant. For instance, during the summer when the in-car temperature is higher than 23 °C, CCDs can make the interpretation that users will feel more comfortable with the air-conditioner turned on, whereas in the winter, it might communicate the information that turning the heater is unnecessary.

Candela per square meter (cd/m2), or “nit”. However, since this study is more concerned with visual perception rather than physical properties, brightness was used to replace luminance. Brightness, as one of photometric term, refers to the perceptual quantity of the luminious strength of luminance [6]. This brightness of luminescent surfaces was divided into two levels; strong and weak. The strong level was the greatest level of luminescence that the RGB LED could generate, and the weak level was the half level of luminescence. Purity, which indicates the vividness of luminescent surface, was also taken into consideration. As a color is closer to the boarder line (i.e. purity 100%), it becomes more vivid. Oppositely, the closer it is to the white point (x= 0.330, y= 0.330; i.e. purity 0%), the less vivid a color becomes. In this way, each hue category had four variations: two levels of brightness and two levels of purity, to initially make up 40 chromatic color stimuli respectively. Because the Chromaticity Diagram does not show the Z-axis that corresponds to luminescence level, two dots seem to overlap at one point in Fig. 1.

Next, four whites with four brightness levels were added to the color stimuli, and finally, a dark olive was included to represent the power of the interface display when it was not connected to power. Therefore, a total of 45 color stimuli made up of 40 chromatic and five achromatic stimuli were collected for the user study.

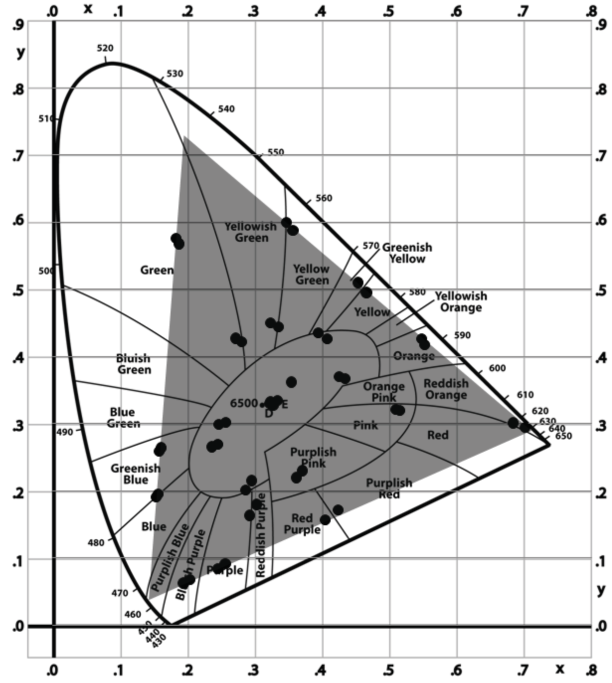


Fig. 1. The 45 color stimuli plotted in CIE Chromaticity Diagram: the shaded area is the color gamut of RGB LED of the climate control device.

In order to present the 45 color stimuli on the interface display, the appropriate R, G, and B input values had to be manually identified. While changing the values of each RGB channel from 0 to 255, a database was constructed to keep record of the x-y coordinates and luminance value of the background color. A Light Meter CS-100A of Minolta was facilitated to take the measurements. By this way, it was possible to determine the RGB input values that would render the desired 45 target color stimuli.

2) Twelve types of in-car climate

The CCD used in the user study had eight different stages of blow level with the temperature range of 5 °C to 40 °C. Four out of the eight different blow levels and three in-car temperature situations (the standard of 23 °C that Koreans find most pleasant, 30°C for the heating condition and 10 °C for the cooling condition) were implemented. Each temperature level allowed the participants to experience four blow levels, and hence, a total of 12 different in-car climate situations were prepared. Each participant was exposed to the twelve in-car climate types in random order, and were asked to find the interface display color that best expressed the given in-car climate type. Because the CCD used in the experiment was a product that was mounted on the H Motor “i30” model, the user test was also conducted in the i30 model car.

3) Matching the in-car climate with display color

In order to make it intuitive for the participants to find the interface display color based on the given in-car climate, a color palette that allows the participants to see all the 45 color stimuli at one glance was composed. Most similar colors to the 45 different types of color stimuli implemented on the CCD interface displays were formed on the laptop screen[3], and a color palette consisting of the 45 display colors was composed respectively.

The participants firstly selected one interface display color from the palette, which they felt most intuitively expressed the given experimental in-car climate condition that they experienced. Afterward, the experimenters took the colors that were most frequently selected, implemented those colors into a real CCD, and qualitatively recorded the opinions of the participants on these CCD interfaces.

C. Results and analysis of User Test

In the User Test, participants were instructed to only match one most appropriate interface display color to each of the twelve different in-car climate conditions. After having 36 participants match a color to each of the conditions, a total of 432 responses (36 selections for each of the twelve conditions) were collected. Some colors were selected several times while some were not selected at all. For example, Weak & Pale Purple was not selected by any of the 36 participants to represent any of the 12 in-car climate conditions. This indicates that Weak & Pale Purple is a color easily associated with typical in-car climate. In this way, based on the frequency of selection, the hue categories, yellow green, green, blue purple, purple, or purple red, should have lower priorities when designing the color scenario of interface display.

[3] The laptop computer used during the User Test was the Samsung Sense Mini-Notebook N310 model and had a 10.1 inch screen with a LED Back Light Unit..

TABLE II
THE FREQUENTLY SELECTED COLOR STIMULI (6 TIMES OR ABOVE) OF INTERFACE DISPLAY FOR EACH OF TWELVE TYPES OF IN-CAR CLIMATE

In-car temperature (°C)	Blow level			
	Low	Medium low	Medium high	High
Cold	Weak & Pale Greenish blue (11)[a], Strong & Pale Greenish blue (7)	Weak & Vivid Greenish blue (7)	None (None above 6)	Strong & Vivid Blue (10), Weak & Vivid Blue (7), Strong & Vivid Greenish blue (7)
Mild	Bright White (7)	Weak & Pale Yellow (7)	None (None above 6)	None (None above 6)
Warm	Strong & Pale Red (8), Weak & Pale Red (8)	Weak & Vivid Red (10)	Weak & Vivid Red (9)	Strong & Vivid Red (14), Weak & Vivid Red (8)

[a] Numbers in parentheses are the frequency

VI. PART III: IMPLEMENTATIN OF "ECO & HEALTHY DRIVING"

To implement the concept of Eco & Healthy Driving, the interface display colors from the results of Part II and the following aspects were taken into considerations: 1) First, similar to what was identified in Part II, the correlation between blue or greenish blue with cold in-car temperature and red with hot in-car temperature can be recalled intuitively. However, using greenish blue rather than blue was more desirable to represent cold temperature as greenish blue provided a greater contrast between the background color and the text and icons expressed on the interface display of a CCD, thereby improving legibility. Also, users dominantly preferred greenish blue especially for low blow level conditions; 2) Second, following what was derived from Part II, there was difficult to find a strong association between a mild 23 °C in-care temperature and hue category. Nevertheless, the association between weak & pale yellow and white and mild in-car temperature tended to exhibit a relatively more strong appeal. Among the two colors, as much as weak & pale yellow can be interpreted as yellowish white, white was selected as the main hue category to represent mild in-car temperature; 3) Third, the stronger the strength of the blow, the more vivid the chroma of the interface display color becomes, given the white hue category remains fixed.

Accordingly, to make it possible for the engineer to implement the design solution, a complete guideline with information regarding color and temperature was completed as shown in Fig. 2.

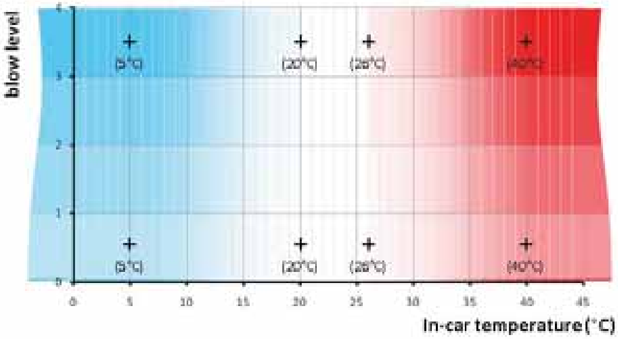


Fig. 2. The color scenario of "Eco & Healthy Driving".

VII. CONCLUSION

The challenge for designers is to find the parallel balance between diverging thinking for acquiring creative ideas and convergent thinking for finding scientific evidence. Part I of this experiment focuses on the divergent way of thinking for exploring the potential of in-car CCD to express relevant information to its users, whereas Part II focuses on the convergent way of thinking for implementing one of the potential ideas resulting from Part I and deriving reliable data. In Part II's case, it is possible to extract a clear research topic focused on color and climate perception that can be packaged as a research in the field of cognitive science. Yet there are limitations when trying to incorporate certain experiment designs and analysis methods for better experimental results into design practice. Therefore, results of Part II were directly used as resources for implementing ideas derived from Part I into Part III. It is only when the designers are able to foster their competence to persuasively implement creative ideas on their own that the creativity of the designers' ideas can be firmly recognized by others.

Furthermore, to support the new domain of design practice anchored to new technologies and LED as a new opportunity in design, designers must develop an intellectual curiosity for other academic fields. In all, the use of lighting surface as a new competitive factor in product design is greatly anticipated.

REFERENCES

[1] D. Eby, and L. P. Kostyniuk, "Driver distraction and crashes: An assessment of crash databases and review of the literature (Report No. UMTRI-2003-12)", Ann Arbor, MI.: University of Michigan Transportation Research Institute, 2003.

[2] J. Y. Choi, Y. S. Kim, S. W. Bahn, M. H. Yun, and M. W. Lee, M. W., "A study on optical layout of control buttons on center fascia considering human performance under emergency situations", *Journal of the Ergonomics Society of Korea*, vol. 29(3), 2010, pp. 365-373.

[3] P. Green, "Motor-vehicle driver interfaces". In *The human –computer interaction handbook*, J. Jacko and A. Sears, Eds. Hillsdale, NJ: Lawrence Erlbaum Associates. 2002, pp. 844-860.

[4] B. N. Baek, H. J. Suk, and M. S. Kim, "Context Information Representation Applying Light Attributes -Representation of Abstracted Context Information Using LED," *Korean Journal of Design Studies*. vol. 25(1), 2012, pp. 207-218.

[5] ISO 7730:2005(E), Ergonomics of the thermal environment.

[6] D. L. DiLaura, K. W. Houser, K.W., R. G. Mistrick, and G. R. Steffy, *The Lighting Handbook,* 10th ed., New York, NY: Illuminating Engineering Society. 2011. pp. 5. 22.

Efficient Construction of Database by Indexing and Correcting Algorithms for Personal Computed Indoor Positioning System

Jong-In Jung, Hyuk-Won Cho, Jin Cha, Jong-Kyun Hong, and Sang-Sun Lee*

***Abstract*-- To address location privacy and database maintenance problems of handheld devices, this paper proposes a personal computed indexing and correcting algorithm for an indoor positioning system.**

I. INTRODUCTION

Increasingly, personalized services are being provided based on a user's location and circumstances. Various wireless LAN-based positioning technologies are being commercialized for this purpose [1], primarily using laptops with wireless adaptors or dedicated tags and server calculations [2], [3].

We experienced commercial engine EKAHAU applied for transportation services test-bed in airport. This system is stability and comparatively high accuracy, useful tool of control for system operator. So we approach proposed system by benchmarking the EKAHAU system. However, this approach has some problems when applied to gather information on mobile devices, such as the need to install drivers and the lack of availability for private applications [4]. The RSSI-based fingerprinting positioning method employed in this paper is able to increase position accuracy by reflecting the differential signal propagation environment of each point [5], [6]. This requires that a database be constructed to cover what may be a very wide service area with a large number of wireless signals. In addition, the information must be updated each time a device's signal propagation environment changes.

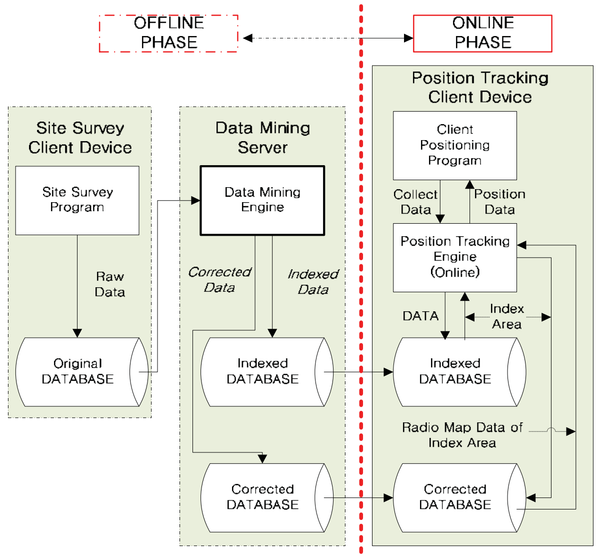


Fig. 1. Diagram of propose system.

Therefore, we performed the current study proposing an effective positioning method for handheld devices. Consequently, we propose the three database constructer algorithms that are shown in Fig.1: one for the handheld device in the online phase and two for the device in the offline phase. In order for the database to efficiently trace the position in a building-unit-wide area, we propose an indexing algorithm. When RSSI values are received in the offline phase, radio map data are created; these are then compared to the RSSI values received in the online phase to perform the positioning.

This paper describes the theoretical development of the proposed algorithms, the simulation procedure used to verify them, and the simulation evaluation results that demonstrate the performance of our proposed design.

II. CLUSTERING-BASED INDEXING ALGORITHM

In order to improve on this inefficient method, we propose an indexing data approach through RSSI signal clustering.

We use the expectation-maximization (EM) algorithm to do the clustering in our design. The EM algorithm is an efficient iterative procedure to compute the maximum likelihood (ML) estimate in the presence of missing or hidden data. It provides the most likely estimates of model parameters. Each iteration procession of the EM algorithm consists of two stages: the E-step and the M-step [7].

The EM algorithm is used for RSSI signal clustering as follows. We conducted a simulation in a 7-floor building using a smartphone in real time. We analyzed the clustering for one AP. The simulation is shown in Fig. 2. These results showed that the signal strength was decreasing, and this point was divided into 3 clusters. The number of clusters should be decided after clustering, and it will be different depending on the environment. In our test, 3 clusters were appropriate.

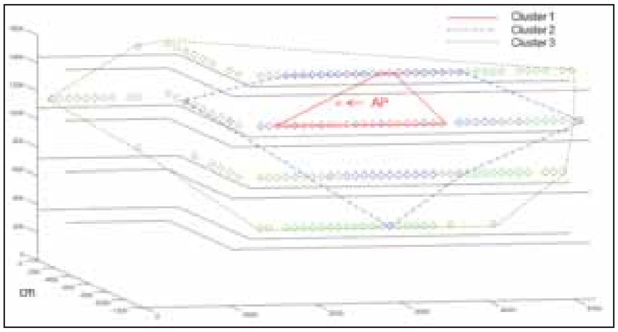


Fig. 2. Clustering result of set the three clusters by used one AP.

This work was supported by the IT R&D program of MKE/KEIT. [KI10040990, A Development of Communication Technology with UTIS & Vehicle Safety Support Service for Urban Area] This work was supported by the Brain Korea 21 Project.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

However, due to the unstable character of wireless signals, the signal strength often changes suddenly. In our test, the cluster order changed suddenly at several points.

We propose solving this problem by using the following algorithm to perform indexing based on the center of each cluster. For positioning during the online phase, the received AP propagation signal with the highest strength is used to track the location area, and, in the offline phase, the received RSSI values are treated as the central values for clustering which is done again for each AP. The final area can be recognized through the overlapped clustering area, and the accurate location within this area is calculated through indexing. It cannot be discussed here for want of space.

III. Radio Map Data Correcting Algorithm

In this section, we describe the radio map data correcting (RMDC) algorithm. The key advantage of the RMDC algorithm is that it makes selectable probabilities similar. The RMDC algorithm modifies the data based on a database analysis, which is saved through the survey. When tracking using the nearest neighbor (NN) algorithm, the classified gap value (GV) among the surrounding sample points (SPs) should be large. To evaluate each SP of the database, the gap ratio (GR) is defined by the GV. The GR is a ratio of the standard deviations of GVt and GVn. In our proposed algorithm, based on an analysis of surveyed data, poor positioning determinations are erased or modified. This erased or modified data is decided based on the RSS value of a specific MAC address.

Fig. 3 shows the results of each SP for an original database produced through the survey, in which the test tracking was performed on one day. The results in Fig. 3 show that GRs for 2, 8, 10, 12, 16, 26, 28, 32, 34, and 36 m are under 50%.

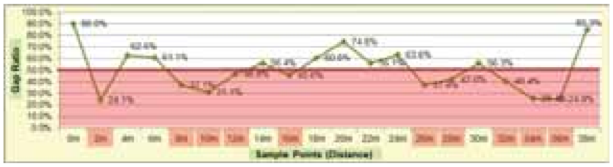


Fig. 3. Results of test tracking from the original database.

Fig. 4 shows the results of each SP for a database processed by the RMDC algorithm. As shown at the bottom of the figure, the GRs for 8, 10, 12, 16, 26, 28, 32, and 36 m are improved to greater than 50%, except for the GRs of 2 and 34 m.

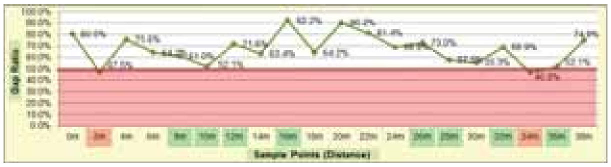


Fig. 4. Results of test tracking from the corrected database.

The RMS-error results of the total SP, shown in Fig. 5, indicate the effectiveness of the methods used in the RMDC algorithm. As shown at the bottom of the figure, on the first day of test tracking, the RMS errors for both the OD and MD were 0 m. On the last day of test tracking, the RMS error of the OD was 2.24 m, while that of the MD was about 1.79 m, which is also quite an improvement.

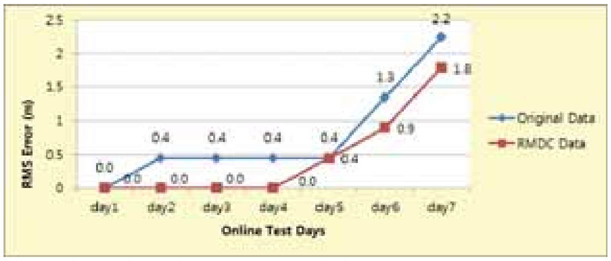


Fig. 5. The results of the RMS errors each day.

IV. Conclusion

This paper has proposed two algorithms for an efficient indoor positioning system in handheld devices. First, RSSI signals are clustered based on their signal propagation strengths. Then, the overlap area of each AP with the largest RSSI received by the user is selected, and an accurate location is calculated by the proposed database indexing constructer.

Through this indexing algorithm, the database calculated from the fingerprinting algorithm, which is similar to the K-NN, is compressed and fast approach. Our RMDC algorithm also improves the database for survey data by erasing or modifying the data that is thought to be unnecessary or error prone. Our experimental results regarding RMS errors demonstrated up to a 1.7-fold improvement in accuracy and improvements in the gap ratio of 31.5% to 37.3%, at about a month after the initial survey.

In future work, the parameters of the algorithm should be optimized. Additionally, to determine the optimal time at which to perform another survey based on a perceived change in the environment, research establishing a proper standard should be performed.

Reference

[1] Jonathan Ledlie, “Mol´e: a Scalable, User-Generated WiFi Positioning Engine,” IPIN 2011, September 2011.

[2] Y. Kong, y. Kwon, and G. Park, “Robust Localization over Obstructed Interferences for In building Wireless Applications,” IEEE Trans. Consumer Electronics, Vol. 55, No. 1, pp. 105-111, Feb. 2009.

[3] S. Y. Cho, "Localization of the Arbitrary Deployed APs for Indoor Wireless Location-Based Applications," IEEE Trans. Consumer Electronics, Vol. 56, No. 2, pp. 532-539, May. 2010.

[4] Arvin Wen Tsui & Yu-Hsiang Chuang & Hao-Hua Chu, “Unsupervised Learning for Solving RSS Hardware Variance Problem in WiFi Localization,” MOBILE NETWORKS AND APPLICATIONS, Springer, Vol 14, Number 5, pp. 677-691, Jan. 2009.

[5] A. Ladd, K. Bekris, G. Marceau, A. Rudys, L. Kavraki, and D. Wallach, “Robotics-Based Location Sensing Using Wireless Ethernet,” Proc. ACM MobiCom ’02, pp. 227-238, Sept. 2002.

[6] M. Youssef and A. Agrawala, “Handling Samples Correlation in the Horus System,” Proc. IEEE INFOCOM ’04, pp. 1023-1031, Mar.2004.

[7] Moon, T.K., "The expectation-maximization algorithm," IEEE Signal Processing Magazine, Volume: 13, Issue: 6, pp. 47-60, Nov. 1996.

OUTAGE PROBABILITY OF NETWORK-CODING-BASED COOPERATIVE COMMUNICATION SYSTEM

Yi-Lin Sung, Aldo Morales, *Senior Member, IEEE,* and Sedig Agili, *Senior Member, IEEE*

***Abstract*—Relay users play a significant role in cooperative communication system improving the overall performance by increasing reliability. First, based on code division multiple access (CDMA) techniques, we propose a network coding (NC) based CDMA cooperation, which consumes less time slots in either synchronous or asynchronous transmission. Further, NC-based cooperative system saves transmission bandwidth, in other words, we show that it reduces outage probability as compared to the traditional cooperative communication system. These results are shown considering a Nakagami-m fading channel.**

***Index Terms*—Outage probability, network coding, cooperative, code division multiple access (CDMA), relay, diversity.**

I. Introduction

TRADITIONAL cooperative communication system has been widely investigated in many papers and used in many consumer electronic mobile devices [1]-[4]. The advantage of using cooperative network coding aims at transmission performance, including high speed transmission. However, the conventional cooperation system has several problems, such as the use of time slots and high bandwidth. In a previous paper [5], we proposed a network based cooperation system. The main idea of NC-based cooperation is saving bandwidth while reducing outage probability. Moreover, based on a CDMA transmission technique [6], we solved the time slots consuming problem so that the mobile users can broadcast their messages to the receiver under asynchronous transmission without signal collision. In this paper, we further analyze the NC-based cooperation system using Nakagami-m fading channel, which is more realistic. Following this assumption, we show, by using MATLAB simulations, reduction in outage probability while saving channel bandwidth.

II. SYSTEM MODEL

In this section, first, we use CDMA based on BPSK transmission technique to broadcast signals from source t_1 to relay user and base station, as shown in Fig. 1, which can be expressed as

$$\begin{aligned} s_{t_i}(t) &= \sum_{m=1}^{M} \sqrt{\frac{2E_b}{T_b}} c_{t_i,m}(t) \cos(2\pi f_m t) \\ &\quad \times \sum_{n=-\infty}^{\infty} a_{t_i}(n) g(t - nT_b), \end{aligned} \tag{1}$$

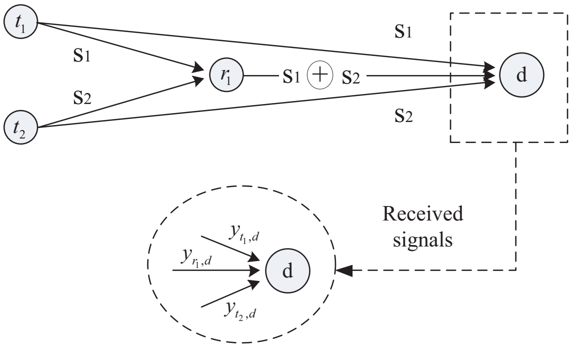


Fig. 1. A simple case depicts NC-based CC-CDMA cooperative system. Two sources t_1 and t_2 send their own message to the relay node r_1 and the destination node d at phase 1, while at phase 2, r_1 helps sources to transmit new message to d. The new message is the combined signals from sources using binary sum operation.

where E_b is the energy per bit, $c_{t_i,m}$ is the spreading code, and a_{t_i} is the digital signal. Further, after demodulation and decoding at relay, the received signals can be written as:

$$\begin{aligned} \tilde{y}_{t_i,r_j} &= \sum_{m=1}^{M} \tilde{y}^{(m)}_{t_i,r_j}(t) \\ &= \sqrt{E_b} a_{t_i}(n') \alpha_{t_i,r_j,0} e^{-j\theta_{t_i,r_j,0}} + \xi^{(DF)}_{t_i,r_j}, \end{aligned} \tag{2}$$

where α_{t_i,r_j} is a Gamma distributed random variable., θ_{t_i,r_j} is the phase, and $\xi^{(DF)}_{t_i,r_j}$ is the AWGN noise. Finally, following by maximum ratio combining (MRC), we have the maximum output SNR which is as shown:

$$(\mathrm{SNR})_o = \frac{\left(\frac{\Psi^2 \sqrt{\Upsilon}}{\sqrt{\frac{N_0}{2}}} + \frac{\Phi^2 \sqrt{\frac{N_0}{2}}}{\sqrt{\Upsilon}} \right)^2}{\Psi^2 \Upsilon + \Phi^2 \frac{N_0}{2}}. \tag{3}$$

Where Ψ, Υ, and Φ are shown in Appendix.

III. OUTAGE PROBABILITY

Following [7] and [8], the mutual information for NC-based CDMA cooperative communication in this paper is given as:

$$\begin{aligned} I_{DF} &= \frac{1}{2} \log \left(1 + \sum_{i=1}^{K} \mathrm{SNR} |\alpha_{s_i,r_j}|^2 \right) \\ &\quad + \frac{1}{2} \log \left(1 + \sum_{i=1}^{K} \mathrm{SNR} |\alpha_{s_i,d}|^2 + \sum_{j=0}^{Z} \mathrm{SNR} |\alpha_{r_j,d}|^2 \right), \end{aligned} \tag{4}$$

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

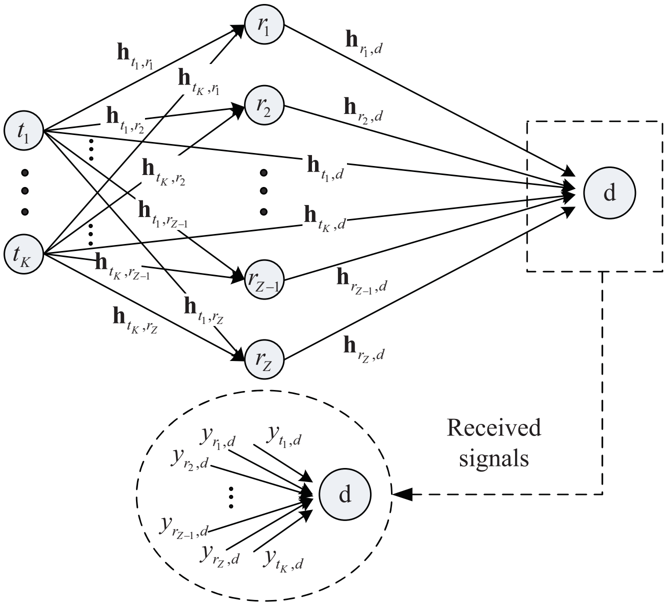


Fig. 2. General case: K source nodes with Z relay nodes. Users in this case cannot be relay.

and in Fig. 2, the general form of outage probability can be written as shown:

$$
\begin{aligned}
& P_{DF}^{\text{out}}(\text{SNR}, R) := \Pr[I_{DF} < R] \\
= \; & \Pr\left[\sum_{i=1}^{K} |h_{s_i,r_j}|^2 < \frac{2^{2R}-1}{\text{SNR}}\right] \\
& + \Pr\left[\sum_{i=1}^{K} |h_{s_i,r_j}|^2 \geq \frac{2^{2R}-1}{\text{SNR}}\right] \\
& \times \Pr\left[\sum_{i=1}^{K} |h_{s_i,d}|^2 + \sum_{j=0}^{Z} |h_{r_j,d}|^2 < \frac{2^{2R}-1}{\text{SNR}}\right].
\end{aligned}
\tag{5}
$$

IV. Numerical Analysis

Using MATLAB and considering Nakagami-m fading channel, the outage probability results for four relays are shown in Fig. 3 and Fig. 4. While at transmission rate $R = 1$ bit/sec/Hz, the performance is slightly worse than $R = 0.5$ bit/sec/Hz, thus, it shows that the higher the transmission rate ($R = 1$ bit/sec/Hz), the higher probability of failure of the system.

V. Conclusions

Consequently, to compare with [9] and [10], the numerical analysis proved that network coding does help cooperative communication while saving channel bandwidth. The new proposed system can improve communication between mobile consumer electronic devices.

VI. Appendix

For the symbols at the first section, we have Ψ, Υ, and Φ as shown:

$$\Psi = \sqrt{E_b} a_{t_1}(n')\alpha_{t_1,d,0} e^{-j\theta_{t_1,d,0}}, \tag{6}$$

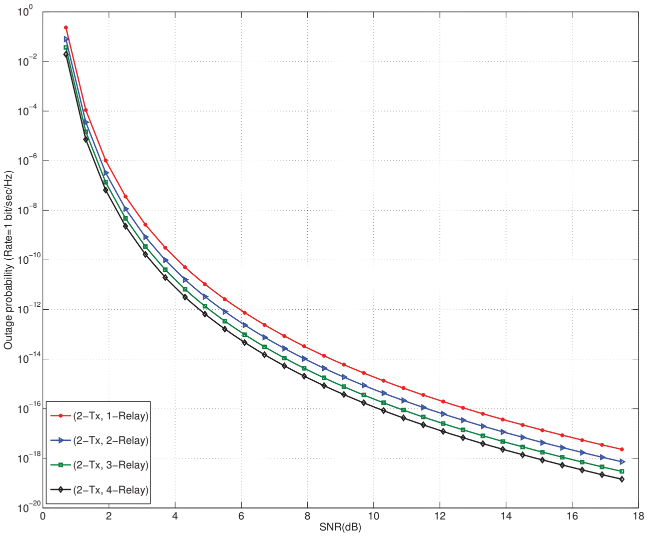


Fig. 3. Outage probability vs. SNR with $R = 1$ bit/sec/Hz, while the Gamma distribution with parameters $\alpha = \beta = 1$.

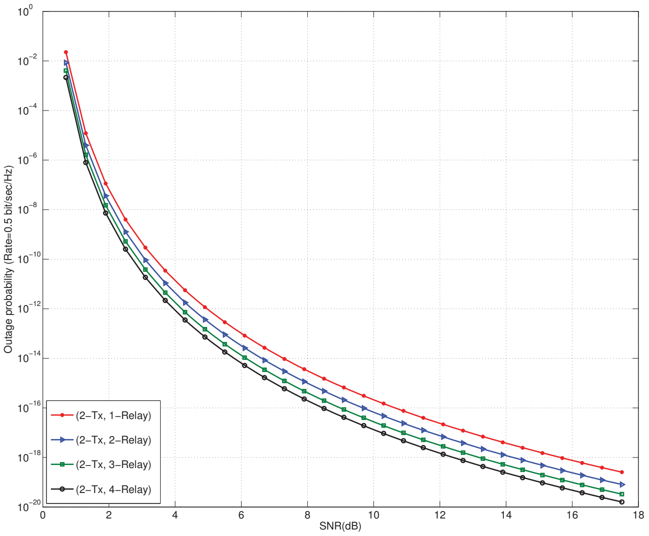


Fig. 4. Outage probability vs. SNR with $R = 0.5$ bit/sec/Hz, while the Gamma distribution with parameters $\alpha = \beta = 1$

and

$$
\begin{aligned}
\Upsilon \; = \; & E_b a_{r_1}^2(n')\alpha_{r_1,d,0}^2 e^{-2j\theta_{r_1,d,0}} \frac{N_0}{2} \\
& + E_b a_{t_2}^2(n')\alpha_{t_2,d,0}^2 e^{-2j\theta_{t_2,d,0}} \frac{N_0}{2} + \frac{N_0^2}{4},
\end{aligned}
\tag{7}
$$

and

$$\Phi = E_b a_{r_1}(n') a_{t_2}(n') \alpha_{r_1,d,0} \alpha_{t_2,d,0} e^{-j(\theta_{r_1,d,0}+\theta_{t_2,d,0})}. \tag{8}$$

References

[1] J. N. Laneman and G. W. Wornell, ”Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks,” IEEE Transaction on Information Theory, vo. 49, no. 10, October 2003.

[2] Todd E. Hunter, ”Diversity through coded cooperation,” IEEE Transactions on Wireless Communications, vol. 5, no. 2, pp. 283-289, Feb. 2006.

[3] G. Jakllari, S. V. Krishnamurthy , M. Faloutsos, and P. V. Krishnamurthy, ”On broadcasting with cooperative diversity in multi-hop wireless networks,” IEEE Journal on Selected Areas in Communication, JSAC, vol. 25, no. 2, pp. 484-496, Feb. 2007.

[4] Y. W. Hong, W. J. Huang, and C.C. Kuo, ”Cooperative Communications and Networking,” Springer, 2012.

[5] H. H. Chen, Y. L. Sung, A. Morales, and S. Agili, ”Cooperative communication with network coding,” Proceeding of ISCE, 2012.

[6] H. H. Chen, ”The next generation CDMA technologies (Hardcover),” ISBN-10: 0470022949, ISBN-13: 978-0470022948, Wiley, Sep. 17 2007.

[7] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, ”Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior,” IEEE Transaction on Information Theory, vo. 50, no. 12, December 2004.

[8] K. G. Vardhe and D. Reynolds, ”The Space-Time Coded Cooperative Diversity in An Asynchronous Cellular Uplink,” IEEE MILCOM, August 2006.

[9] Y. Zhao, R. Adve, and T. J. Lim, ”Outage Probability at Arbitrary SNR with Cooperative Diversity,” IEEE Communications Letters, vol. 9, no. 8, August 2000.

[10] H. Suraweera, P. Smith, and J. Armstrong, ”Outage Probability of Cooperative Relay Networks in Nakagami-m Fading Channels,” IEEE Communication Letter, vol. 10, no. 12, 2006.

Device-to-Device Communication Assisted Interference Mitigation for Next Generation Cellular Networks

Wonjae Shin, *Member, IEEE,* KyungHun Jang *Member, IEEE,* and Hyun-Ho Choi, *Member, IEEE*

***Abstract*--Advanced interference management schemes have become crucial for achieving the required cell edge spectral efficiency targets and to provide ubiquity of user experience throughout the network, actively discussed for 4G and beyond 4G cellular standards. On the other hands, it requires the significant channel state information (CSI) feedback overhead on consumer units so as to promote the cell-edge performance. In this paper, we explore the interplay between interference management, device-to-device (D2D) communication, and CSI feedback. To show this, a novel interference management method, *D2D communication assisted interference alignment (DIA)* is proposed, which is inspired by subspace intersection property. We believe that the proposed technology can be well applied to consumer mobile communication devices over cellular networks such as smart phones, tablets, etc.**

I. INTRODUCTION

Inter-cell interference has become the major limiting factor in next generation cellular networks with frequency reuse factor one. A promising approach to mitigate the effects of inter-cell interference is cooperative communications between the neighboring cell sites, called coordinated multipoint transmission/reception (CoMP) [1]. On the other hands, it requires the significant channel state information (CSI) feedback overhead so as to promote the cell-edge performance.

Recently, device-to-device (D2D) communications underlying a cellular coverage has recently been proposed as a means of improving the user throughput. In general, collaboration levels of D2D links could be classified into two categories: received-signal forwarding and CSI sharing scenarios. Most prior works on D2D communication have considered the received signal forwarding cases, which act as a cooperative relay in various channel models such as broadcast channel [2], interference channel [3].

In this paper, we propose an advanced interference management scheme exploiting D2D link, operating in different frequency bands compared to cellular links. If D2D cooperation can be exploited to share the CSI among user terminals through the Wi-Fi Direct or Bluetooth networks, which is one scenario of promising future wireless network [1], the proposed interference mitigation scheme needs much smaller amount of channel feedback compared to the conventional schemes with global CSI available at the BS.

Our contribution in this paper is as follows. First, we newly introduce interference mitigation method via D2D links for the future cellular networks, called *D2D communication assisted interference alignment (DIA)*. Based on the numerical results, we demonstrate that the proposed DIA outperforms the conventional multi-user MIMO schemes on the cell outer area.

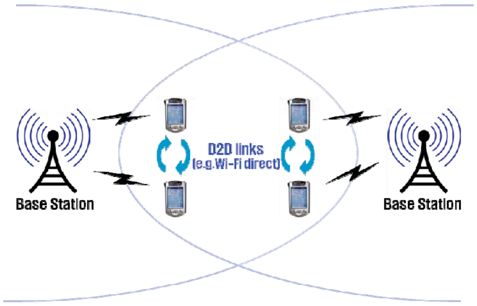


Fig1. D2D communication assisted next generation cellular networks

II. SYSTEM MODEL

Fig. 1 shows the system model of a cellular network assisted by D2D links. The system consists of two BSs with M antennas per BS and K users with N receive antennas per user in each cell, and we will cover general network topology in the full version of the paper. For notation convenience, we refer to the k-th user in the i-th cell as user [k,i].

III. NEW INTERFERENCE MITIGATION WITH HELP OF D2D LINK

In this section, we introduce a new interference mitigation method for both inter-cell interference (ICI) and inter-user interference (IUI) caused by the broadcast nature of wireless medium in the two-cell two-user networks, i.e., $K=2$ with help of D2D connections, and investigate the benefits of exploiting of D2D links compared with existing schemes in terms of multiplexing gain and the amount of channel feedback.

A. Motivating example for $(M,N,K)=(3,2,2)$

To explain our DIA scheme and its benefits clearly, we start with a simple case of $(M,N,K)=(3,2,2)$ as shown in Fig. 2. The BS 1 wants to deliver two symbols, $s^{[1,1]}$ and $s^{[2,1]}$, to the user [1,1] and user [2,1] using the transmit beamforming vectors $\mathbf{v}^{[1,1]}$ and $\mathbf{v}^{[2,1]}$, respectively. In general, for given receive beamforming vectors, the minimum number of transmit antennas is 4 so that the transmit beamforming vectors cancel out all ICI and IUI. On the contrary, our proposed interference alignment scheme can remove both ICI and IUI with 3 transmit antennas by performing interference alignment. In the following four steps, we present our transmit and receive beamforming design method (i.e., DIA) without the need of global CSI.

Step 1: Sharing interfering channels among users

In order to share CSI between co-located cell-edge users, user terminals perform D2D communication using WPAN such as Wi-Fi Direct or Bluetooth. To be specific, the user [1,2] and the user [2,2] exchange CSI of interfering channels to cooperatively design receive beamforming vectors

Wonjae Shin, KyungHun Jang are with the Signal & Systems Lab., Samsung Electronics Co. Ltd., Korea (email: {wonjae.shin, khjang}@samsung.com)
Hyun-Ho Choi is with Department of Electrical, Electronic and Control Engineering, Hankyong National Univ., Korea (email: hhchoi@hknu.ac.kr)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Step 2: Designing the receive beamforming vectors

By using CSI of interfering channels acquired in the *step 1*, the user [1,2] and user [2,2] design the receive beamforming vectors $\mathbf{w}^{[1,2]}$ and $\mathbf{w}^{[2,2]}$, so that the effective ICI channels from the BS 1 are aligned with each other, which is

$$s\mathrm{pan}\left(\mathbf{H}_1^{[1,2]^\dagger}\mathbf{w}^{[1,2]}\right) = s\mathrm{pan}\left(\mathbf{H}_1^{[2,2]^\dagger}\mathbf{w}^{[2,2]}\right)$$

where span($\mathbf{A}$) and $\mathbf{A}^\dagger$ denotes the space spanned by the column vectors of a matrix $\mathbf{A}$ and the conjugate transpose matrix of $\mathbf{A}$, respectively. Note that the channel matrix $\mathbf{H}_j^{[k,i]}$ is the N×M matrix from the BS j to the user [k,i]. We can find out the intersection subspace satisfying the above condition by solving the following matrix equation,

$$\underbrace{\begin{bmatrix} \mathbf{I}_M & -\mathbf{H}_1^{[1,2]^\dagger} & \mathbf{0} \\ \mathbf{I}_M & \mathbf{0} & -\mathbf{H}_1^{[2,2]^\dagger} \end{bmatrix}}_{6\times 7} \begin{bmatrix} \mathbf{h}_1^{ICI} \\ \mathbf{w}^{[1,2]} \\ \mathbf{w}^{[2,2]} \end{bmatrix} = \mathbf{M}_1\mathbf{x}_1 = \mathbf{0}$$

where $\mathbf{h}_1^{ICI}$ implies the direction of aligned effective interference channels from the BS 1 to the user [1,2] and user [2,2] after applying the receiver beamforming vectors. Since the size of the matrix $\mathbf{M}_1$ is 6x7, it has one dimensional null space. Hence, the receive beamforming vectors for ICI channel alignment can be obtained explicitly with probability one.

Step 3: Feedback the effective channels to the BS

Each user feeds back equivalent channels after applying the receive beamforming vectors determined in the *step 2* instead of channels matrix itself through uplink feedback channels for the cellular networks to its corresponding BS. To be specific, the user [1,2] is required to feed back both the effective serving and interfering channel vectors after applying the receive beamforming vectors.

Step 4: Choosing the transmit beamfroming vectors

Since the effective ICI channels are aligned with each other, the BS 1 can consider two different ICI channel vectors as a one ICI channel vector which spans one dimensional subspace as shown in Fig. 2. Thus, the transmit beamforming vectors should be designed with the effective channel as follows:

$$\mathbf{v}^{[1,1]} \subset \mathrm{null}\left(\left[\left(\mathbf{w}^{[2,1]^\dagger}\mathbf{H}_1^{[2,1]}\right)^\dagger \quad \mathbf{h}_1^{ICI}\right]^\dagger\right)$$

$$\mathbf{v}^{[2,1]} \subset \mathrm{null}\left(\left[\left(\mathbf{w}^{[1,1]^\dagger}\mathbf{H}_1^{[1,1]}\right)^\dagger \quad \mathbf{h}_1^{ICI}\right]^\dagger\right)$$

where null($\mathbf{A}$) denotes a basis for the null space of a matrix.

IV. Simulation Results and Discussions

We consider a linear cell layout in which two BSs exist and the cooperating MS pair in each cell moves on the line connecting these two BSs. For comparison, we adopt typical multi-user MIMO schemes: block diagonalization (BD) and zero-forcing (ZF) beamforming. Based on the evaluation methodology of 3GPP [1], we choose the inter-BS distance of 500 m, BS transmission power of 43 dBm, and the distance-dependent pathloss L=128.1+37.6log10(R) [dB] where R is a distance in km. Fig. 3 shows the achievable sum rate as the cooperating MS pair moves from the middle of two BSs to its serving BS. As the MSs move into the inside of its serving cell (i.e., the MS position increases), the rate of all schemes is improved by the increase of SNR. Compared to the conventional BD and ZF, the proposed DIA shows better performance up to 150 m from the cell boundary, but shows worse performance as the MSs go to the cell center. This is because the per-cell multiplexing gain (i.e., the number of independent data streams transmitted simultaneously) of the proposed IA is two on the overall cell area, however that of the conventional schemes is three on the cell inside but approaches zero at the cell boundary since they are originally designed without respect to inter-cell interference. This leads us to conclude that the proposed DIA should be adaptively applied according to the level of inter-cell interference.

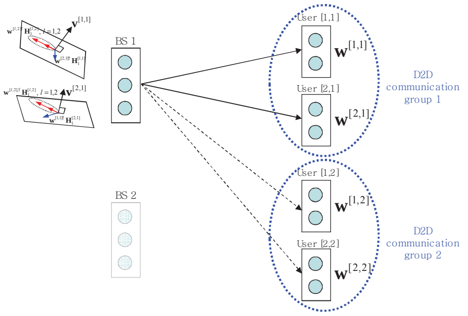

$\mathbf{w}^{[l,2]\dagger}\mathbf{H}_1^{[l,2]}$, $l=1,2$
$\mathbf{v}^{[1,1]}$
$\mathbf{w}^{[2,1]\dagger}\mathbf{H}_1^{[1,1]}$
$\mathbf{w}^{[l,2]\dagger}\mathbf{H}_1^{[l,2]}$, $l=1,2$
$\mathbf{v}^{[2,1]}$
$\mathbf{w}^{[1,1]\dagger}\mathbf{H}_1^{[2,1]}$
BS 1
BS 2
User [1,1]
$\mathbf{w}^{[1,1]}$
User [2,1]
$\mathbf{w}^{[2,1]}$
D2D communication group 1
User [1,2]
$\mathbf{w}^{[1,2]}$
User [2,2]
$\mathbf{w}^{[2,2]}$
D2D communication group 2

Fig. 2. The concept of the proposed IA scheme for two-cell scenarios

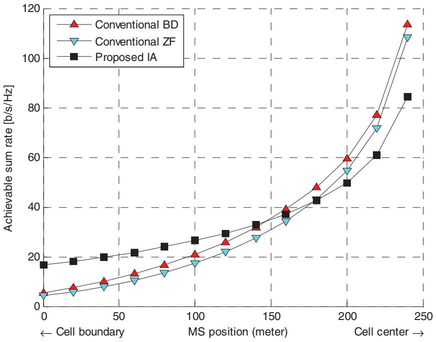


Fig. 3. Simulation results for linear cell layout as a function of MS position

V. Conclusion

In this paper, we propose a novel cross-network cooperative protocol between D2D and cellular networks, one scenario of promising future wireless networks. Based on the shared CSI through D2D links, transmit and receive beamforming vectors can be jointly constructed so that every ICI and IUI is completely removed with lower CSI feedback overhead, thereby achieving much higher data rate for cell-edge users, such as smart phones, tablets.

Reference

[1] 3GPP TR 36.819 V2.0.0, “Technical specification group radio access network: coordinated multi-point operation for LTE physical layer aspects (Release 11),” Mar. 2010.

[2] R. Dabora and S. D. Servetto, “Multi-user MISO broadcast channel with user-cooperating decoder,” IEEE Trans. Info. Theory, vol. 52, pp. 5438-5454, Dec. 2006.

[3] B. W. Khoueiry and M. R. Soleymani, “Destination cooperation in interference channel,” *in Proc of IEEE ICCE*, Las Vegas, USA, Jan. 2012.

Deafness-aware MAC Protocol for Directional Antennas

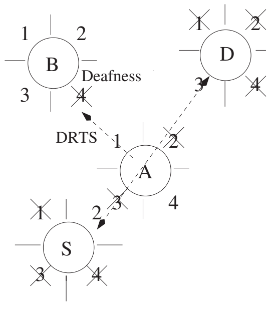
Woongsoo Na, Laihyuk Park, and Sungrae Cho

***Abstract*—In this paper, we propose a new directional MAC protocol referred to as deafness-aware MAC (DA-MAC) distinguishing the deafness from the collision. Although a significant number of directional MAC protocols have been proposed including [1]–[3], they have not comprehensively resolved the deafness problem. With the expense of an additional narrow-band antenna, we provide distinction between the deafness and the collision. Through performance analysis, we show our DA-MAC protocol can significantly outperform the existing techniques with respect to the throughput, energy consumption, and deafness duration.**

I. Introduction

IN the near future, wireless technologies supporting more than 3Gbps will emerge in order to accommodate uncompressed HDTV video and consumer electronic (CE) applications with very high throughput such as HD digital cameras and high-volume storages. One of the core technologies that industries are interested in is to use directional antennas, through which CE devices can obtain benefits such as better spatial reuse and longer transmission range. For this reason, standardization organizations such as IEEE 802.11ad and IEEE 802.15.3c have much attentions on MAC protocols using directional antennas (or *directional MAC*). Despite of these merits, directional MAC protocols suffer from the deafness problem. The deafness problem occurs if a node does not answer a directional RTS (DRTS) frame addressed to it. Consequently, the originator of the DRTS will try more DRTS frames while increasing the contention window, during which messages toward the other nodes are subject to be blocked.

Although a significant number of directional MAC protocols have been proposed, they have not comprehensively resolved the deafness problem. For instance, in [3], tried to solve the deafness problem through disclosing one's transmission information to all of its neighboring nodes. Although this technique costs multiple RTS/CTS overheads, it might not solve the deafness problem. Fig. 1 is an example of the deafness problem caused in [3] (see the explanations in Fig. 1.). Scheme in [2] exploited a local table that maintains potential senders which previously transmitted an advance notice. The advance notice informs the receiver that the sender will transmit data next time. [2] used an additional RTS (A-RTS) frame to notify potential senders to wait until a node finishes transmission so that it avoids deafness. Even if [2] tries to avoid the deafness problem by the advance notice, they cannot resolve the deafness problem in some cases. For example (in Fig. 1), although A transmits the DRTS to B, node B cannot hear the frame since nodes S and D are in communication and thus beam 4 of node B is blocked, causing the deafness problem. The other way to mitigate the deafness problem is proposed in [1]. [1] tried to distinguish the deafness from the collision using a tone signal. In [1], the sender and the receiver transmit a tone omnidirectionally after their communication. The problem of [1] is that if communication time increases, the deafness problem will be deteriorated.



At t=t$_0$, S has data to D.
At t=t$_1$, S transmits DRTS to D.
At t=t$_2$, beam 4 of B and 3 of A are blocked.
At t=t$_3$, D responses DCTS to S.
At t=t$_4$, beam 2 of A is blocked and S and D are in communication.
At t=t$_5$, A has data to B.
At t=t$_6$, A transmits DRTS to B however, B will not hear the DRTS since beam 4 of B is blocked.

× indicates the corresponding beam is blocked since the beam is not engaged in communication; or the beam causes interference to on going communication.

Fig. 1. An example of the deafness problem (4 antenna beams).

In this paper, we propose a new directional MAC referred to as deafness-aware MAC to completely filter out the deafness problem by distinguishing the deafness from the collision.

II. The Impact of Deafness

To verify the impact of the deafness, we analyze the deafness probability of the original DMAC protocol [4] with the Markov model. Fig. 2 shows the state transition diagram of a node. Table I presents the symbols which have been used in analysis. Due to space limitation, we omit the detailed derivations of the the steady state probabilities that a node stays in each of the individual states. Then, the steady-state probability is calculated as

$$S_c = \frac{\tau}{1-p_{cc}}S_i = \frac{\tau}{(1-\tau)^{N-2}(1-(1/M))-f^m}S_i \quad (1)$$

$$S_t = \frac{\tau(1-\tau)^{N-2}(1-(1/M))}{(1-\tau)^{N-2}(1-(1/M))-f^m}S_i \quad (2)$$

$$S_d = \tau(1-\tau)^{N-2}(1-\frac{1}{M})S_i = S_r \quad (3)$$

where N denotes the number of nodes and M the number of beams of a node. By (1)-(3), we calculate the deafness probability as

$$T_{deaf} = \frac{S_cE[T_c]}{S_cE[T_c]+S_iE[T_i]+S_tT_t+S_rT_r+S_dT_d} \quad (4)$$

In (4), we can observe that a node *almost cannot* communicate (i.e., T_{deaf} = 0.981) in a fairly normal topology (N = 30 and M = 4) where T_{deaf} is interpreted as the ratio of total communication time in deafness state.

The authors are with the School of Computer Science and Engineering, Chung-Ang University, South Korea; E-mail: srcho@cau.ac.kr.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

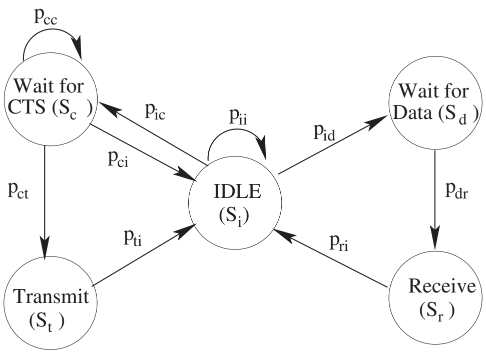


Fig. 2. State diagram of the DMAC protocol [4].

TABLE I
THE GLOSSARY

Symbol	Definition
S_x	The steady-state probability of the Markov chain in state x
p_{xy}	The transition probability where state x changes to y
T_x	The time duration that a node stays in the state x
τ	The probability that a node transmits in a random slot time
f	The probability of network failure
m	The retransmission limit

III. The DA-MAC Protocol

The proposed scheme is based on dual channel: data and control channels. We assume these two channels do not interfere with each other. Therefore, the same frame can be transmitted through the dual channel simultaneously. For simultaneous transmission, dual directional antennas are attached to the MAC unit. On the control channel, only the DRTS and DCTS frames are exchanged while DRTS/DCTS/DATA/ACK can be exchanged on data channel. If a sender has data to send, it transmits the DRTS on both channels to the receiver. The receiver of the DRTS responds with DCTS on both channels and continuously senses the control channel while using the data channel to receive data. On successful transmission, the receiver transmits ACK to the sender on the data channel. The reason for using both data and control channels to transmit DRTS is to distinguish the deafness from the collision. Suppose that a node has data to a destination, it transmits a DRTS frame on the data and control channels. Even if the destination cannot hear the DRTS frame on the data channel (destination may be engaged in communication with other node), it can hear the DRTS frame on the control channel. The destination then replies to the sender with the DCTS frame on the control channel. Since the sender receives only one DRTS frame on the control channel, it can determine that the destination is a deafness node. Now, if there is a collision with the DRTS frame at the destination, the DRTS frame can never be received at the destination on both channels. Therefore, the DCTS frame will not be transmitted to the sender. The sender can determine this case as a collision.

IV. Performance Evaluation

In this section, the DA-MAC is compared with Tone-DMAC [1], AN-DMAC [2], and CRCM [3]. The packet size and inter-arrival time is assumed to be 1024 bytes and 30 usec, respectively. The data rate is 54Mbps. We also assume that a node consumes 9.6mW/sec in the busy state and 3.0mW/sec in the idle state. Table II shows the performance of each scheme. As shown in the table, the DA-MAC can significantly reduce the deafness duration compared with the other schemes. Also, the aggregate throughput of the DA-MAC achieves the best performance compared with other schemes. One of interesting points is that, our scheme has benefits of deafness avoidance capability and energy-efficiency. Fig. 3 shows the aggregate throughput versus arrival rate. The aggregate throughput increases as the number of nodes and antennas increases for all schemes. This is because growth of number of nodes generates more traffic load and more narrower beam stimulates the spatial reuse.

TABLE II
PERFORMANCE EVALUATION (DATA RATE = 54MBPS, M = 4, N = 20)

Scheme Metric	DA-MAC	[1]	[2]	[3]
Deafness Duration (msec)	0.8	4	18.3	16
Aggregate Throughput (Mbps)	174	111	105	69
Energy Consumption (mW/frames)	0.004	0.007	0.007	0.012

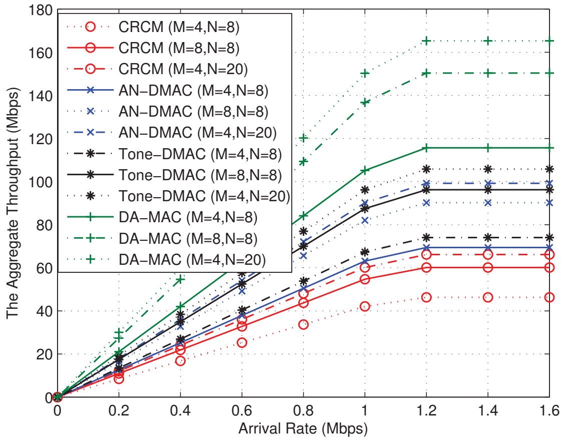


Fig. 3. The aggregate throughput vs. arrival rate.

V. Conclusion

In this paper, we proposed the DA-MAC which distinguishes the deafness from the collision by exploiting dual channel. Performance evaluation shows that the aggregate throughput can be improved in the DA-MAC against the existing schemes by up to 56%, and the deafness duration can be reduced by up to 80%. Although our proposed scheme uses dual interface, the energy consumption also can be reduced by up to 43%.

References

[1] R. R. Choudhury *et al.*, "Deafness: A MAC Problem in Ad Hoc Networks when using Directional Antennas," *in Proc. of ICNP,* 2004.

[2] J. Feng *et al.*, "A deafness free MAC protocol for ad hoc networks using directional antennas," *in Proc. of ICIEA,* 2009.

[3] G. Jakllari *et al.*, "Handling asymmetry in gain in directional antenna equipped ad hoc networks," *in Proc. of PIMRC,* 2005.

[4] Y. B. Ko *et al.*, "Medium access control protocols using directional antennas in ad hoc networks," *in Proc. of INFOCOM,* 2000.

Fast Three-Dimensional Node Localization in UWB Wireless Sensor Network Using Propagator Method
Digest of Technical Papers

Hong JIANG, *Member*, *IEEE* , Yu ZHANG, Haijing CUI, and Chang LIU

Abstract—**This paper presents a fast 3D node localization algorithm for UWB wireless sensor network employing modified propagator method for multipath time delay estimation and 3D Chan algorithm for determining the position of sensor nodes. It enhances the estimation accuracy while requires neither spectral peak searching nor covariance matrix estimation.**

I. Introduction

Ranging and localization of unknown sensor nodes in wireless sensor networks (WSN) [1]-[3] have drawn considerable attention in environmental monitoring, health tracking, smart home, M2M, etc. So far, most localization algorithms in WSN are only applicable for two-dimensional (2D) networks, However, in many actual environment, nodes are placed in three-dimensional (3D) terrains, such as workshops, forests, oceans, etc. Although some 3D localization algorithms have been proposed [4],[5], many aspects should be improved in accurate localization, small computational amount, robustness in multipath, energy saving, fast executing, etc. Since the energy and the processing power carried by sensor nodes are limited, the research on effective methods for 3D node positioning is of great significance.

This paper investigates 3D localization problem for ultra wideband (UWB)-based WSN using multipath time-delay measurement. Time-delay estimation problem has been studied with a variety of super-resolution techniques, such as Matrix Pencil, MUSIC, TLS-ESPRIT. Compared with correlator methods, they can increase time-resolution even if time delay is smaller than a pulse width. Unfortunately, these techniques increase the complexity of WSN implementation. The propagator method (PM), developed in [6], is a subspace method for direction-of-arrival (DOA) estimation. It can avoid the estimation and eigen-decomposition of the covariance matrix of the received signals which is the main computational burden in traditional subspace methods. However, spectral peak searching through all the space is needed in PM method, which increases computational complexity. Under the principle of low cost, low complexity and low power consumption of node equipment, considering multipath effect, we put forward a fast range-based node localization method for UWB wireless sensor networks in this paper. We develop a modified propagator method (MPM) for time-of-arrival (TOA) estimation in frequency domain, which enhances the estimation accuracy and requires neither spectral searching nor covariance matrix estimation. The computational load is lower than traditional subspace methods. Furthermore, 3D Chan algorithm combined with multilateral localization instead of trilateral localization is developed in the paper to accurately determine the physical coordinates of a group of sensor nodes.

II. Signal Model

Assume that a UWB pulse is transmitted from an unknown node to an anchor node by L paths. In the q-th snapshot, $q=1,\cdots,Q$, the received signal can be expressed as

$$y^{(q)}(t)=\sum_{l=1}^{L}\beta_l^{(q)}p(t-\tau_{\mathrm{TOA}}-\Delta\tau_l)+w^{(q)}(t)=\sum_{l=1}^{L}\beta_l^{(q)}p(t-\tau_l)+w^{(q)}(t) \quad (1)$$

where $p(t)$ is UWB waveform. τ_{TOA} denotes the TOA of the unknown node, $\Delta\tau_l$ and $\beta_l^{(q)}$ represent the relative delay and time-varying complex fading amplitude of the l-th path, respectively. $\Delta\tau_1=0$. $\tau_l=\tau_{\mathrm{TOA}}+\Delta\tau_l$ denote the propagation time of the l-th path. $w^{(q)}(t)$ is noise. The discrete frequency-domain representation of the identified channel is written as

$$H^{(q)}(k)=\sum_{l=1}^{L}\beta_l^{(q)}e^{-j\frac{2\pi k\tau_l}{KT}}+V^{(q)}(k)=\sum_{l=1}^{L}\beta_l^{(q)}z_l^k+V^{(q)}(k) \quad (2)$$

for $k=0,1,\cdots,K-1$. $z_l=e^{-j\frac{2\pi\tau_l}{KT}}$ contains the estimated parameter τ_l. Collecting the data of Q snapshots from (2) yields

$$\mathbf{H}=\mathbf{Z}(\tau)\mathbf{B}+\mathbf{V} \quad (3)$$

where $\mathbf{B}\in\mathbb{C}^{L\times Q}$, $\mathbf{H}\in\mathbb{C}^{K\times Q}$, $\mathbf{V}\in\mathbb{C}^{K\times Q}$, and $\mathbf{Z}(\tau)\in\mathbb{C}^{K\times L}$. The problem interest is to estimate parameter τ_l based on (3).

III. The Algorithms

A. *Multipath Delay Estimation Algorithm Using Modified Propagator Method* (*MPM*)

The steps of MPM algorithm are as follows:

i) Divide $\mathbf{H}$ into $\mathbf{H}_1$ and $\mathbf{H}_2$ consisting of the first and last $K-1$ rows of $\mathbf{H}$, respectively.

ii) Compose $\mathbf{H}_1$ and $\mathbf{H}_2$ to form a $2(K-1)\times Q$ matrix $\mathbf{X}=\left[\mathbf{H}_1^T\mathbf{H}_2^T\right]^T$. Partition $\mathbf{X}$ into two matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, such that $\mathbf{X}=\begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{bmatrix}\begin{matrix}\}L\\ \}2(K-1)-L\end{matrix}$.

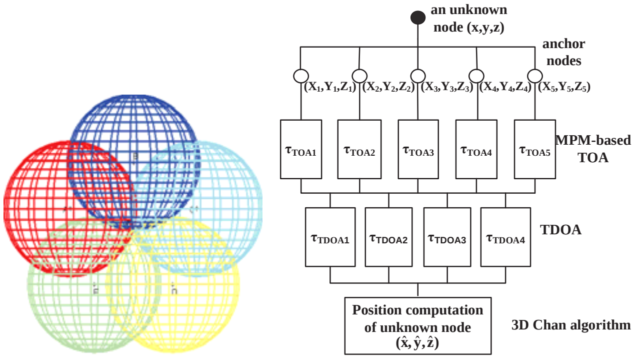
iii) Solve $\mathbf{P}^H$ by $\mathbf{P}^H=\mathbf{X}_2\mathbf{X}_1^H\left(\mathbf{X}_1\mathbf{X}_1^H\right)^{-1}$, where $\mathbf{P}^H$ is a propagator operator. Let $\tilde{\mathbf{P}}=\begin{bmatrix}\boldsymbol{I}_L\\ \mathbf{P}^H\end{bmatrix}$, evenly divide $\tilde{\mathbf{P}}$ into two $(K-1)\times L$ matrices $\tilde{\mathbf{P}}_1$ and $\tilde{\mathbf{P}}_2$, and solve $\mathbf{\Psi}=\tilde{\mathbf{P}}_2\left(\tilde{\mathbf{P}}_1^H\tilde{\mathbf{P}}_1\right)^{-1}\tilde{\mathbf{P}}_1^H$.

iv) Obtain the eigen-value matrix $\mathbf{\Phi}$ of $\mathbf{\Psi}$. Therefore, $\hat{\tau}_l=\frac{KT}{2\pi}\mathrm{angle}(\lambda_l)$, where λ_l is the eigen value of $\mathbf{\Psi}$, $l=1,\cdots,L$.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. 3D CHAN Algorithm and Multilateral Computation for 3D Node Localization

According to the results of MPM-based multipath delay estimation, the TOAs and distances from a node to multiple anchor nodes can be determined. Here, Chan algorithm [7] is extended to 3D, the nonlinear equations are accurately solved to obtain 3D positions of unknown nodes. The geometry relationship of cross points of spheres using five anchor nodes and the localization process are shown in fig.1.(a)(b).



(a) 3D localization geometric relationship (b) Localization process

Fig.1. Node localization based on TOA and 3D Chan algorithm

IV. SIMULATIONS

A. MPM-based multipath delay estimation results

Fig. 2 shows the transmitted UWB signal and the received 5-path superposition signal.

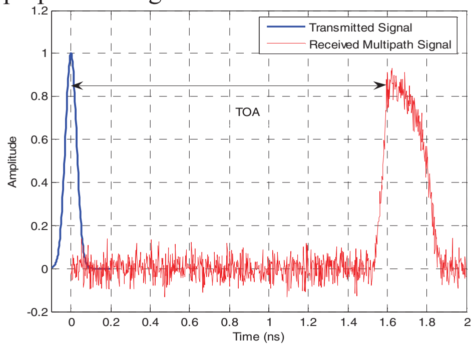


Fig. 2. Transmitted signal and received multipath signal of UWB.

Table.1 shows the estimation results of 5-path time delays using the proposed MPM algorithm. SNR=10dB. Q=100. K=1000. T=0.02ns. L=5. The simulations confirm good resolution of the proposed algorithm in multipath.

Table.1 MPM-based delay estimation results when SNR=10dB

$\hat{\tau}_l$	True value (ns)	Estimation value (ns)
Path 1	1.60	1.576
Path 2	1.65	1.629
Path 3	1.70	1.680
Path 4	1.75	1.731
Path 5	1.80	1.784

B. Position Computation Results using 3D Chan Algorithm

In fig. 3, we randomly generate 100 unknown nodes in a $2m\times2m\times2m$ space, and locate them using 3D Chan algorithm. It shows that the coordinates of these nodes can be estimated with high accuracy even with low density of anchor nodes.

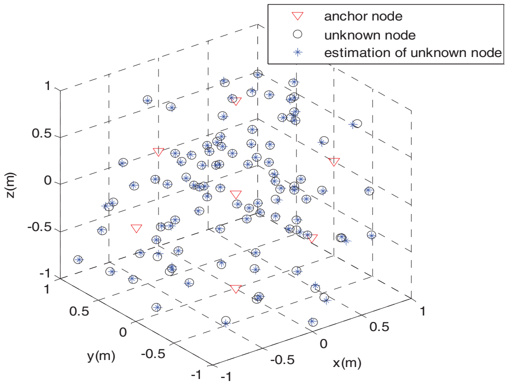


Fig. 3. 3D Node Localization (100 unknown nodes, 7 anchor nodes)

C. Accuracy of the Proposed Algorithm

In fig.4, SNR=10 dB, the number of anchor nodes varies from 5 to 8. We generate randomly 100 unknown nodes and localize them using traditional Matrix Pencil-based estimation algorithm and the proposed MPM algorithm.

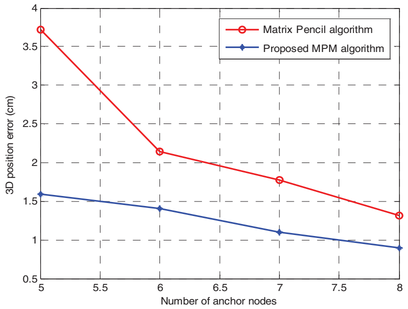


Fig. 4. Effect of the number of anchor nodes

The simulations shows better performance of the proposed MPM algorithm compared with Matrix Pencil algorithm. In addition, the performance is promoted by increasing the number of anchor nodes. Our method greatly enhances the accuracy and reduces the computational complexity.

ACKNOWLEDGMENT

The research is supported by Chinese National Natural Science Fund (61071140), as well as Jilin Provincial Natural Science Foundations of China (201215014).

REFERENCES

[1] Long Cheng, Cheng-Dong Wu, Yun-Zhou Zhang, "Indoor robot localization based on wireless sensor networks," IEEE Transactions on Consumer Electronics, vol. 57, 3, pp.1099 – 1104, 2011.

[2] Byoung-Suk Choi, Ju-Jang Lee, Sensor network based localization algorithm using fusion sensor-agent for indoor service robot. IEEE Transactions on Consumer Electronics, vol. 56 , 3, pp.1457–1465, 2010.

[3] Yu-Yi Cheng, Yi-Yuan Lin, A new received signal strength based location estimation scheme for wireless sensor network. IEEE Transactions on Consumer Electronics, vol.55, 3, pp.1295 – 1299, 2009.

[4] E. Kim, S. Lee, C. Kim, and K. Kim, “Mobile Beacon-Based 3D-Localization with Multidimensional Scaling in Large Sensor Networks”, IEEE Communication Letters, vol.14, pp.647-649, 2010.

[5] M. T. Isik, and O. B. Akan, “A three dimensional localization algorithm for underwater acoustic sensor networks”, IEEE Trans. Wireless Communications, vol. 8, pp.4457-4463, 2009.

[6] S. Marcos, A. Marsal, and M. Benidir, “The propagator method for source bearing estimation,” Signal Processing, vol. 42, pp.121-138,1995.

[7] J. W. Zhang, C. L.Yu, B.Tang, Y.Y.Ji, Chan Location Algorithm Application in 3-Dimension Space Location, Computing, Communication, Control, and Management (CCCM), 2 (2008), 622-624.

3Gbit/s Transmission over Plastic Optical Fiber with Adaptive Tomlinson-Harashima Precoded Systems

Yixuan Wang
University Stuttgart
Institute of Telecommunications
Stuttgart, Germany +49711-68567925
Email: ywang@inue.uni-stuttgart.de

Julian Müller
University Stuttgart
Institute of Telecommunications
Stuttgart, Germany
Email: mail@julian-mueller.eu

Joachim Speidel
University Stuttgart
Institute of Telecommunications
Stuttgart, Germany +49711-68568017
Email: speidel@inue.uni-stuttgart.de

***Abstract*—This paper presents a 3 Gbit/s transmission scheme for an automobile optical physical layer, which is based on plastic optical fibers (POF) and low cost light emitting diodes (LED). Tomlinson Harashima Precoding (THP) and an adaptive feedforward equalizer (FFE) are used to cope with the strong inter-symbol-interference (ISI). By computer simulations, the system is proven to be a cost-effective and reliable solution for high-speed in-car communications.**

I. Introduction

The up-to-date in-car infotainment backbone MOST150 (the Media Oriented Systems Transport 150) offers a data-rate of 150 Mbit/s in its optical physical layer. The future MOST systems, however, are expected for a 2 $\sim$ 3 Gbit/s transmission capability to serve various automobile applications, like side and back cameras or driver assist applications. Meanwhile, a smooth upgrade from the current MOST150 is required due to cost reasons, i.e., the car-makers can still use plastic optical fibers (POF) and low cost light emitting diodes (LED) in the MOST physical layer [1]. Since the limited bandwidths of POF and LED cause very strong inter-symbol-interferences (ISI) during multi-Gbit/s transmissions, advanced signal processing techniques are necessary for compensating the ISI.

There are several papers on this topic proposing solutions utilizing either an analog prefilter/peaking, or an advanced modulation scheme like discrete multitone modulation (DMT). In [2], a 1.25 Gbit/s transmission with pre-filtered four-level pulse amplitude modulation (4PAM) signaling and fractionally spaced (FS) equalization is presented. However, the prefilter scheme is quite impractical to meet the demand for ever-higher data rates, because the dynamic range of a pre-distorted input signal increases along with the data-rate. So a large DC offset is needed for the LED, which leads to more power consumption and a decrease in the receiver sensitivity. In [3], a 1 Gbit/s transmission is demonstrated using DMT. As is well known, DMT requires high linear transceivers over a wide range of the input optical power, which is difficult for the circuit design. Moreover, the increased cost from using QAM modulator/demodulator and FFT/IFFT components in DMT is a sensitive factor for the car-makers. Until now, most research works targeting at 3 Gbit/s used costly laser diode (LD) or complex modulation schemes. Our work, on the other hand, reports on attempts to achieve the 3 Gbit/s data-rate with a low cost LED and a practical cost-effective system layout. In [4], a 3 Gbit/s transmission is first achieved with a decision feedback equalizer (DFE) and a cheap LED.

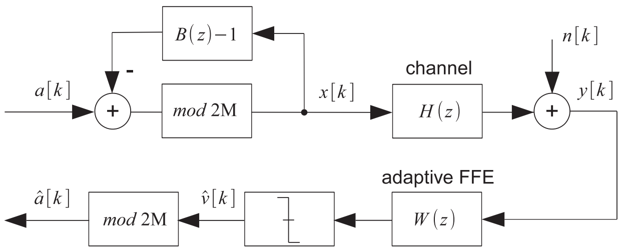


Fig. 1. System block diagram

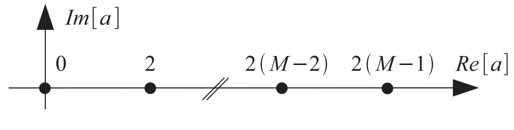


Fig. 2. Modified signal constellation diagram

To reach better performance than [4], while avoiding drawbacks in [2] and [3], this paper combines the Tomlinson Harashima Precoding (THP), which is a nonlinear form of pre-equalization, with a linear feedforward equalizer (FFE) at the receiver, in order to mitigate the error propagation problem of the DFE as well as to restrict the transmit power. The complexity of THP depends linearly on the channel length. The FFE taps are either optimized by the minimum mean squared error (MMSE) criterion or updated by the least mean square (LMS) algorithm. It will be shown that THP-FFE is superior to DFE at 3 Gbit/s, and is robust against variations in a transmission.

II. Overview of the system design

The system block diagram at the symbol level is depicted in Fig. 1. The information bits in the bit level are first encoded by Reed Solomon (RS) coding and then mapped into M-level unipolar pulse amplitude modulation (MPAM) symbols $a[k]$. By taking the bandwidth efficiency, cost and system linearity into account, 8PAM is chosen as the mapping scheme.

$H(z)$ is the discrete form of the electrical equivalent base-band channel of the MOST optical physical layer. The con-

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

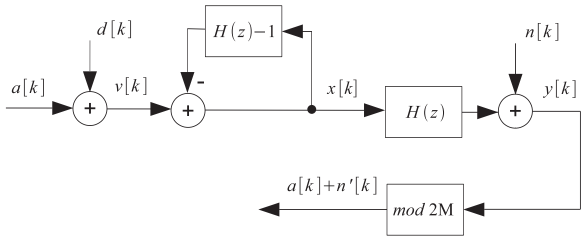


Fig. 3. Linearized THP model

sidered optical layer is a cascade of an LED, a 10 m POF and a photo diode (PD). According to the MOST150 optical physical layer specification [5], its electrical equivalent transfer function can be modeled by a Gaussian low-pass filter:

$$H_a(f) = Ae^{-2(\pi\sigma f)^2} e^{-j2\pi f \tau L_{pof}}; \qquad (1)$$

where L_{pof} is the fiber length, A is the linear fiber loss, $\sigma = \frac{0.132}{B}$ is the standard deviation, $B = 1009 \cdot (\frac{L_{pof}}{m})^{-0.8747}$ MHz is the 3 dB bandwidth, and $\tau = 4.97 \cdot 10^{-9}$ s/m. For a 10 m POF, the 3 dB channel bandwidth is about 134 MHz.

The information symbols $a[k]$ are first encoded by the THP, which consists of a modulo 2M (mod 2M) device and a feedback loop. Here, a change has been made to the classical THP and bipolar MPAM combined scheme. That is, we used the unipolar MPAM and shifted the output range of the mod 2M device from $[-M, M)$ to $[0, 2M)$, as shown in Fig. 2. The reason is that a photo diode is only capable of detecting the power of the received signal, and a loss of phase information is inevitable. Note that this modification does not change the properties of THP.

The mod 2M device is the key element which makes THP outperform the linear prefilter in [2]. Its algorithm can be better explained by a linearized THP model as shown in Fig. 3, where we assume $H(z)$ is a monic channel and no $W(z)$ is required at the receiver. The mod 2M operation is equivalent to adding an unique value $d[k] \in 2M \cdot N$, where M is the order of MPAM and N is an arbitrary integer number, to the symbol $a[k]$ at each time instant k, so that the channel input signal $x[k]$ lies in the interval $[0, 2M)$. By operating non-linearly, the amplitude of $x[k]$ is bounded, thus an amplification of the transmit power in [2] is avoided. To be more precise, THP also introduces certain amplification to the transmit power, but it is small and limited. We will prove this in the later section. Without the presence of noise, $y[k] = v[k] = a[k] + d[k]$ is received after the channel, because the feedback loop perfectly equalizes the channel $H(z)$. So the receiver can simply repeat the mod 2M operation to remove the additive sequence $d[k]$ in order to recover $a[k]$.

To initialize the system, the channel estimation is done within a training block before the data transmission. Based on the channel estimate $\hat{H}(z)$, the feed forward equalizer $W(z)$ is calculated at the receiver by forcing a $B(z) = \hat{H}(z) \cdot W(z) \cdot z^{k_0}$ as close as possible to a causal and monic channel, where k_0 is the index of the main tap. The post-cursors of $B(z)$ are then fed back to the transmitter to determine the coefficients of THP, i.e., the feedback loop of THP is set to $B(z)-1$. Once the THP coefficients are determined, they are fixed. While $W(z)$ serves as the adaptive part in the system, which is constantly updated to handle transmission errors like mismatch of the THP coefficients, channel variation or channel estimation errors. Note that the pre-cursors of $B(z)$ can not be handled by THP, which will severely decrease the performance of THP when they come into play.

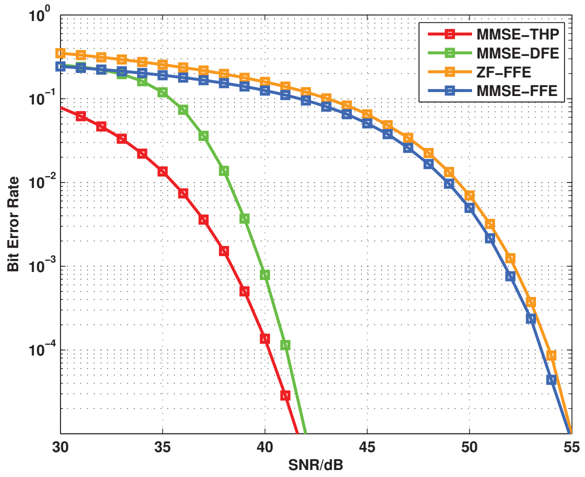


Fig. 4. Comparison of BER performances for THP, DFE, ZF-FFE and MMSE-FFE transmission schemes as a function of SNR at 3Gbit/s

III. Simulation results

The system is evaluated by simulations of bit error rate (BER) as a function of signal-to-noise ratio (SNR). The area-of-interest for SNR is $20.4 \sim 59.5$ dB based on [4] according to the MOST optical physical layer link budget [5]. The area-of-interest for BER is 10^{-4} to 10^{-3}, so the RS coding used in the bit level can decrease the final BER to 10^{-9} and below.

The FFE $W(z)$ at the receiver is fractionally-spaced with twice the symbol rate. Its taps are either optimized in terms of mean squared error for a specific SNR value, or updated by the least mean squares (LMS) algorithm if $W(z)$ is adaptive. Because of the slow fading nature of the optical channel, a small step-size of LMS is sufficient for the variance tracking.

Simulations are first run to identify the improvement of MMSE THP-FFE in comparison to linear FFE and MMSE-DFE, respectively. Then, the robustness of the adaptive THP-FFE as well as its performance under reduced channel bandwidths are investigated. Finally, THP-FFE is compared to an ideal DFE without error propagation.

A. Comparison of THP-FFE, FFE and DFE at 3Gbit/s

Fig. 4 shows a 3 Gbit/s transmission with different system layouts, where taps of THP-FFE and DFE are calculated with the MMSE criterion. Taps of FFE are calculated with zero forcing (ZF) and MMSE criterion, respectively. It can be seen that the THP-FFE outperforms other schemes by reaching the target BER $10^{-4} \sim 10^{-3}$ with a SNR less than 40 dB.

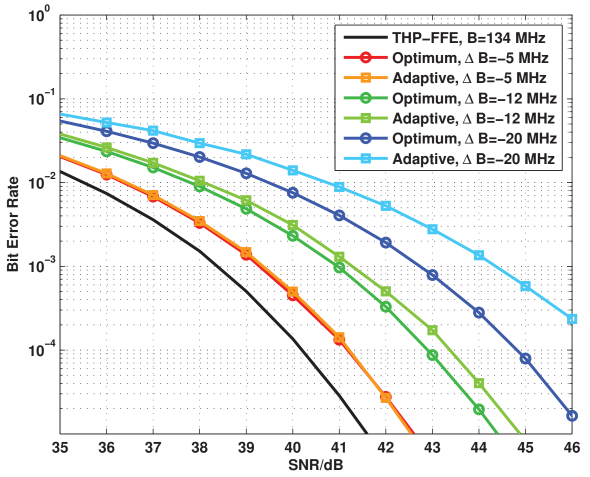


Fig. 5. BER performance of the adaptive THP-FFE transmission scheme as a function of SNR for various bandwidth reductions ΔB

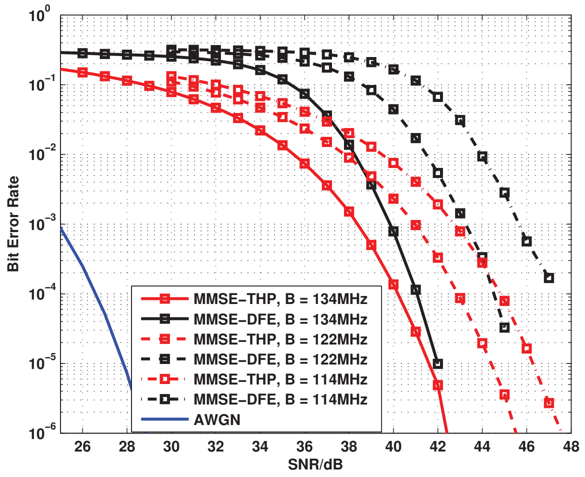


Fig. 6. Comparison of BER performances for MMSE THP-FFE and MMSE DFE transmission schemes with different channel bandwidths B

According to the power budget, a power margin of nearly 20 dB is reserved. Comparing to the DFE solution in [4], THP-FFE presents a performance gain of $1 \sim 2$ dB.

At the same time, a huge improvement of the nonlinear equalizers (THP-FFE and DFE) in comparison to the linear equalizers (ZF-FFE and MMSE-FFE) can be observed. In the BER range of interest, the nonlinear equalizers have 10 dB SNR gain over the linear ones. Because both DFE and THP manage to avoid the large noise enhancement of a linear FFE.

B. Performance of the adaptive THP-FFE under a temperature change

In a practical environment, a channel bandwidth decrease might happen due to the increase of the operating temperature during the transmission. Robustness of the adaptive THP-FFE against such a situation is examined in this section.

According to measurements in [6] under a temperature change in the automobile environment from -40°C to 105°C, the maximum bandwidth decrease is $\Delta B_{max} \approx -20$ MHz. Hence, we test the adaptive system with three different bandwidth drops respectively: a small one $\Delta B = -5$ MHz, a middle one $\Delta B = -12$ MHz, and a big one $\Delta B = -20$ MHz. During the transmission, ΔB is reached gradually.

The results are depicted in Fig. 5. The black curve is the BER performance of MMSE THP-FFE under room temperature with an initial channel bandwidth $B_{\text{-3 dB}} = 134$ MHz. The colored curves demonstrate the performances with bandwidth decreases. For each value of ΔB, two set-ups are compared: an optimum set-up where the coefficients of MMSE THP-FFE are re-calculated once the bandwidth decreases, and an adaptive set-up where LMS algorithm is used at the receiver FFE to follow the channel variation, while the THP coefficients at the transmitter remain fixed.

As expected, performances for both set-ups get worse with the decreasing bandwidth. However, comparing the adaptive set-up with the optimum one, we notice that for a small bandwidth decrease of 5 MHz, the difference between them is negligibly small. In such a situation, a simple adaptive filter at the receiver side is completely sufficient. For a bandwidth drop of 12 MHz, they still have similar performances. Only under the situation that the bandwidth is decreased by 20 MHz, the difference is noticeable. But even so, the adaptive filter performs very stable.

C. Effects of a reduced channel bandwidth

Based on the consideration that an aging degradation in hardware or use of a longer fiber will lead to a permanent decrease of the channel bandwidth, this section investigates the THP performances with a channel bandwidth that is smaller than the theoretical value. Fig. 6 compares the MMSE THP-FFE with MMSE DFE for different channel bandwidths at 3 Gbit/s. It clearly shows that THP-FFE outperform DFE. The reason is that a smaller channel bandwidth produces stronger ISI for a fixed data-rate. In order to combat the ISI, DFE enlarges correspondingly its tap length and values, which increases the probability of an erroneous decision and prolongs the propagation of an error. Contrarily, THP does not suffer from the error propagation, because the encoder uses the actual past symbols instead of the decided past symbols for getting rid of the ISI. Therefore, THP is more advantageous in applications with e.g., a longer fiber, a slower LED or a higher data-rate. Note that the reference BER curve for AWGN channel belongs to the biased MPAM.

D. Discussion of THP-FFE losses

It is well known that THP-FFE suffers from modulo loss and precoding loss, whereas DFE suffers from error propagation. Whether THP-FFE or DFE is more beneficial depends on which loss is dominant. Therefore, we first calculate numerically the THP-FFE losses, and then compare THP-FFE with an ideal DFE without error propagation. The ideal DFE can be considered as the upper bound for THP-FFE, because THP-FFE is somehow equivalent to a DFE by moving the feedback filter in DFE to the transmitter [7], it can never be superior to an ideal DFE without error propagation.

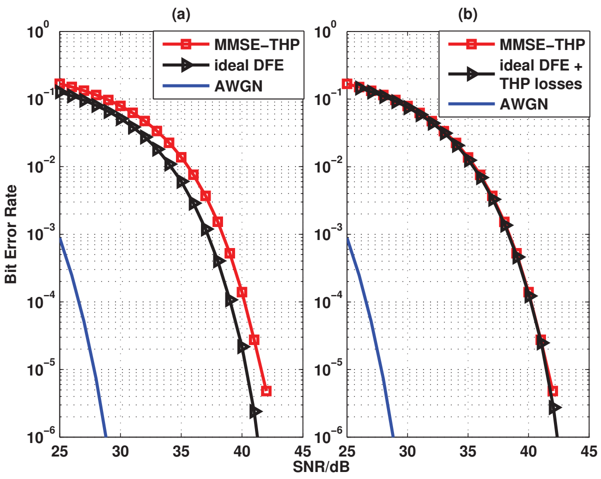


Fig. 7. a) Comparison of a MMSE THP-FFE and an ideal DFE; b) Comparison of a MMSE THP-FFE and an ideal DFE added by THP losses

The losses of THP-FFE are calculated as follows:

- the modulo loss:
  When introducing the congruent signal constellation for THP-FFE, the MPAM signal constellation is periodically repeated, hence the symbol error probability of the edge symbols are now doubled. The increased symbol error rate from $P_{S,MPAM}$ to $P_{S,THP}$ can be calculated as:

$$\gamma_{modulo} = \frac{P_{S,MPAM}}{P_{S,THP}} = \frac{2\frac{M-1}{M}Q(\frac{b}{\sigma_n})}{2Q(\frac{b}{\sigma_n})} = \frac{M-1}{M} \qquad (2)$$

  with b being the distance from a signal point to its neighborhood decision threshold, σ_n being the noise standard deviation, and $Q(x)$ being the Q-function. For $M = 8$, we get $\gamma_{modulo} = 1.1429$. In the logarithmic plot this factor corresponds to a shift of $\log(1.1492) = 0.058$ upwards.
- the precoding loss:
  The precoding loss comes from an increased transmit power of THP-FFE comparing to DFE. However, we use the received SNR as the figure of merit. So the precoding loss at the transmitter has to be transferred to the receiver. Unlike an unbiased (zero mean) transmission, the simple shift of the constellation diagram at the transmitter changes the distribution of the modulo output. Therefore, calculation of the power penalty at the receiver, which is the ratio of the expected received signal power for THP-FFE to the expected received signal power for DEF, has to be done explicitly. For the fractionally spaced system, we have:

$$\begin{aligned} \mathrm{E}[y^2[k]] \quad &= \quad \frac{1}{2}\sigma_s^2 \sum_{i=0}^{P-1} h^2[i] + \frac{1}{2}\mu_s^2 (\sum_{\substack{i=0 \\ i \text{ even}}}^{P-1} \sum_{\substack{j=0 \\ j \text{ even}}}^{P-1} h[i]h[j] \\ &\quad + \sum_{\substack{i=0 \\ i \text{ odd}}}^{P-1} \sum_{\substack{j=0 \\ j \text{ odd}}}^{P-1} h[i]h[j]) \qquad (3) \end{aligned}$$

  with h being the fractionally sampled channel, μ_s and σ_s^2 being the mean value and the variance of the transmit signal $s[k]$, respectively. So the power penalty from the precoding loss is:

$$\gamma_{power-penalty} = \frac{\mathrm{E}[y_{DFE}^2[k]]}{\mathrm{E}[y_{THP}^2[k]]} \qquad (4)$$

  For the biased DFE, $\mu_s = \mathrm{E}[a[k]] = M - 1$, and $\sigma_s^2 = \sigma_a^2 = (M^2 - 1)/3$. For the biased THP-FFE, $\mu_s = \mathrm{E}[x[k]] = M$, and $\sigma_s^2 = \sigma_x^2 = M^2/3$. According to (4), we get $\gamma_{power-penalty} = 1.2744$ for our channel. This power penalty can be regarded as a shift of 1.05 dB to the right on the SNR axis.

When we bring in the calculated losses on the BER plots in Fig. 7, we see: in Fig. 7 (a), an ideal DFE without error propagation performs always better than THP-FFE. However, by adding the two expected losses of THP-FFE to its BER curve, we get almost the curve of THP-FFE, as shown in Fig. 7 (b). Therefore our analysis and calculation of the expected losses are very reliable. When an actual DFE suffers from severe error propagation under strong ISI, THP-FFE has only a certain loss to the ideal DFE. This also well explains the increasing gap between the BER curves of THP-FFE and DFE in Fig. 6 for a decreasing channel bandwidth.

IV. Conclusion

As a conclusion, THP-FFE offers a larger SNR margin than DFE and linear FFE for the sake of data transmission at low error rates. THP-FFE is also more energy-efficient than a pure prefilter. In addition, the adaptive THP system is robust against transmission errors and channel bandwidth reductions caused by temperature variations or hardware degradation. It is a cost-effective, simple and reliable solution for high speed in-car communications based on the MOST optical physical layer.

References

[1] A. Grzemba, Ed., *MOST - The Automotive Multimedia Network - From MOST25 to MOST150*. Franzis, 2011.

[2] F. Breyer, S. Lee, S. Randel, and N. Hani, “PAM-4 signalling for gigabit transmission over standard step-index plastic optical fibre using light emitting diodes,” in *Proc. ECOC*, 2008, pp. 1–2.

[3] S. Lee, F. Breyer, S. Randel, O. Ziemann, H. van den Boom, and A. Koonen, “Low-Cost and Robust 1-Gbit/s Plastic Optical Fiber Link Based on Light-Emitting Diode Technology,” in *Proc. OFC/NFOEC*, 2008, pp. 1–3.

[4] Y. Wang, “Investigation and Simulation of high speed Gigabit Transmission for POF based MOST Networks,” *Elektronik Automotive (Special issue MOST)*, pp. 18–33, Apr. 2011.

[5] *MOST150 oPHY Automotive Physical Layer Sub-Specification, Rev.1.1.* MOST Cooperation, 2010.

[6] J. Krapp, “Dependencies of Bandwidth of Polymer Optical Fiber for MOST Systems,” in *Electronik Automotive, special issue MOST*, 2010.

[7] R. Fischer, *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, 2002. [Online]. Available: http://books.google.de/books?id=jtputBwrGvcC

Variable Block Size Motion Estimation Implementation on Compute Unified Device Architecture (CUDA)

Dong-Kyu Lee and Seoung-Jun Oh
Department of Electronics Engineering at Kwangwoon University

Abstract--**This paper proposes a highly parallel variable block size full search motion estimation algorithm with concurrent parallel reduction (CPR) on graphics processing unit (GPU) using compute unified device architecture (CUDA). This approach minimizes memory access latency by using high-speed on-chip memory of GPU. By applying parallel reductions concurrently depending on the amount of data and the data dependency, the proposed approach increases thread utilization and decreases the number of synchronization points which cause latency. Experimental results show that the proposed approach achieves substantial improvement up to 92 times than the central processing unit (CPU) only counterpart.**

I. INTRODUCTION

Block based motion estimation (ME) is an important part of the video coding standards, such as MPEG-4, H.264/AVC, and the next generation video coding standard, high-efficiency video coding (HEVC). Although ME provides high coding efficiency, it typically requires extensive computation, which makes it difficult to implement a real-time encoder.

With tremendous progress, graphics processing unit (GPU) has been widely used for "General purpose computing on GPUs" (GPGPU). In 2006, a general purpose parallel computing architecture called compute unified device architecture (CUDA) which can easily describe data parallelism and enables GPGPU is introduced.

A few GPU-based ME methods have been proposed [1], [2]. In [1], full search motion estimation algorithm in H.264/AVC was implemented on GPU using CUDA. Although [1] achieves 12 times speed-up in comparison with the central processing unit (CPU) counterpart, it has high memory access latency because the intermediate results are stored not in high-speed on-chip memory of GPU but in device DRAM. Zhou Jing, Jiao Liangbao, and Cao Xuehong presented another approach to take full advantage of on-chip memory of GPU, such as registers and shared memory for memory access latency reduction and achieved improvement up to 50 times more than CPU implementation [2]. However, this approach still has no consideration of the parallel reduction (PR) for the least sum of absolute differences (SAD). A PR inherently has very low thread utilization because the number of active threads is halved between iterations. Furthermore, since data hazard may occur during the PR, a synchronization process is required between iterations, which can cause latency.

In this paper, we propose a highly parallel variable block size full search ME algorithm with concurrent parallel reduction (CPR) on Fermi architecture GPU using CUDA. Specifically, we focus on the variable block size ME in the H.264/AVC standard. The proposed algorithm maximizes the number of active threads and minimizes the number of synchronization points by grouping the available task threads based on the amount of data as well as data dependency. Furthermore, we efficiently use on-chip memory of GPU to minimize memory access latency.

II. PROPOSED APPROACH

The basic unit in the H.264/AVC motion estimation is a 16x16-macroblock (MB). A MB can be split into 16x8-, 8x16-, and 8x8-blocks. The 8x8-block can be further partitioned into 8x4-, 4x8-, and 4x4-blocks. Thus, a MB has total 41 different blocks, each of which has a motion vector (MV). The SAD is used as a matching criterion. As there are several blocks in a MB, hierarchical SAD computing can be used for data reuse. After 4x4-SADs are initially calculated, the 8x4- and the 4x8-SADs are generated by adding two 4x4-SADs. The 8x8-SADs are generated by adding two 4x8-SADs and so on.

Only one kernel is designed to obtain all of the integer-pel motion vectors (IMV) for a current frame which is being encoded. Using just one kernel in our algorithm, storing intermediate results into the device DRAM for the next step is not needed. Before the kernel is launched, the current and a reference frame are transferred to the texture memory of GPU. Since the individual MBs in the current frame can be processed independently, a MB is mapped to a thread block. In our work, the search area (SA) is set to 32x32. Thus, the number of candidates for an IMV is equal to 1024. In order to assign one candidate in the SA to each thread within a thread block, the thread block is set to 32x32. Threads within the thread block calculate SADs and find the least SAD for each of 41 blocks. The indexes of the least SADs become the IMVs.

The proposed algorithm consists of four steps: 1) Calculate 4x4-SADs. 2) Find the least SAD for each 4x4-block using a new concurrent parallel reduction called CPR. 3) Generate SADs for the other blocks using hierarchical SAD computing, and find the least SAD for each corresponding block using CPR. 4) Repeat step 3 until all IMVs are obtained.

A. 4x4-SAD Calculation

Threads within a thread block transfer a MB in the current frame and the 48x48 pixels of an SA in the reference frame from the texture memory to the shared memory which is the on-chip memory of GPU. Then, 1024 SADs for a 4x4-block are computed by 1024 threads. The 1024 calculated SADs are

This work is financially supported by the Ministry of Knowledge Economy(MKE) and Korea Institute for Advancement of Technology(KIAT) through the Workforce Development Program in Strategic Technology and the Research Grant of Kwangwoon University in 2012.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

denoted as a SAD group. Since the number of 4x4-blocks in a MB is 16, 16 SAD groups are sequentially calculated. The 16 SAD groups are stored into the shared memory.

B. *Concurrent Parallel Reduction*

After the 4x4-SAD calculation, we find the least SAD in each SAD group. Since the size of a SA is 32x32, there exists 1024 SADs in a SAD group.

To increase the number of active threads and to decrease the number of synchronization points in PR, we propose the CPR which can handle all of the available SADs, that is, 16 SAD groups in parallel. In CPR, 1024 threads are grouped into 16 thread groups. Thus, each group has 64 threads. Assigning a SAD group to a thread group, 16 PRs can be concurrently executed. In the view of one SAD group, 64 threads within a thread group are available to deal with 1024 SADs. Thus, each thread finds a subminimum SAD among 16 SADs. Since each thread conducts comparisons independently, any synchronization is not required. After the 64 subminimum SADs are found, a PR can be applied. Now, only 32 threads are required for the PR. In the CUDA programming model, 32 threads called a warp are inherently synchronized and execute one common instruction at a time [3]. Thus, we can apply warp unrolling [4], that is, neither a synchronization operation for the next iteration nor a conditional statement for selecting active threads is required. At the end of the CPR, the 16 IMVs for 16 4x4-blocks are obtained and stored into the shared memory.

C. *Hierarchical SAD Computing and Concurrent Parallel Reduction for the Other Blocks*

By the hierarchical SAD computing, 8 8x4-SAD and 8 4x8-SAD groups can be generated. Then, 16 IMVs for 8x4- and 4x8-blocks are obtained by CPR. In the same way, we can generate 9 SAD groups for the other blocks and obtain 9 corresponding IMVs. The total 41 IMVs in the shared memory are finally transferred to the host.

III. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithm on CUDA, the following development environment is used: 8-core CPU with 3.4 GHz with 8 GB memory, Fermi architecture GPU with CUDA capability 2.0 with 1.2 GB DRAM, and CUDA toolkit 4.2. The popular MPEG test sequences, CIF, 4CIF, HD720p, and 1080p, are examined with a 32x32 search area. For CIF and 4CIF videos, the sequences “City”, “Harbour”, and “Soccer” are used. For HD 720p videos, the sequence “Night”, “City”, and “Crew” are used. For 1080p video, the sequence “Rush hour” is used.

The ME algorithm is executed on the host CPU only as well as mainly on GPU. We compare the performance in terms of two cases: one frame and 100 frames. The data transfer time between host and device is considered in the second case whereas not being considered in the first case. The results are shown in tables I and II.

From tables I and II, we can see that our implementation on CUDA demonstrates substantial improvement up to 92 and 82 times than CPU counterpart, respectively. Furthermore, the tables show that the CPR-based algorithm is much faster than the conventional PR-based algorithm. Also, the speed-up will be greater if the application is more computing intensive. For example, speed-up obtained for HD 1080p format is higher than any other format used in this experiment.

IV. CONCLUSIONS

TABLE I
PERFORMANCE COMPARISONS BETWEEN CPU AND GPU (1 FRAME)

Sequence size	CPU (ms)	GPU			
		PR		CPR	
		Time(ms)	Speed-up	Time(ms)	Speed-up
CIF	116.4	5.99	19.4	1.37	85
4CIF	462.2	23.61	19.6	5.37	86.1
HD 720p	1066	53.58	19.9	12.15	87.7
HD 1080p	2512.8	121.08	20.8	27.09	92.8

TABLE II
PERFORMANCE COMPARISONS BETWEEN CPU AND GPU (100 FRAMES)

Sequence size	CPU (ms)	GPU			
		PR		CPR	
		Time(ms)	Speed-up	Time(ms)	Speed-up
CIF	11779.5	633.75	18.6	160.83	73.2
4CIF	46527.3	2434.68	19.1	593.19	78.4
HD 720p	106754.1	5444.65	19.6	1312.27	81.4
HD 1080p	253659.1	12387.43	20.5	3071.49	82.6

In this paper, we proposed the new highly parallel variable block size full search ME algorithm on Fermi architecture GPU using CUDA. We proposed the concurrent parallel reduction called CPR to minimize the inherent problems in conventional PR in terms of thread utilization and synchronization. In our algorithm, hierarchical SAD computing is used to reduce computation time by data reuse. Experimental results show that the proposed approach can achieve up to 92 times and 4 times faster than the traditional CPU and the PR-based CUDA approaches, respectively.

REFERENCES

[1] Wei-Nien Chen and Hsueh-Ming Hang, “H.264/AVC motion estimation implementation on compute unified device architecture (CUDA*)", IEEE International Conference on Multimedia and Expo 2008 (ICME'08)*, pp. 697-700, April 2008.

[2] Zhou Jing, Jiao Liangbao, and Cao Xuehong, “Implementation of parallel full search algorithm for motion estimation on multi-core processors”, *The 2nd International Conference on Next Generation Information Technology (ICNIT)*, pp. 31-35, June 2011.

[3] NVIDIA CUDA, “NVIDIA CUDA C Programming Guide”, v.4.2, pp. 61-62, April 2012.

[4] Mark Harris, “Optimizing parallel reduction in CUDA”, *NVIDIA Developer Technology*, pp. 21-22, 2007.

Ultra-Fast Live Video-in-Video Insertion for H.264/AVC

Dan Grois[1], *Senior Member, IEEE,* Maoz Loants[1], Ofer Hadar[1], *Senior Member, IEEE,*
Rony Ohayon[2] and Noam Amram[3]

**Abstract — ***In this work, an ultra-fast live video-in-video (ViV) insertion scheme is presented, according to which a predefined video content, such as a video advertisement, can be inserted in real-time into a predefined location within a pre-encoded video stream. According to the proposed scheme, the video insertion process is performed in two steps. The first step includes the modification of a conventional H.264/AVC video encoder to support the visual content insertion by using either the Flexible Macroblock Ordering (FMO) technique or by using the Variable Length Coding (VLC)/Variable Length Decoding (VLD). In the second step, the ViV insertion is performed separately for each overlay, while operating in a compressed domain. The presented scheme provides significant improvements in terms of both the bit-rate and insertion run-time: the time period of the proposed ViV insertion process is extremely fast, i.e. up to 8000 times faster compared to JM 17.2 reference software and more than 100 times faster compare to the commercial VSS® (Vanguard Software Solutions®) Streaming Codec Pack 4.5. Also, the bit-rate overhead is very low.*

Index Terms — video-in-video insertion, compressed domain, Flexible Macroblock Ordering (FMO), Variable Length Coding (VLC)/Variable Length Decoding (VLD), H.264/AVC.

I. INTRODUCTION

Recently, the content distribution network industry has become exposed to significant changes. The reduction of cost of digital video cameras along with development of user-generated video sites (e.g., iTunes™, YouTube™) have stimulated a video content sector. As a result, the Video-in-Video (ViV) insertion became a very desirable feature for various future applications, including various TV services for mobile device users (such as the commercial video insertion, subtitling, advertising, and the like). However, traditional approaches failed to provide an efficient solution for supporting a fast real-time ViV insertion of overlays. Thus, for example, [1] presents a method for a logo insertion into a H.264 encoded video stream, while providing a logo as a static image only. Also, [2] present a Picture-in-Picture embedding scheme, while reducing the computational time by an average factor of five (in comparison to full re-encoding) that is still not fast enough for real-time video applications, similarly to [3] and [4], which present the most recent solutions in this field.

In the next section, a detailed description of the proposed scheme is presented, which presents dramatic improvements over authors' previous research [3], followed by experimental results and conclusions.

II. PROPOSED LIVE VIV INSERTION SCHEME

The proposed live ultra-fast ViV insertion system for H.264/AVC, which is presented in *Fig. 1*, contains three main units: *Encoder 1, Encoder 2* and *Inserter*.

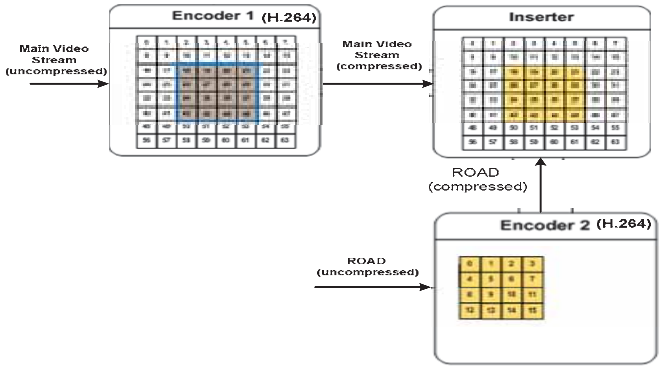


Fig. 1. The proposed live ultra-fast ViV insertion scheme.

Encoder 1 is a H.264\AVC-based modified encoder, which receives a Main video stream as an input and allows terminating spatial and temporal dependencies between the macroblocks at the border between the Main and ROAD (Region of Added Data) video streams - this step is relatively computationally intensive, however, it is performed only once. On the other hand, *Encoder 2* is also a H.264\AVC-based modified encoder, which receives a secondary video stream (e.g., a video advertisement) to be inserted into the Main stream. Further, *Inserter* receives, as an input, two encoded video streams from *Encoder 1* and *Encoder 2*, respectively, and then performs ultra-fast ViV insertion in a compressed domain, without re-encoding the video streams.

A. Live ViV Insertion by Using the FMO

According to one proposed approach, the Flexible Macroblock Ordering (FMO) technique is used for performing the live ViV insertion. FMO is an advanced tool of H.264/AVC that defines the information of slice groups and enables to employ different macroblocks to slice groups of mapping patterns, while each slice within a frame can be encoded independently from its neighbors. In our proposed scheme, each frame is divided onto two slices: one for the Main video stream and another one for the ROAD, while suppressing motion vectors pointing from the Main video stream toward the ROAD (the motion vectors are terminated at a border between the two slices). Also, corresponding parameters of the Main video stream only, such as the Picture Parameter Set (PPS), Sequence Parameter Set (SPS), and slice headers are decoded, followed by insertion of the ROAD FMO slice into the Main video stream.

B. Live ViV Insertion by Using the VLC/VLD

According to another proposed approach, only the entropy coded portions of the encoded Main and ROAD video streams are decoded, and then upon performing the insertion and termination of the spatial and temporal dependencies between the macroblocks, the decoded data is re-encoded again.

This work was supported by the Negev consortium, MAGNET Program of the Israeli Chief Scientist, Israeli Ministry of Trade and Industry under Grant no. 85265610. Also, this work was partially supported by the MEDIEVAL project (Multimedia Transport for Mobile Video Applications) of the 7th Framework Program of the European Commission (Ref.: 258053).

[1] Dr. D. Grois, Maoz Loants and Dr. O. Hadar are with the Communication Systems Engineering Department, Ben-Gurion University of the Negev, Israel (correspondence e-mail: grois@cse.bgu.ac.il, to Dan Grois).

[2] Rony Ohayon is with the Engineering Faculty, Bar-Ilan University, Israel, and [3] Noam Amram is with LiveU® Ltd., Israel.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. Experimental Results

The test conditions are as follows: the spatial resolution of the Main and ROAD video streams are CIF and QCIF, respectively; the frame rate is 30fp/s; Intra-period is 15; Search range is 16; a number of reference frames is 2; the overall number of frames is 180; GOP size is 15; GOP structure is *IPPP*. The tests were carried out on computers with Intel Core 2 Duo CPU, 2.33 GHz, 2GB RAM with Windows® XP, by using the JM 17.2 reference software and VSS® (Vanguard Software Solutions®) Streaming CODEC Pack 4.5. Also, the ROAD stream was inserted in three different locations within the Main stream, i.e. at the top-left corner, at the center, and at the bottom-right corner. *Tables I* and *II* show the overall overhead (compared to the Main video stream) for *PARIS*, *NEWS*, and *MOBILE* video sequences, by using the methods of *Section II.A* and *II.B*, respectively. As seen from these tables, the bit-rate overhead is very low.

TABLE I
The Overall Overhead by Using FMO-based ViV Method

	PARIS		TEMPETE		MOBILE	
Position	**Bit-Rate *(Kb/sec)***	**Bit-Rate Overhead *(%)***	***Bit-Rate (Kb/sec)***	***Bit-Rate Overhead (%)***	**Bit-Rate *(Kb/sec)***	**Bit-Rate Overhead *(%)***
Top-Left	1132.0	1.3	2022.1	0.4	3007.6	1.6
Center	1141.5	2.1	2025.0	0.6	3004.9	1.5
Bottom-Right	1133.1	1.4	2027.3	0.7	2997.4	1.3
Average	1135.5	1.6	2024.8	0.6	3003.3	1.5

TABLE II
The Overall Overhead by Using VLC/VLD-based ViV Method

	PARIS		TEMPETE		MOBILE	
Position	**Bit-Rate *(Kb/sec)***	**Bit-Rate Overhead *(%)***	**Bit-Rate *(Kb/sec)***	**Bit-Rate Overhead *(%)***	**Bit-Rate *(Kb/sec)***	**Bit-Rate Overhead *(%)***
Top-Left	1132.5	1.3	2019.2	0.3	3006.2	1.6
Center	1139.9	2.0	2023.9	0.5	3005.9	1.6
Bottom-Right	1135.1	1.6	2027.4	0.7	2996.9	1.2
Average	1135.8	1.6	2023.5	0.5	3003.0	1.5

In addition, *Tables III* and *IV* present the insertion time of methods, which are proposed in *Section II.A* and *II.B* above (i.e. the live/real-time FMO-based and VLC/VLD-based ViV insertion, respectively). As clearly seen from these tables, the FMO-based ViV insertion provides significant improvements in terms of both the bit-rate and insertion run-time: the time period of the ViV insertion process is extremely fast, i.e. up to *8000 times faster* compared to the JM 17.2 reference software and more than *100 times faster* compare to the VSS® Streaming Codec Pack 4.5. On the other hand, the VLC/VLD-based ViV insertion is more than *200 times faster* compared to the JM 17.2 reference software and more than *4 times faster* compare to the VSS® Streaming Codec Pack 4.5.

IV. Conclusion

In this work, we presented two methods for the ultra-fast live ViV insertion, thereby providing significant improvements in terms of both the bit-rate and insertion run-time: the time period of the proposed ViV insertion process is extremely fast, i.e. up to *8000 times faster* compared to JM 17.2 reference software and more than *100 times faster* compare to the VSS® Streaming Codec Pack 4.5. Also, the bit-rate overhead is very low.

TABLE III
Insertion Time Comparison Between The Proposed FMO-Based ViV Method (*Section II.A*), JM and VSS® Streaming Pack

Input Video Sequences *Main/ROAD*	**Bit-Rate *(K/sec) Main/ ROAD***	**JM Inser tion *(sec)***	**VSS Inser tion *(sec)***	**Proposed Method by Using FMO *(sec)***	**Compared to JM *(Times Faster)***	**Compared to VSS *(Times Faster)***
CREW / COASTGUARD	1468/ 458	185	3.5	0.031	**5966**	**113**
NEWS / GRANDMA	470/ 110	132	2.1	0.016	**8246**	**131**
PARIS / COASTGUARD	1084/ 458	155	2.8	0.032	**4840**	**88**
ICE / GRANDMA	581/ 110	126	3.3	0.031	**4078**	**108**
CREW / FOREMAN	1468/ 261	191	3.5	0.032	**5975**	**109**
NEWS / COASTGUARD	470/ 458	130	2.1	0.031	**4200**	**68**
PARIS / FOREMAN	1084/ 261	145	2.8	0.032	**4544**	**88**
ICE / COASTGUARD	581/ 458	137	3.3	0.031	**4440**	**108**

TABLE IV
Insertion Time Comparison Between The Proposed VLC/VLD - Based ViV Method (*Section II.B*), JM and VSS® Streaming Pack

Input Video Sequences *Main/ROAD*	**Bit-Rate *(K/sec) Main/ ROAD***	**JM Inser tion *(sec)***	**VSS Inser tion *(sec)***	**Proposed Method by Using FMO *(sec)***	**Compared to JM *(Times Faster)***	**Compared to VSS *(Times Faster)***
CREW / COASTGUARD	896/ 317	157	3.5	1.5	108	2.4
NEWS / GRANDMA	497/ 84	124	2.1	0.5	247	4.2
PARIS / COASTGUARD	741/ 317	151	2.8	1.2	123	2.3
ICE / GRANDMA	400/ 84	117	3.3	0.8	149	4.3
CREW / FOREMAN	896/ 193	157	3.5	1.4	113	2.5
NEWS / COASTGUARD	497/ 317	124	2.1	0.8	149	2.5
PARIS / FOREMAN	741/ 193	151	2.8	1.1	136	2.5
ICE / COASTGUARD	400/ 317	117	3.3	1.1	105	3.0

Acknowledgment

We gratefully thank Elina Sklarevich, Yoav Preisler, Shlomi Fraiman, and Yoav Naaman for their assistance in testing an evaluation. We also thank Udi Levy, Michael Degtyar, Alexey Minevich and Evgeny Kaminsky for their contribution to our previous work in this field.

References

[1] T. Tsuji, A. Yoneyama, H. Yanagihara, and Y. Takishima, "High quality logo insertion algorithm for H.264/AVC," *Consumer Electronics, 2008. ICCE 2008. Digest of Technical Papers. International Conference on*, pp.1-2, 9-13 Jan. 2008.

[2] Y. Michalevsky, and T. Shoham, "Fast H.264 Picture in Picture (PIP) transcoder with B-slices and direct mode support," *MELECON 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference*, pp.862-867, 26-28 Apr. 2010.

[3] D. Grois, E. Kaminsky, and O. Hadar, "Efficient real-time Video-in-Video insertion into a pre-encoded video stream," *ISNR Signal Processing*, vol. 2011, Article ID 975462, 11 pages, 2011.

[4] N. Roma, and L. Sousa, "A tutorial overview on the properties of the discrete cosine transform for encoded image and video processing," *Signal Processing*, vol. 91, iss. 11, pp. 2443-2464, Nov. 2011.

Fast Coding Unit Size Decision Algorithm for Intra Coding in HEVC

Jongho Kim[1], Yoonsik Choe[1], and Yong-Goo Kim[2]
[1]School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
[2]Newmedia Department, Korean German Institute of Technology, Seoul, Korea

Abstract—**The recursive quad-tree structure of Coding Unit (CU) in HEVC can improve coding efficiency while the encoding complexity is increased significantly. In this paper, a fast CU size decision algorithm to reduce the complexity of HEVC intra coding is proposed. It introduces the early termination method based on the statistics of rate-distortion costs in CU splitting process. Experimental results show the encoding time can be reduced by 24% on average without significant loss of BD-rate.**

I. INTRODUCTION

The demands on higher resolution or better quality of video beyond high definition (HD) are increasing. The H.264/AVC standard may not provide the best compression performance for such high resolution or high quality video data. To improve compression efficiency, ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG) have jointly developed a new video coding standard, named High Efficiency Video Coding (HEVC) [1][2].

The HEVC is also based on hybrid coding architecture as the H.264/AVC. Compared to the H.264/AVC, the HEVC employs new coding modes, extended block sizes and a flexible hierarchical coding structure such as Coding Unit (CU) and Prediction Unit (PU) [2]. In particular, the CU is the basic unit which can be recursively quad-tree split into sub-CUs unlike the macroblock in H.264/AVC. One CU can be encoded in the best mode achieving the smallest Rate-Distortion (RD) cost among all the possible modes (35 modes for intra coding) or recursively split into four CUs with halved sizes. Split case is selected for the CU if the sum of the RD costs of four sub-CUs is smaller than the RD cost of large CU. This recursive quad-tree splitting process starts from the Largest CU (LCU, 64x64, depth=0) and continues until the Smallest CU (SCU, 8x8, depth=3) is reached. The recursive quad-tree structure of CU can improve coding efficiency while the computational complexity on encoder to decide the optimal CU size is increased significantly.

There have been a few approaches on a fast CU size decision in the HEVC [3][4]. Kim et al. in [3] proposed an early CU termination method based on the adaptive weighted average RD cost of previous skipped CUs. Shen et al. in [4] proposed a fast CU size selection algorithm based on Bayesian decision rule. They provide early termination methods for the specific inter coding modes such as SKIP and INTER_2Nx2N and cannot be directly applied to intra coding with different prediction modes. This paper presents a fast CU size decision algorithm for intra coding in the HEVC which is based on the statistics of RD costs in CU splitting process. By analyzing the statistical distribution of RD costs for each CU in different sizes, we develop the early termination technique to skip further splitting into sub-CUs.

II. PROPOSED ALGORITHM

A. Statistical analysis and motivation

For the optimal CU size decision, the HEVC performs the recursive CU splitting process by evaluating the RD cost. Such an exhaustive search approach is obviously impractical due to its complexity. One can easily perceive that large sized CU is more suitable for a homogeneous area, while a region with edges or object boundaries is usually split into small CUs. The CUs without being split into sub-CUs are expected to have relatively smaller RD costs than the CUs with being split.

To verify this intuition, simulations were conducted by using various test sequences under different QP settings for investigating the RD cost distribution for CUs resulted from the exhaustive search of HEVC Test Mode (HM), and the results are shown in Fig. 1. We can observe that the RD cost distribution for CUs without being split (Non-SPLIT) is highly skewed toward low values, while the RD costs for CUs with being split (SPLIT) are widely distributed to relatively higher values. This observation implies that an early termination chance for unnecessary recursive splitting the CU is given by evaluating whether the resulted RD cost is below a certain threshold to reduce computational complexity effectively.

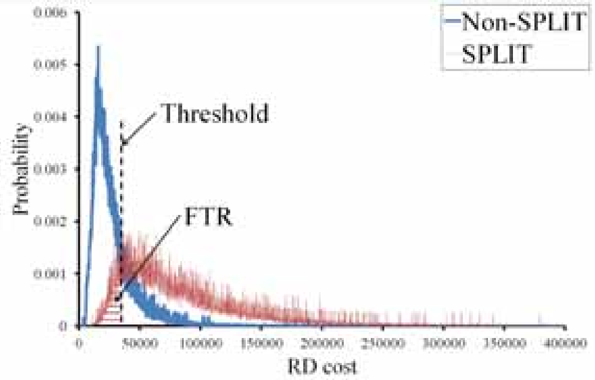


Fig. 1. The RD cost distribution of CUs without being split and CUs with being split for 32x32 CUs in BasketballDrive sequence (QP=37).

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the “ITRC(Information Technology Research Center)” support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2012-H0301-12-1008)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

B. Proposed early termination algorithm

We propose the fast CU size decision algorithm for intra coding to early terminate the recursive CU splitting process before further checking the RD cost with being split into sub-CUs. If the RD cost computed for one CU is small enough, that is, below a pre-set threshold, then the current CU is selected as the best CU with the optimal size, and the encoding process proceeds to the next LCU. This early termination process is applied to each CU with different sizes by using different threshold values.

The threshold values play a critical role to control the trade-off between complexity reduction and RD degradation. To determine the reliable threshold values, we define False Termination Rate (FTR), which means the percentage of CUs being falsely encoded into Non-SPLIT case (Fig. 1), and experiments were conducted with 100 frames from different test sequences, ParkScene, BasketballDrive, and BQTerrace. Fig. 2 shows the results according to varying the FTR. The results indicate that BD-rate loss increases linearly while encoding time saving ratio decreases more rapidly as FTR increases. From these results, we consider the threshold values with satisfying 5% of FTR to reduce complexity without significant RD degradation (BD-rate loss less than 1%).

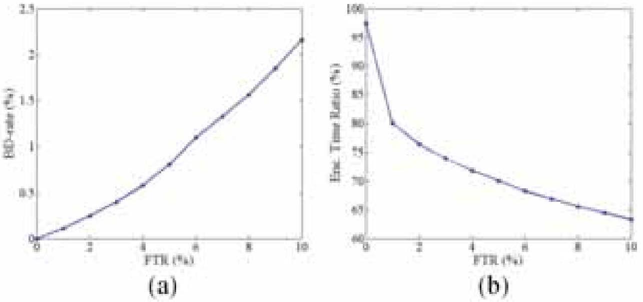


Fig. 2. (a) BD-rate loss and (b) encoding time ratio according to varying FTR.

Since different QPs yield the different statistics of RD cost distribution, the thresholds must vary with QPs. The threshold values for different sizes of CUs versus the various QP values are shown in Fig. 3. Each curve can be fitted with the simple exponential function of QP as the following equations.

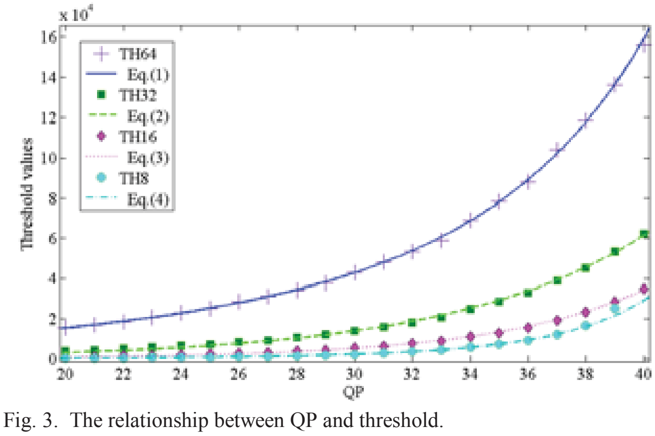


Fig. 3. The relationship between QP and threshold.

$$TH_{64} = 982.7e^{0.126\times QP} \quad (1)$$

$$TH_{32} = 164.6e^{0.148\times QP} \quad (2)$$

$$TH_{16} = 19.75e^{0.187\times QP} \quad (3)$$

$$TH_{8} = 1.054e^{0.254\times QP} \quad (4)$$

Therefore, the proposed algorithm performs the early termination for every CU with different sizes by using the corresponding RD cost thresholds.

III. EXPERIMENTAL RESULTS & CONCLUSION

To evaluate the performance of the proposed algorithm, it was implemented on HM5.2rc1 software [5]. The experimental condition is "Intra Only, High Efficiency" in the common test conditions [6]. The experimental results are shown in Table 1. The proposed algorithm saves 24% encoding time on average up to 37% compared to HM. Its BD-rate [7] increment is only 0.83% on average.

The proposed algorithm can effectively reduce the encoding computational complexity of the HEVC intra coding without significant RD performance degradation by using the simple early termination technique.

Table 1. BD-rate and Encoding time ratio of the proposed algorithm.

Class	Sequences	BD-rate (%)	Enc.Time Ratio (%)
A (2560x1600)	Traffic	1.24	74
	PeopleOnStreet	1.43	77
B (1920x1080)	Kimono	1.7	63
	ParkScene	0.56	76
	Cactus	0.63	76
	BasketballDrive	0.89	68
	BQTerrace	0.23	80
C (832x480)	BasketballDrill	0.63	81
	BQMall	1.10	77
	PartyScene	0.15	87
	RaceHorces	0.60	79
Average		0.83	76

REFERENCES

[1] "Joint call for proposals on video compression technology," ISO/IEC MPEG 91st meeting, N11113, Kyoto, Japan, Jan. 2010.

[2] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, and T. Wiegand, "WD5: Working draft 5 of High-Efficiency Video Coding," JCTVC-G1103, Geneva, CH, Nov. 2011.

[3] J. Kim, S. Jeong, S. Cho, and J. Choi, "Adaptive coding unit early termination algorithm for HEVC," in Consumer Electronics (ICCE), 2012 IEEE International Conference on, pp. 261-262, Jan. 2012.

[4] X. Shen, L. Y, and J. Chen, "Fast coding unit size selection for HEVC based on Bayesian decision rule," in Picture Coding Symposium (PCS) 2012, pp. 453-456, May 2012.

[5] JCT-VC HEVC reference software version 5.2, available online at https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-5.2rc1.

[6] F. Bossen, "Common HM test conditions and software reference configurations," ISO/IEC JTC1/SC29/WG11, JCTVC-G1200, Geneva, CH, Nov. 2011.

[7] G. Bjontegarrd, "Calculation of average PSNR differences between RD curves," in ITU-T SC16/Q6 13th VCEG meeting, No.VCEG-M33, Austin, TX, Apr. 2001.

Using a Wireless LAN to Perform Motion Detection

Cade CASHEN, *Student Member, IEEE,* Samuel RUSS, *Member, IEEE,* and Thomas THOMAS, *Member, IEEE*

***Abstract*—In a wireless home network, multipath signal components are made up of reflections from walls, ceilings, floors, furniture, and other objects. Any motion in the vicinity will change the magnitude of the multipath components arriving at the antenna. This variation in signal strength intensity is statistically significant and does not rely on line-of-sight between the transmitter and receiver. In this project, a set of wireless sensor nodes were developed that are capable of utilizing the RSSI variance between path links to detect motion and control electrical outlets. The result is a stable, reliable room of "smart outlets" that can sense whether or not a room is occupied, and change their state accordingly.**

I. Introduction

Smart grid technology promises to revolutionize power management by enabling the real-time monitoring energy consumption of homes and inside homes. One method for monitoring power consumption is to install wireless sensor nodes that sense power consumption at individual outlets and enable monitoring of the power consumption of individual appliances [3].

The wireless network traffic passing between nodes also passes through and bounces off of objects in the home. As a result, the measured signal strength of each received packet is a function of the objects in proximity, and any motion in the area can have the effect of changing the measured signal strength.

By measuring the variance in received signal strength indication (RSSI) at each node, motion detection becomes possible. This document outlines a system that was implemented and tested to add motion-detection capability to a wireless "smart grid" sensor network. It is important to understand that this motion-detection capability can added to *any* wireless network that has the ability to report the RSSI of each packet received.

II. Survey of Current Technology

A. Wireless network localization

Much work has been done on using wireless networks to locate moving wireless clients. The process of locating is called *localization* and there are numerous methods that have been proposed and used to perform it [5],[6].

This work focuses on a different, more modest application, that of motion detection. The goal in this work is not to locate wireless clients that are moving but simply to sense whether people (e.g.) are moving around inside a space that is monitored by a wireless network. One important distinction is that localization generally involves tracking wireless clients (that is, clients that are active wireless-network nodes) but the motion detection performed in this work is not restricted to wireless nodes but to almost any object moving in the vicinity.

B. Wireless network motion detection

Variance in RSSI has been used to develop a system that can accurately localize the source of motion within a room [1]. The system uses dozens of nodes, numerous packets, and a large, matrix-based tomographic process to determine the source of motion from variance in the measured RSSI. The system demonstrated considerable success, being able to localize motion from objects that were not wireless clients. Similar work has also been carried out by using 802.11 [4].

The approach taken in this work is to combine the RSSI-variance-based approach with a much smaller number of nodes and a much simpler algorithm to achieve a more modest goal: motion detection. The rationale is that the smaller number of nodes is more practical and cost-effective, and the goal (motion detection) is adequate for power management (e.g. determining if a room or dwelling is occupied) and home health care (e.g. determining if a home health patient is ambulatory).

C. Wireless Network Protocols

Numerous wireless network protocols are potentially usable for motion detection. Two popular protocols in low-cost, low-power settings are Zigbee and 802.15.4 [2]. The latter was selected for this work as it is a simpler, lower-level network protocol but essentially any protocol could be used. The only requirement for a usable protocol is that nodes be able to read the received signal strength and source address of each packet.

III. Development of Sensor Node

A. Hardware

An experimental node was developed to test its ability to perform motion detection. The node is composed of a microprocessor board, a wireless-LAN add-on board, and an antenna. The node uses the 802.15.4 IEEE wireless protocol, and operates using an ad-hoc PAN network at a frequency of 2.4 GHz. The network was configured using the software that was provided with the Zigbee add-on board.

A 16x2 RGB LCD display is used to display values for power consumption and signal variance, while also changing backlight color with the intensity of RSSI. The LCD display

The authors are affiliated with the University of South Alabama in Mobile, Alabama, USA.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

provided a real-time indication of RSSI measurement for field testing.

The node is shown below in Figure 1.

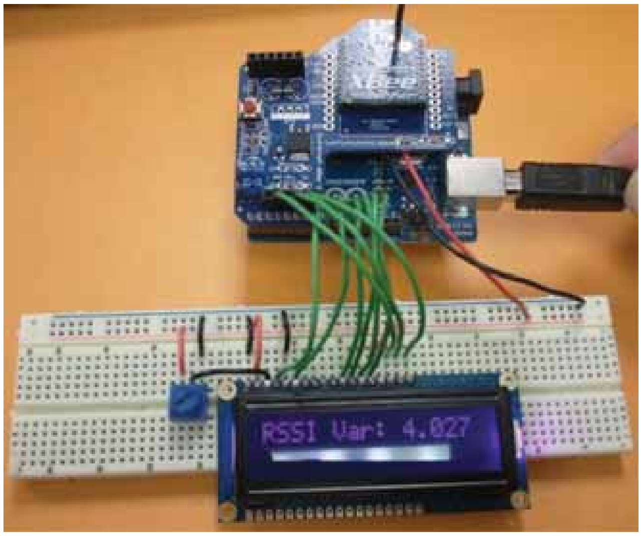


Fig. 1. Prototype Sensor Node. The RSSI variance is displayed on the LCD panel for field-testing purposes.

To emulate more sophisticated capabilities, a relay was added to the design so that, if motion is sensed, the relay could be activated. The relay might be used in an actual home, for example, to keep the lights in the room on when the room is occupied.

Two types of software was written for the node -- one "base station" node, with code to broadcast data or commands to all nodes in the network, and numerous "multipoint" nodes, with code to receive the base station transmissions, measure RSSI, calculate the variance in RSSI, then control the state of the relay and measure power consumption [3].

B. *Algorithm*

The base station node is programmed to send 10 broadcast packets and then wait for one reply from each multipoint node. The multipoint node is programmed to receive 10 packets and record the RSSI of each received packet. Once all 10 packets are received, the variance in RSSI across the 10 packets is calculated and then transmitted back to the base station node.

Additionally, each multipoint node reports the variance on the LCD panel and, if there is motion in the vicinity, turns on a relay. The appropriate variance threshold was determined empirically.

IV. TESTING AND RESULTS

For proof of concept, a single multipoint node was tested with a base station node in different locations and the measured variance in RSSI observed.

In a simple two-node setup in a household living room, approximately 10 feet apart without line-of-sight between the two nodes, significant differences in RSSI variance were observed with just one person between them. If the person sat very still (no head/torso movement), the observed RSSI variance fell to under 0.2. In fact, in many cases, it dropped all the way to zero. With just slight head/torso movement, the RSSI variance climbed to 0.5, reaching values as high as 0.7. If the person stood up and walked across the room, the RSSI variance skyrocketed to values as high as 50. Even sporadic motion while seated induced RSSI variances of 5 or more. Unlike conventional infrared motion detectors, there did not need to be line of sight from the moving objects to the sensor nodes, nor did there need to be line of sight between the sensor nodes.

A similar setup was duplicated in an undergraduate laboratory at the University of South Alabama. With the nodes about six feet apart, motion in the vicinity was detected consistently. Additional testing continues with different geometries and motion vectors.

V. CONCLUSIONS AND FUTURE WORK

RSSI-based motion detection is easy to implement and appears to produce a wide dynamic range for motion measurements. The next steps are to make additional measurements to quantify the thresholds, to run the testbed with a larger number of "multipoint" nodes, and to develop a strategy to minimize the wireless network traffic that is needed to maintain a coherent picture of RSSI variance as the number of nodes increases. Beyond that, the ability to localize motion and generate motion vectors, as in [1], will be explored.

The addition of motion detection to an existing wireless network is a powerful capability. Many institutional and home settings, such as daycare, fire and rescue, and elderly living alone, can benefit from motion-detection capability, and, because it leverages a wireless network already in place, the incremental cost is minimal.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of the Department of Electrical and Computer Engineering.

REFERENCES

[1] Wilson, J.; Patwari, N.; , "See-Through Walls: Motion Tracking Using Variance-Based Radio Tomography Networks," *IEEE Transactions on Mobile Computing,* , Vol.10, No.5, pp.612-621, May 2011.

[2] Gascon, D., "802.15.4 vs. ZigBee", Wireless Sensor Networks Research Group White Paper, Published Nov. 17, 2008, Available via http://sensor-networks.org/index.php?page=0823123150.

[3] S. Russ, T. Thomas, and C. Cashen, "Leveraging Smart Grid Technology for Home Health Care," Accepted to 2013 ICCE.

[4] Moustafa Youssef, Matthew Mah, and Ashok Agrawala. 2007. "Challenges: device-free passive localization for wireless environments", *Proceedings of the 13th annual ACM international conference on Mobile computing and networking* (MobiCom '07). ACM, New York, NY, USA, pp. 222-229.

[5] Hui Liu; Darabi, H.; Banerjee, P.; Jing Liu; , "Survey of Wireless Indoor Positioning Techniques and Systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* , vol.37, no.6, pp.1067-1080, Nov. 2007.

[6] Guvenc, I.; Chia-Chin Chong; , "A Survey on TOA Based Wireless Localization and NLOS Mitigation Techniques," *Communications Surveys & Tutorials, IEEE* , vol.11, no.3, pp.107-124, 3rd Quarter 2009

Human Mobile-Device Interaction on HEVC and H.264 Subjective Evaluation for Video Use in Mobile Environment

Ray Garcia, *Member*, IEEE, Hari Kalva, *Senior Member*, IEEE
Florida Atlantic University, Boca Raton, Florida, United States

***Abstract*-- High Efficiency Video Coding (HEVC) is the next coding standard that is being finalized by ITU's Joint Collaborative Team on Video Coding (JCT-VC). This paper compares the H.264/AVC, current coding standard, and HEVC (aka H.265) in mobile compute environments. In this study, the focus, within the mobile compute environment, are smart phones. The major smart phone elements are smaller screen size, which is typically 3.5 inches diagonal to 5.0 inches diagonal for high end smart phones and typical cellular network bandwidth, which is typically 3G or faster. There is compelling subjective test feedback that indicates human device interaction focused on user's experience is very similar between HEVC and H.264 encoding standards for mobile environment screen size and higher cellphone bandwidth (such as 400kbps constant bit rate). This suggests the benefits of HEVC over H.264 in mobile environment are not as clear.**

I. INTRODUCTION

The mobile compute environment has evolved rapidly in the last few years and smart phones have penetrated the consumer market extensively. Display technologies, in the mobile environment market space, have benefited from strong design investment by smart phone manufacturers and significant research and development by liquid crystal display (LCD) manufacturers. This has enabled the mobile LCDs to improve steadily in performance, such as: (a) resolution, (b) power consumption, (c) viewing angles, and other aspects.

The evolution of encoding methods from H.264 to HEVC targets mainly larger screen displays, which are described as 10.4 inch and above, with the main beneficiary being the broadcast industry. However, HEVC is expected to provide compression gains over H.264 in the mobile environment. However, the significance of these gains in mobile devices has to be evaluated to determine whether the additional gains are perceivable by end users on mobile device displays. This study will provide guidance on user experience for the targeted bitrate.

The economic incentive for mobile industry to adopt HEVC may not be a strong motivator when considering user experience with encoding method for a given bitrate. This will allow this market segment to leverage existing H.264 infrastructure and allow HEVC standard to mature some more before adopting. This will reduce early adoption risk with little risk to reduced consumer experience.

II. METHODS

Mobile encoding method recommendations, within this study, revolved around recommendations from Brightcove and Apple that are widely adopted by content providers. Smart phone display performance has progressed significantly and cellular network bandwidth has improved in recent years. These events have led to the use of higher quality resolutions and bitrates for mobile environments. The higher resolution and bitrate will eventually be common for consumer-grade mobile products in the near future. For Brightcove, "Higher quality and resolution" [1] which lists resolution for 640x360 (16:9) at 400kbps. For Apple HTTP Live Streaming [2] recommendations are listed in Table 2-1 (Encoder settings for iPhone, iPod touch, iPad, and Apple TV, 16:9 aspect ratio) were used for test direction. A combination of WiFi resolution (640x360) and higher bandwidth cellular (400 kbps) were selected. Basically, the display technology progress strongly dictated the WiFi resolution (640x360) over the high end cellular resolution (480x224), as currently observed by popular mobile phones with resolutions up to 1280x720 for 5.0" diagonal screen sizes.

The encoder implementations used for the comparison are H.264/AVC Software Coordination version: JM18.3 [3] and High Efficiency Video Coding (HEVC) version HM 6.0[4]. H.264 was configured to closely mimic HEVC coding (based on HM-like configurations available in JM 18.3). The benefit is to reduce the variability between the two video encoding protocols.

Video test sequences chosen are the 30fps HEVC test sequences used in HEVC standards development. The video test sequence files are scaled and cropped to 640x360 resolutions. A total of 5 sequences are in the experiment pool. These are: (a) Flowervase, (b) Keiba, (c) People on Street, (d) Race Horses, and (e) Traffic.

III. EXPERIMENTS

The 640x360 video test sequences were encoded for various quality levels. For H.264, the QP encoded was for QP 27-51. For HEVC, the QP encoded was for QP24-51. For objective comparison the PSNR(Y) was chosen.

Video sequences were shown on a 4.3" LCD with 480x272 resolution. The observer is approximately 12" to 18" from the display and viewing angle is approximately 10 degree above normal as shown in figure 1.

Seven observers participated in the test. The observers are 18 to 50 years of age. All observers are in good health. Corrective lenses were used during the test, if the lenses were prescribed.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

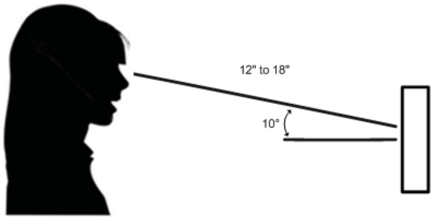


Fig. 1. Observer to LCD viewing definition

For subjective comparison, the H.264 and HEVC quality sequence closest to 400kbps bitrate was chosen. The impaired sequences test methods were configured as described by Oelbaum[5]. Comparisons between both impaired video sequences are in accordance with Double-stimulus impairment scale (DSIS) as defined by ITU-R BT.500-13[6]. Variant II was used for presentation structure of test material. This allows the user two viewings of each video sequence (reference and impaired) before subjective grading. Also, video sequence was shown randomly in order to reduce observer bias. Grade scores are on a scale from 1 to 5. Grade of "one" is poor (very annoying) and "five" is excellent (imperceptible) as defined by ITU-R BT.500-13.

IV. RESULTS

"Keiba" test sequence results yielded PSNR(Y) is about 1.5dB better than H.264. However, the subject results indicate subjective performance is about the same. Actually, H.264 was shown to perform better by one observer. The next two figures show test results on "Keiba" test sequence.

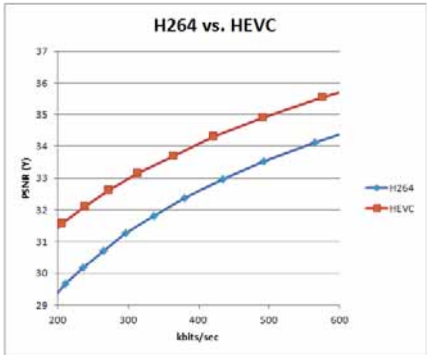


Fig. 2. Keiba PSNR(Y) data. HEVC performed approximately 1.5dB better than H264.

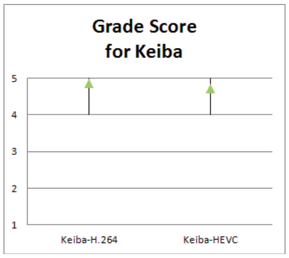


Fig. 3. Keiba subjective data. H.264 performed, as well as, HEVC performance. Chart shows minimum to maximum grade range and average.

"People On Street" test sequence yielded PSNR(Y) is about 0.5dB better than H.264. The subject results indicate subjective performance is about the same.

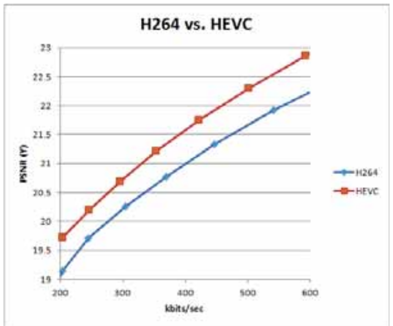


Fig. 4. PeopleOnStreet PSNR(Y) data. HEVC performed approximately 0.5dB better over H264.

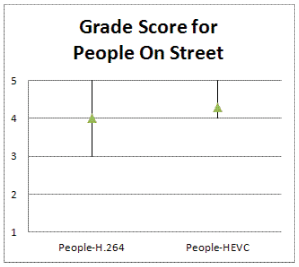


Fig. 5. PeopleOnStreet subjective data. H.264 performed, as well as, HEVC performance. Chart shows minimum to maximum grade range and average.

Subjective results showed 90% grade scores ratings of "5" or "4", which indicates the observer feedback is "imperceptible" or "perceptible, but not annoying", respectively. The observer feedback is the impaired video sequence is acceptable for the mobile (i.e. smaller screen) environment.

V. CONCLUSION

For the mobile conditions performed in the study, which is 4.3" screen size and 400kbps constant bit rate, the user (i.e. observer) experience is not severely affected when comparing H.264 and HEVC impaired video sequences. Observations show both encoding methods are adequate in the mobile environment.

VI. REFERENCES

[1] Encoding for Mobile Delivery. Brightcove. Retrieved 2012July08 from http://support.brightcove.com/en/docs/encoding-mobile-delivery

[2] iOS Developer Library – Preparing Media for Delivery to iOS-Based Devices. Apple. Retrieved 2012July08 from https://developer.apple.com/library/ios/#documentation/NetworkingInternet/Conceptual/StreamingMediaGuide/UsingHTTPLiveStreaming/UsingHTTPLiveStreaming.html#//apple_ref/doc/uid/TP40008332-CH102-SW8

[3] H.264/AVC Software Coordination Version: JM18.3. Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz Institut. Retrieved 2012May09 from http://iphome.hhi.de/suehring/tml/

[4] High Efficiency Video Coding (HEVC) version 6.0. Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz Institut. Retrieved 2012May06 from http://hevc.hhi.fraunhofer.de/

[5] T. Oelbaum, V. Baroncini, T. K. Tan, and C. Fenimore, "Subjective Quality Assessment of the Emerging AVC/H.264 Video Coding Standard"

[6] ITU, "Methodology for the subjective assessment of the quality of television pictures", ITU-R BT.500-13, 2012Jan.

Block Boundary Filtering for Intra Prediction Samples

Akira Minezawa, Kazuo Sugimoto and Shun-ichi Sekiguchi
Information Technology R&D Center, Mitsubishi Electric Corporation, Kamakura, JAPAN

***Abstract*—This paper presents a new intra prediction technique for HEVC, which is next-generation video coding standard that has collaboratively been developed by ISO/MPEG and ITU-T/VCEG. On intra coding for HEVC, extension of MPEG4-AVC/H.264-like intra prediction scheme has been studied. In this conventional scheme, the prediction block is generated by simply copying adjacent reference samples, interpolating the reference samples toward prediction direction or averaging the reference samples. Thus, signal discontinuity between prediction block and adjacent reconstructed blocks could occur, which causes loss of coding efficiency of residual signal. In this paper, we propose an adaptive filtering scheme for intra prediction samples generated by either DC or directional prediction that solves the above issue. According to our evaluation, the proposed scheme achieves approximately 0.6% bitrate savings on average compared to the conventional AVC-style intra coding.**

I. INTRODUCTION

Recently, a next-generation video coding scheme called HEVC(High Efficiency Video Coding) has been developed by JCT-VC, which is a collaborative working group of ISO/MPEG and ITU-T/VCEG. HEVC is aiming twice compression performance compared with MPEG-4 AVC/H.264[1] and a lot of new coding tools have been studied for HEVC on top of AVC-like hybrid coding structure. The technical advances relative to AVC are as follows:

- Employment of extended coding block size
- Inter prediction based on multi-layer hierarchical partitioning with various partition shapes
- Merging prediction blocks sharing unique motion parameter
- Adaptive derivation of motion vector predictors
- Multi-directional intra prediction
- Adaptive transform with multiple block sizes
- In-loop adaptive de-blocking and de-noising filters

In the multi-directional intra prediction as adopted in AVC, the prediction block is generated by simply copying adjacent reference samples, interpolating the reference samples toward prediction direction or averaging the reference samples. Thus, signal discontinuity between prediction block and adjacent reconstructed blocks could occur, which causes loss of coding efficiency of intra residual signal. To overcome this problem, we propose a new adaptive filtering scheme for intra prediction samples generated by DC or directional prediction to improve the accuracy of prediction.

This paper is organized as follows. The background of the proposed scheme is described in Section 2. In Section 3, the proposed technical elements are described. The coding performance of the proposed scheme on top of HEVC Test Model (HM) is discussed in Section 4 followed by the conclusion in Section 5.

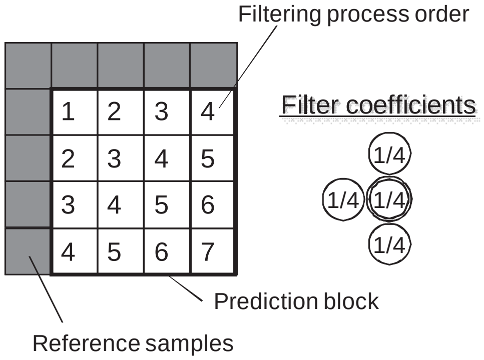


Fig. 1. MPI for 4x4 intra prediction block

II. PRIOR ARTS

To solve the problem of AVC-like intra prediction which can be posed signal discontinuity between prediction block and adjacent ones, several filtering schemes have been proposed [2][3]. In [2], a recursive filtering process, named as MPI (Multi parameter intra), using predicted samples generated by the intra prediction and filtered samples is conducted as shown in Fig. 1. This scheme has a drawback that its filtering process cannot be performed in parallel. In [3], a filtering that utilizes property of DCT is proposed. This scheme requires a memory for reconstructed samples which has three times the size of memory for prediction block and its filtering process relatively complex. In addition, both the filtering schemes require a selection whether the filtering process performs or not for each prediction block and this selection information must be encoded as coding parameter. Given these issues, we developed a simpler smoothing filter to be applied to block boundary of intra prediction samples, while having implicit adaptivity based on intra prediction mode.

III. PROPOSED FILTERING SCHEME

This section provides description of the proposed filtering scheme that improves intra prediction efficiency with a simple solution. The proposed scheme applies simple smoothing filter to block boundary between prediction samples, those are generated by simply applying directional intra prediction, and spatially adjacent reconstructed samples (i.e., reference samples). The smoothing filter is designed to be adaptive depending on prediction block size and intra prediction modes covering up to 34 directions per each prediction block.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

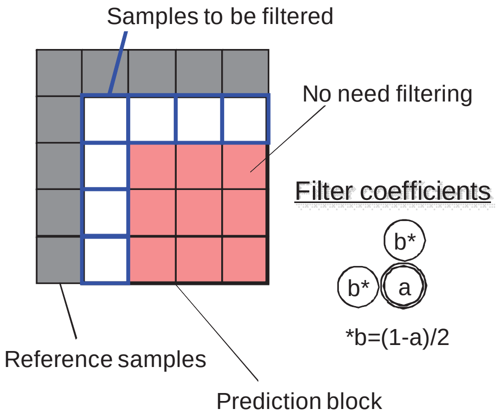


Fig. 2. DC prediction filtering for 4x4 intra prediction block

TABLE I
PROPOSED FILTER COEFFICIENTS

Prediction block size	Value of “a”
4x4	1/4
8x8	1/2
16x16	3/4
32x32	1(Not filtering)

A. *For DC prediction*

In the case of DC prediction, a simple linear filtering that can be performed in parallel is proposed. This filtering process uses 3-tap linear filter, shown in Fig. 2, using predicted samples generated by the intra prediction and adjacent reconstructed samples. Since the proposed filtering can be independently performed for each sample and all samples in DC prediction block are set to an averaging value of adjacent reference samples as a prediction value, samples to be filtered can be limited to block boundary samples as depicted in Fig. 2 with identical output in case that all samples in the prediction block are filtered. As for the filter coefficients, we employed several fixed set of coefficients depending on prediction block size shown in Table 1 determined through offline training.

B. *For directional prediction*

For directional predictions, prediction block is generated by copying adjacent reference samples or interpolating the reference samples toward prediction direction. From a viewpoint that simpler directional prediction modes would produce more signal discontinuity around block edge, we design different filtering methods for vertical and horizontal prediction modes and the other 31 directional modes, respectively. It is noted that the proposed scheme does not perform any filtering for planar prediction mode.

To reduce the block boundary discontinuity for vertical and horizontal prediction modes as shown in Fig. 3, we propose a gradient filter approach depicted in Fig. 4. Equations of the proposed scheme are described as follows:

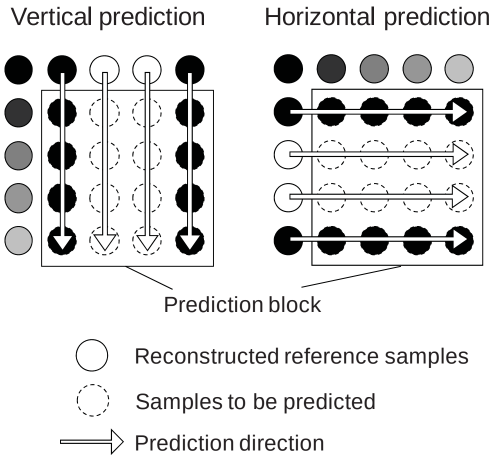


Fig. 3. An example of occurrence of block boundary discontinuity

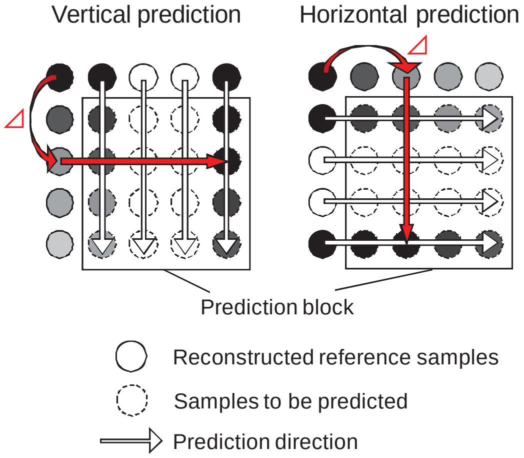


Fig. 4. The proposed gradient filter approach

- Vertical prediction:

$$Pred(x, y) = Rec(x,-1) + (Rec(-1, y) - Rec(-1,-1)) \cdot u(x) \quad (x, y = 0,1,\dots, N-1) \tag{1}$$

- Horizontal prediction:

$$Pred(x, y) = Rec(-1, y) + (Rec(x,-1) - Rec(-1,-1)) \cdot v(y) \quad (x, y = 0,1,\dots, N-1) \tag{2}$$

where (*x*, *y*) denotes the pixel position in prediction block, *Pred*(*x*, *y*) is prediction sample located in (*x*, *y*), *Rec*(*x*, *y*) is reconstructed sample located in (*x*, *y*), *u*(*x*) and *v*(*y*) are scaling parameter for gradient component and *N* indicates width or height of prediction block.

In the case of non-vertical/horizontal directional predictions, we introduce 2-tap smoothing filter for 10 directional predictions around 45 degree shown in Fig.5. Its filter coefficients are set to (a_0, a_1) = (3/4, 1/4), where a_0 indicates a coefficient of the sample to be filtered and a_1 indicates a coefficient of the reference sample, which are determined based on offline simulations.

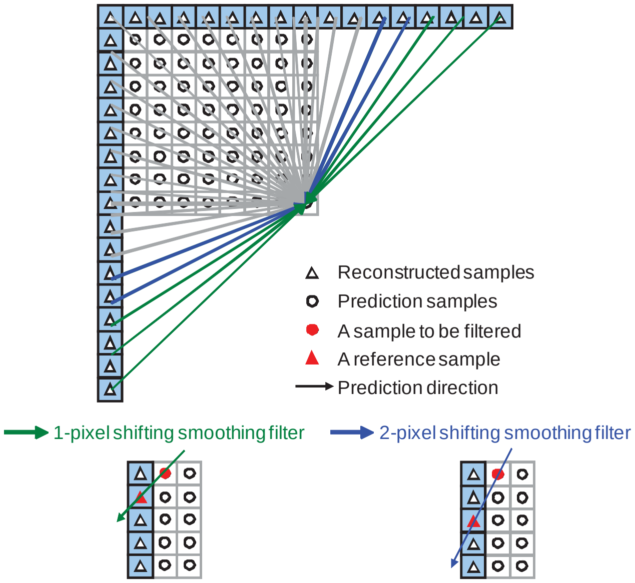


Fig. 5. Proposed smoothing filter for directional predictions

IV. EXPERIMENTAL RESULTS

This section evaluates coding performance of the proposed filtering schemes described in Section 3. First of all, we investigate appropriate function $u(x)$ and $v(y)$ using the gradient filter based prediction defined by equation (1) and (2) for vertical and horizontal prediction modes. Methods to be evaluated are specified as follows:

- Method1:

$$u(x) = v(y) = 1 \quad (3)$$

- Method2:

$$u(x) = 1/2^{x+1} \quad (4)$$

$$v(y) = 1/2^{y+1} \quad (5)$$

- Method3:

$$u(x) = \begin{cases} 1/2^{x+1}\ (N \le 16) \\ 0\ (otherwise) \end{cases} \quad (6)$$

$$v(y) = \begin{cases} 1/2^{y+1}\ (N \le 16) \\ 0\ (otherwise) \end{cases} \quad (7)$$

- Method4:

$$u(x) = \begin{cases} 1/2\ (x = 0\ and\ N \le 16) \\ 0\ (otherwise) \end{cases} \quad (8)$$

$$v(y) = \begin{cases} 1/2\ (y = 0\ and\ N \le 16) \\ 0\ (otherwise) \end{cases} \quad (9)$$

In this investigation, we implemented these methods on top of HM software that is reference implementation of HEVC draft 3.0[4] (HM-3.0) and conducted coding experiments using JCT-VC test sequences with various resolutions listed in Table 2. Experimental conditions are shown in Table 3 in accordance with JCT-VC common test conditions of all intra / high efficiency configuration case[5]. As a metric to assess coding efficiency, BD-bitrate[6] was used. Coding results of these four methods compared with HM-3.0 are shown in Table 4. It is noted that HM-3.0 software used for obtaining the result of Table 4 includes the DC prediction filter proposed in Section 3.A. From the results on this table, it can be seen that method 2 achieves coding improvement compared with the anchor, although the degradation of coding performance occurred in the case of method 1. Then, method 3 has better coding gain compared with method 2 due to restriction on size of the block to be filtered. Moreover, method 4 achieved almost the same coding performance against method 3 although samples to be filtered are only limited to 1 line / column in method 4. Through these investigation, $u(x)$ and $v(y)$ are set to 1/2 where $Pred(1, y)$ and $Pred(x, 1)$ for vertical and horizontal prediction up to 16x16 prediction block, respectively. For the other coordinates (x, y), $u(x)$ and $v(y)$ are set to 0.

TABLE II
TEST SEQUENCES

Class	Picture size [pels]	Bit depth [bits/pel]	Frame rate [fps]	Sequence name
A	2560 x1600	10	60	NebutaFestival, SteamLocomotiveTrain
		8	30	Traffic, PeopleOnStreet
B	1920 x1080		24	Kimono, ParkScene
			50	Cactus, BasketballDrive
			60	BQTerrace
C	832 x480		30	RaceHorses
			50	BasketballDrill, PartyScene
			60	BQMall
D	416 x240		30	RaceHorses
			50	BasketballPass, BlowingBubbles
			60	BQSquare
E	1280 x720		60	Vidyo1, Vidyo3, Vidyo4

TABLE III
EXPERIMENTAL CONDITIONS

Coding structure	IIIII…(All Intra)
Entropy coding	CABAC
Coding block size	64x64~8x8
Transform block size	32x32~4x4
The number of Intra prediction mode	33 directional modes, DC mode, Planar mode
RD optimization	ON
QP	22, 27, 32, 37

To evaluate overall coding performance of the proposed filtering scheme for all intra prediction modes, we performed coding experiments using HM-3.0 for JCT-VC test sequences in Table 2 and experimental conditions in Table 3. Coding

TABLE IV
COMPARISON OF SEVERAL METHODS

Sequence	BD-bitrate (Y signal) [%]			
	Method1	Method2	Method3	Method4
Traffic	0.55	-0.75	-0.77	-0.60
PeopleOnStreet	0.18	-0.74	-0.74	-0.51
NebutaFestival	0.01	-0.07	-0.07	-0.05
SteamLocomotive	0.08	0.06	-0.07	-0.06
Kimono	0.05	-0.10	-0.18	-0.08
ParkScene	0.17	-0.34	-0.37	-0.31
Cactus	0.27	-0.39	-0.42	-0.39
BasketballDrive	0.94	-0.06	-0.27	-0.16
BQTerrace	0.41	-0.22	-0.24	-0.23
Vidyo1	0.14	-0.24	-0.39	-0.33
Vidyo3	1.22	-0.08	-0.31	-0.33
Vidyo4	0.85	-0.28	-0.43	-0.32
BasketballDrill	0.09	-0.10	-0.10	-0.13
BQMall	1.22	-0.61	-0.67	-0.58
PartyScene	1.43	-0.27	-0.28	-0.32
RaceHorses	0.06	-0.13	-0.12	-0.14
BasketballPass	0.21	-0.38	-0.46	-0.33
BQSquare	0.12	-0.27	-0.27	-0.35
BlowingBubbles	0.62	-0.14	-0.14	-0.19
RaceHorses	0.10	-0.19	-0.19	-0.25
2560x1600	0.21	-0.38	-0.41	-0.31
1920x1080	0.37	-0.22	-0.30	-0.23
1280x720	0.74	-0.20	-0.38	-0.33
832x480	0.70	-0.28	-0.29	-0.29
416x240	0.26	-0.24	-0.26	-0.28
Average	***0.44***	***-0.26***	***-0.32***	***-0.28***

TABLE V
COMPARISON OF CODING PERFORMANCE

Sequence	Size [pel]	BD-bitrate[%]		
		Y	U	V
Traffic	2560 x1600	-1.12	-0.89	-0.79
PeopleOnStreet		-1.14	-0.92	-0.90
NebutaFestival		-0.18	-0.10	-0.29
SteamLocomotive		-0.20	0.07	-0.20
Kimono	1920 x1080	-0.26	-0.02	-0.07
ParkScene		-0.81	-0.34	-0.06
Cactus		-0.76	-0.46	-0.47
BasketballDrive		-0.32	-0.40	-0.43
BQTerrace		-0.57	-0.69	-0.54
Vidyo1	1280 x720	-0.66	-0.53	-0.73
Vidyo3		-0.48	-0.68	-0.55
Vidyo4		-0.56	-0.64	-0.69
BasketballDrill	832 x480	-0.37	-0.49	-0.49
BQMall		-0.81	-0.91	-0.88
PartyScene		-0.56	-0.48	-0.49
RaceHorses		-0.54	-0.38	-0.38
BasketballPass	416 x240	-0.60	-0.63	-0.66
BQSquare		-0.49	-0.34	-0.38
BlowingBubbles		-0.51	-0.44	-0.45
RaceHorses		-0.61	-0.62	-0.59
2560x1600		-0.66	-0.46	-0.54
1920x1080		-0.54	-0.38	-0.31
1280x720		-0.57	-0.62	-0.66
832x480		-0.57	-0.56	-0.56
416x240		-0.55	-0.51	-0.52
Average		***-0.58***	***-0.49***	***-0.50***

performance compared with the anchor configuration that disables all proposed filtering is shown in Table 5. According to the results, it can be seen that the proposed scheme achieves around 0.6% consistent bitrate reduction for luma component. In terms of computational complexity, Table 6 shows run time comparison measured on the same CPU. This table indicates run time of proposed scheme is almost the same as that of anchor that disables proposed filtering since the processing impact is small due to limitation of filtered area only to block boundary. It is also noted that the proposed filter does not require additional memory for storing samples other than those used for original intra prediction process.

V. CONCLUSION

In this paper, new filtering schemes depending on intra prediction modes to reduce prediction mismatch around bock boundary have been proposed. According to the simulation results, the proposed scheme achieves consistent bitrate savings over wide range of picture resolutions, compared to the conventional intra prediction design.

TABLE VI
COMPARISON OF EXECUTION TIME

	CPU run time[sec]		Rates of increase[%]
	HM-3.0 w/o proposed filter	HM-3.0 with proposed filter	
Encoding	7488.46	7544.23	0.74
Decoding	65.80	65.78	-0.03

REFERENCES

[1] ISO/IEC 14496-10 | ITU-T Recommendation H.264, March 2005.

[2] K. McCann, et al., “Samsung’s Response to the Call for Proposals on Video Compression Technology,” JCTVC-A124, April 2010.

[3] S. Sekiguchi, et al., “DCT-based noise filtering for intra prediction samples,” JCTVC-B067, July 2010.

[4] T. Wiegand, et al., “WD3: Working Draft 3 of High-Efficiency Video Coding,” JCTVC-E603, March 2011.

[5] F. Bossen, “Common test conditions and software reference configurations,” JCTVC-E700, March 2011.

[6] G. Bjontegaard, “Calculation of average PSNR differences between RD-Curves,” VCEG-M33, April 2001.

Utilization Analysis of Trim-Enabled NAND Flash Memory

Boncheol GU, Jupyung LEE, Brian Myungjune JUNG, Jungmin SEO, and Hyun-Jung Shin
Samsung Advanced Institute of Technology, Yong-in, South Korea

Abstract—**In this paper, we present a novel probabilistic model of the utilization for trim-enabled NAND flash memory devices. This model provides a simple and powerful method to reason about the performance of NAND flash memory, given its capacity and the frequency of write operations.**

I. INTRODUCTION

NAND flash memory has become a standard medium for consumer electronics because it has many appealing features such as small size, low read latencies, low power consumption, shock resistance, and lightweight. However, NAND flash memory has some limitations which pose challenges to the flash memory system. After data is written into a page, the elementary unit for reads and writes, new data cannot be overwritten immediately. The page needs to be erased before it is written. To address this erase-before-write, it is necessary to perform out-of-place update and garbage collection, analogous to those of the log-structured file systems [5]. This garbage collection process can severely degrade the performance and endurance of NAND flash memory. It is known that the overhead of garbage collection is closely related to the utilization of NAND flash memory [3], [4], [6], [7]. In this paper, we analyze the utilization of *trim*-enabled NAND flash memory device using the stochastic methods.

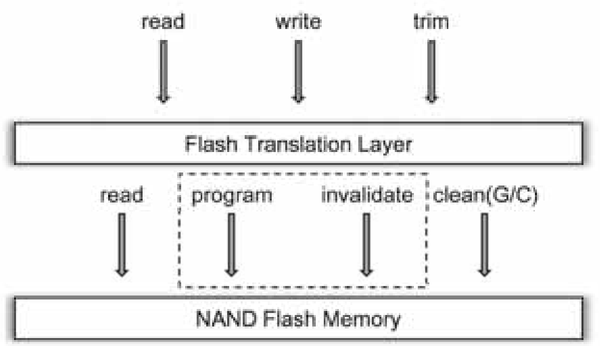


Fig. 1. Flash translation layer, which turns block-level operations into NAND-specific operations. Note that only program and invalidate operations affect the utilization of the NAND flash memory.

II. SYSTEM MODEL

Most NAND flash memory devices are equipped with the dedicated component called Flash Translation Layer (FTL), which is described in Fig. 1. One of main functions of FTLs is converting upper-level block operations into lower-level NAND-flash-memory-specific operations [9]. For example, upper-level write operation to update existing data may cause lower-level invalidate, program and potential erase operations. We assumed that FTLs employ fully page-level mapping algorithms and random write workloads are uniformly distributed over entire available memory space.

In this paper, we focus on the part of the lower-level operations, such as program and invalidate, which influence the utilization of available memory space, i.e. ratio of valid pages. Read and garbage collection doesn't affect the number of valid pages. Moreover, we took *trim* operation into consideration as well as the other upper-level operations. The trim explicitly informs FTL of obsolete pages and alleviates overhead of garbage collection. We think that this is one of the contributions of this paper.

III. UTILIZATION ANALYSIS

A. Notations

In this section, we derive the utilization of trim-enabled NAND flash memory. Some common notations that will be used throughout this section are presented in Table I.

TABLE I
NOTATIONS

Symbol	Description
α	Arrival rate of write operations
β_k	Arrival rate of trim operations, when $X = k$
m	Number of all pages
X	Number of valid pages (r.v.)
P_k	Steady-state probability, $P\{X = k\}$
λ_i	Transition rate from state i to state $i + 1$
μ_i	Transition rate from state i to state $i - 1$
λ	Arrival rate of program operations per system
μ	Arrival rate of invalidate operations per page, $(\alpha + \beta_m) / m$
U	Utilization of NAND flash memory (r.v.), X / m
γ	Trim intensity, β_m / α

B. M/G/s/s Queueing Model

We assumed that write operations arrive to the system according to a Poisson process with rate α. The arrival of trim operations is assumed to be an arbitrary random process with mean time $1/\beta_k$. In the case of trim, β_k depends on the ratio of valid pages in the system. This is one of main differences from [1], which assumed that the trim rate is constant. We assumed that more valid pages exist more likely trim happens.

Each write operation includes one program operation. When the write operation is performed on a valid page, which is named *update*, an additional invalidate operation occurs to mark the original page as invalid. Because we assumed that write operations are uniformly distributed over all m pages, the probability that each write is update is given by X / m. Trim operations can target only valid pages and the rate is proportional to the number of valid pages, which gives $\beta_k = \beta_m * k / m$. Only program and invalidate operations change the

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

number of valid pages in the system, respectively increasing and decreasing by 1.

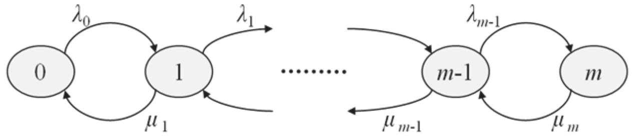


Fig. 2. State-transition-rate diagram for NAND flash memory.

Fig. 2 describes the system as a state-transition-rate diagram, where each state presents the number of valid pages in the system, denoted by the random variable *X*. The transition rate λ_i is equal to the rate of program operations, which can be directly given from the arrival rate of write operations α. The transition rate μ_i is the compound rate of update and trim operations. They are calculated as

$$\begin{aligned} \lambda_i &= \alpha && 0 \le i \le m-1 \\ \mu_i &= \alpha \cdot \frac{i}{m} + \beta_m \cdot \frac{i}{m} && 1 \le i \le m \\ &= \left(\frac{\alpha + \beta_m}{m}\right) i \end{aligned}$$

Again we define

$$\lambda = \alpha, \quad \mu = \frac{\alpha + \beta_m}{m}$$

Here we see that the system is the *M/G/s/s* queueing system, which is also known as *Erlang loss model* [2]. From [2], we can obtain the steady-state probability distribution

$$\begin{aligned} P_0 &= \left[\sum_{n=0}^{m} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}\right]^{-1} \\ P_k &= \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} P_0 \end{aligned}$$

Using *Little's formulas* [8], the expected value of *X* is given by

$$E[X] = \frac{\lambda}{\mu}(1 - P_m)$$

Finally, we obtain the average utilization of the system, that is,

$$\begin{aligned} E[U] &= E\left[\frac{X}{m}\right] = \frac{E[X]}{m} \\ &= \left(\frac{\alpha m}{\alpha + \beta_m}\right)(1 - P_m) \\ &= \left(\frac{m}{1+\gamma}\right) \frac{1 - \left(\frac{m}{1+\gamma}\right)^m \frac{1}{m!}}{\sum_{n=0}^{m} \left(\frac{m}{1+\gamma}\right)^n \frac{1}{n!}} \end{aligned}$$

, where γ is the ratio of β_m to α.

C. *Numerical Results*

As seen above, the average utilization depends on the number of pages *m* and the trim intensity γ. Fig. 3. Shows the expected free space ratio, which is given by 1 – *E*[*U*], varying *m* and γ. Note that the number of pages *m* has much less effect on results than the trim intensity γ when *m* is large enough. It means that we can approximate the model without using *m*. Also, we can verify that the enhancement of trim significantly helps NAND flash memory secure the free memory space. For example, 0.2 of trim intensity makes about 17% more free space ratio than no trim support.

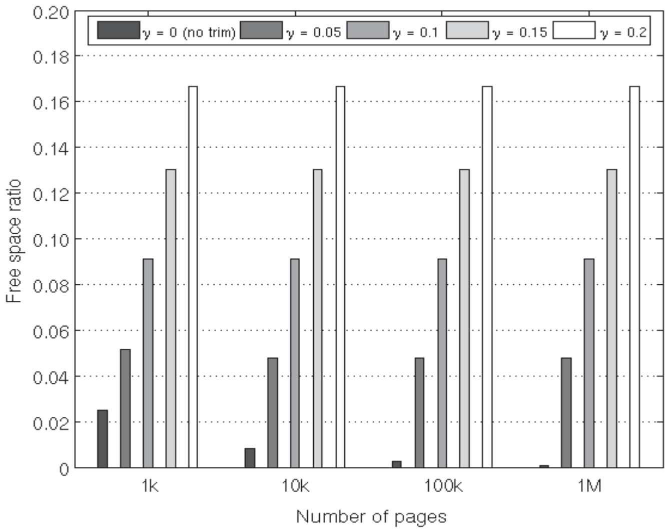


Fig. 3. Trim's effect on NAND flash memory utilization (free space ratio = 1 – *E*[*U*]). Note that γ is the ratio of β_m to α

IV. CONCLUSION

In this paper, we have presented one of our preliminary results of studying NAND flash memory performance. It provides a simple and powerful method to reason about the flash memory utilization when given trim intensity and the number of pages. As a future work, we are planning to extend this work for analyses of other performance metrics such as latency and endurance of trim-enabled NAND flash memory systems.

RERFERENCES

[1] T. Frankie, G. Hughes, and K. Kreutz-Delgado, "A mathematical model of the trim command in NAND-flash SSDs", *Proc. of the 50th Annual Southeast Regional Conference*, N Y, USA, Mar . 2012, pp. 59-64.

[2] L. Kleinrock, *Queueing Systems, Volume 1, Theory*, John Wiley & Sons, 1975, pp. 105-106.

[3] S. Boboila and P. Desnoyers, "Write endurance in flash drives: measurements and analysis", *Proc. of the 8th USENIX conference on File and storage technologies*, CA, USA, 2010.

[4] P. Desnoyers, "Analytic Modeling of SSD Write Performance", *Proc. of the 5th Annual International Systems and Storage Conference*, Haifa, Israel, Jun. 2012.

[5] M. Rosenblum and J. K. Ousterhout, "The design and implementation of a log-structured file system", *ACM Transactions on Computer Systems*, 10(1):26-52, Feb. 1992.

[6] R, Agarwal and M. Marrow, "A closed-form expression for write amplification in NAND Flash", *Proc. of the IEEE GLOBECOM Workshops*, FL, USA, Dec. 2010, pp. 1846-1850.

[7] X. Luojie and B. M. Kurkoski, "An improved analytic expression for write amplification in NAND flash", *Proc. of the International Conference on Computing, Networking and Communications*, HI, USA, Feb. 2012, pp. 497-501.

[8] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, Wiley Interscience, 1998, pp. 10-13

[9] E. Gal and S. Toledo, "Algorithms and data structures for flash memories", *ACM Computing Surveys*, 37(2):138-163, June 2005.

Patching A Patch - Software Updates Using Horizontal Patching

Milosh Stolikj, Pieter J. L. Cuijpers, and Johan J. Lukkien, *Member, IEEE*
Dept. of Mathematics and Computer Science, Eindhoven University of Technology,
P.O. Box 513, 5600 MB, Eindhoven, The Netherlands

***Abstract*—This paper presents a method for optimizing incremental updates of consumer electronic devices running multiple applications, called horizontal patching. Instead of using separate deltas for patching different applications, the method generates one delta from the other. Due to the large similarities between the deltas, this horizontal delta is small in size. In all test cases horizontal patching produced smaller deltas, with compression ratios between 8.02% and 43.38%.**

I. Introduction

Today's consumer applications, such as a DLNA based media system [1], are running on multiple, networked devices. A more pervasive upcoming system is an adaptive ambient lighting system [2], that employs a network of low capacity nodes with different roles. For instance, while some nodes measure luminance, others are responsible for switching the light actuators.

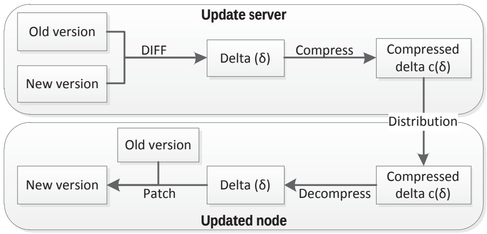


Fig. 1. Overview of an incremental update.

Updating software is an essential feature of modern CE devices, for the purpose of bringing new functionality, or correcting discovered bugs. Since the number of nodes to be updated can be large, the communication medium has limitations and the update should be swift, a software update is a non-trivial task. This is especially true for sensor networks where the lifecycle of nodes depends on small batteries.

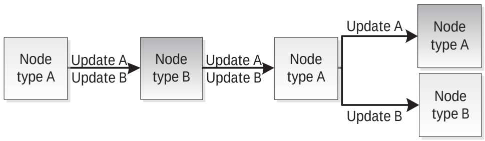


Fig. 2. Updating two different node types in a sensor network.

Software is most effectively updated in an incremental fashion [3] (Figure 1). Incremental updates use small scripts called *deltas* (δ), which contain instructions and data to produce an updated version from a previous one. In networks running multiple applications, incremental updates foresee separate deltas for each application. Since nodes of each type might be distributed throughout the entire network, in all update schemes that use multicasting as a basis, all deltas will be transmitted to almost every node (Figure 2).

In this paper, we present a new method for handling code differences in systems running multiple applications. Instead of distributing separate deltas for each application, one delta is generated from the other one, and both deltas are distributed together. The combined delta is smaller in size, hence less data needs to be transmitted, saving work, bandwidth and energy.

Related work on multiple deltas mainly focuses on incremental updates of a single application. In [4], multiple consecutive deltas for one application are merged to decrease the delta's size. Our work is complementary, focusing on situations where multiple applications need to be updated. In [5], an epidemic propagation protocol is described to handle the distribution of multiple deltas of applications for the Android system. The protocol assumes that one application can evolve into multiple orthogonal versions, hence multiple deltas exist for it. Their approach optimizes the gathering of deltas in an opportunistic fashion. In [6], updates of multiple applications are planned off-line by examining which combination of deltas has the smallest size. The method we present broadens the scope of the two works, by allowing one delta to be the source of another delta, essentially expanding that search space.

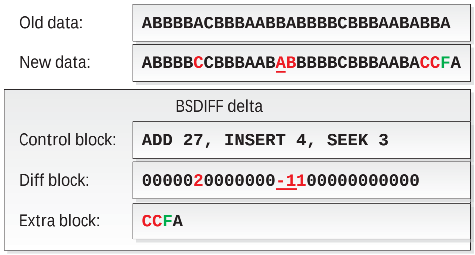


Fig. 3. Example of a BSDIFF delta. ADD specifies that the first 27 bytes from the old data and from the Diff block are summed. Zeroes in the Diff block mean that the corresponding byte from the old data is unchanged. INSERT adds four bytes from the Extra block to the output. SEEK moves the pointer in the old data three places forward, to the end of the stream.

II. BSDIFF Delta encoding

We use the BSDIFF [3] format for delta encoding. An update with BSDIFF is created in two steps (Figure 1). First, a delta (δ) between the two versions is constructed. Then, the delta is compressed using Bzip2 ($c(\delta)$) and sent to the node for update. There, after decompression, the delta is applied to the old version to reconstruct the new version.

BSDIFF has a two-pass algorithm to construct optimized deltas. In the first pass, completely identical blocks are found

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

in the two versions. Next, these blocks are extended in both directions, such that every prefix/suffix of the extension matches in at least half of its bytes. These extended blocks correspond to the modified code.

The BSDIFF delta is built of three parts (Figure 3): a control block of commands; a diff block of bytewise differences between approximate matches and an extra block of new data. When the old and new version are similar, the diff block consists of large series of zeroes, which are easy to compress.

III. Horizontal patching

Consider a network with two node types (A and B) as in Figure 2. Currently, when the operating system needs to be updated, two separate deltas are created, one for each node type. Both deltas are distributed independently.

In horizontal patching, one delta is used as a basis, and the other delta is an update of the first one (Figure 4). The motivation is that the operating system is the largest part of the software in both cases, and that is the part that is changed. Both deltas hold the same modifications, thus the horizontal patch between them is smaller in size than the vertical one.

The combined delta then consists of the basis and the horizontal delta, compressed together ($c(\delta_0+\delta_2)$, or $c(\delta_1+\delta_3)$). E.g., when δ_0 and δ_2 are used, only δ_0 needs to be executed for updating node type A. On node type B, first δ_2 is executed on δ_0, producing δ_1; finally, δ_1 is executed (Figure 5). The savings in space by using the combined delta in the multi-hop part of the network overweighs the loss in using it in the last-hop part.

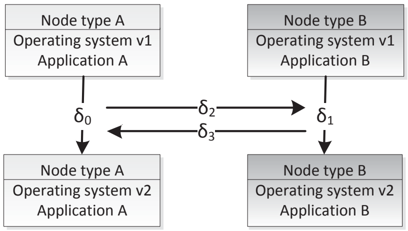


Fig. 4. Possibilities for horizontal patching.

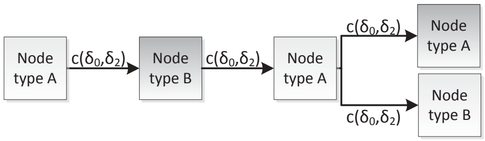


Fig. 5. Horizontal patching in practice.

IV. Evaluation

We tested on a sample set of seven applications for the Contiki operating system [7][1], and two consecutive operating system updates (Table I). We considered all combinations of 2 applications, and for each combination we computed the size of all deltas in Figure 4. The first application (node type A) was always larger than the second one (node type B).

We used the compression ratio, i.e., the percentage of data saved from transmission from the original data, as a metric: $cr = (1 - \frac{delta\_size}{c(\delta_0)+c(\delta_1)})100$. The results in Table II indicate that largest reductions in size are achieved by using horizontal patching from larger to smaller applications ($\delta_0+\delta_2$). Improvements differ depending on the type and size of applications. The gain is the smallest when both applications are larger than the operating system. When both applications are of similar size, horizontal patching gives up to 43% smaller deltas.

TABLE I
Size of sample data in bytes.

Application	Contiki 2.3	Contiki 2.4	Contiki 2.5
1	42,000	41,904	41,504
2	39,396	37,356	39,448
3	29,780	27,524	29,776
4	28,872	26,616	28,868
5	23,432	21,124	23,484
6	23,000	20,700	23,060
7	22,944	20,644	23,000

TABLE II
Compression ratio of compressed horizontal patches relative to vertical patches $(c(\delta_0) + c(\delta_1)$.

Delta type	Minimum	Maximum	Average	STDDEV
$c(\delta_0 + \delta_1)$	7.73%	29.67%	17.17%	5.67%
$c(\delta_0 + \delta_2)$	8.02%	43.38%	25.82%	10.13%
$c(\delta_1 + \delta_3)$	5.78%	43.00%	23.28%	11.20%

V. Conclusion

In this paper we presented horizontal patching, a method for optimizing the size of incremental updates in a multi-application environment. Horizontal patching reduces the size of updates by constructing one delta from another. We validated our hypothesis with experiments for two applications.

As future work, we foresee the analysis of impact of the additional processing in horizontal patching on the total delay of an update. Furthermore, when the number of applications increases, interesting combinatorics come into play for choosing the set of possible and optimal updates. Additional analysis would define which delta should be taken as a basis, and how the update should be formed.

Although the method was demonstrated in a sensor network, it can be used for updating CE devices which share the same code base. We showed that updating the firmware of several CE devices, as television sets or smart phones, can be done by a flooding scheme using smaller updates.

References

[1] Digital Living Network Alliance, *DLNA home networked device interoperability guidelines version 1.0*. 2004.

[2] S. Bhardwaj, A.A. Syed, T. Ozcelebi and J. J. Lukkien, *Power-managed smart lighting using a semantic interoperability architecture*. Conf. Consumer Electronics (ICCE), 2011, pp.759-760.

[3] C. Percival, *Matching with mismatches and assorted applications*. Ph. D. Thesis, University of Oxford, 2006.

[4] R. Kiyohara, K. Tanaka, Y. Terashima, *S/W upgrade for on-vehicle information devices*. Conf. Consumer Electronics (ICCE), 2012, pp.19-20.

[5] T. F. Bissyandé, L. Réveillère, J.-R. Falleri, and Y. Bromberg, *Typhoon: a middleware for epidemic propagation of software updates*. Middleware for Pervasive Mobile and Embedded Computing (M-MPAC), 2011.

[6] A. Shamsaie, and J. Habibi, *Planning updates in multi-application wireless sensor networks*. Symp. Computers and Communications (ISCC), 2011, pp.802-808.

[7] A. Dunkels, B. Grnvall and T. Voigt, *Contiki - a Lightweight and Flexible Operating System for Tiny Networked Sensors*. Work. Embedded Networked Sensors (Emnets-I), 2004, pp.455-462.

[1] The operating system and applications were compiled as one firmware image for commercially available sensor nodes.

Enhancing Application Performance by Memory Partitioning in Android Platforms

Geunsik Lim*, Changwoo Min† and Young Ik Eom‡
Sungkyunkwan University, Korea*†‡ Samsung Electronics, Korea*†
{leemgs*, multics69†, yieom‡}@skku.edu, {geunsik.lim*, changwoo.min†}@samsung.com

Abstract—**This paper suggests a new memory partitioning scheme that can enhance process lifecycle, while avoiding Low Memory Killer and Out-of-Memory Killer operations on mobile devices. Our proposed scheme offers the complete concept of virtual memory nodes in operating systems of Android devices.**

I. Introduction

Recent mobile phone users can use not only the built-in applications that the manufacturers included into the mobile phone, but also the third-party applications obtained from various app-markets. In these systems, due to the memory consumption of the third-party applications, there are frequent situations that the available memory space is insufficient to run those applications efficiently. Especially, in low-end mobile devices that do not have sufficient memory capacity, memory shortage may occur more frequently.

In this paper, we introduce a new memory partitioning scheme to get enhanced application performance during *process lifecycle* [1], while avoiding Low Memory Killer (LMK) and Out-of-Memory Killer (OOMK) operations on mobile devices. We propose a complete memory partitioning framework at the operating system level.

The rest of this paper is organized as follows. In Section II, several technical issues on *process lifecycle* are described. The new memory partitioning scheme for improving *process lifecycle* is suggested in Section III. Section IV shows the evaluation results of the proposed scheme. Finally, Section V concludes the paper.

II. Memory Management in Android Platform

The operating system generally supports page reclamation [2], swap in/out [3], *cgroups* [4], and OOMK [5] to settle the memory shortage problem.

The page reclamation mechanism is useful to obtain available memory in the system. However, the mechanism always finds victim processes heuristically among the processes in the memory.

The mobile device manufacturers do not use the swap in/out technology in their commercial products because of the throughput issues of their applications.

Although the *cgroups* provides a mechanism for aggregating/partitioning the set of tasks into several hierarchical groups, this mechanism does not prevent memory fragmentations because of the logical memory partitioning with private LRU of structure page cgroup per page.

The OOMK attempts to recover memory shortage from the OOM condition by killing low-priority processes [2] [5] which will most likely be the first victim. But, the operation of the OOMK results in the performance damage of new applications because of the thrashing occurred due to the limited memory resource of the mobile devices.

Android platform supports *process lifecycle* mechanism to classify the processes based on the importance of the processes so that new applications can get the needed memory properly even when it reaches the situation of memory shortage. It controls the memory usage of each application via user-space components (Activity manager, Dalvik) and kernel-space components (LMK, OOMK) [1] [5] to secure available memory stably.

Even under the system with large memory, memory shortage can happen when high-capacity and high-performance user applications come to run. Therefore, it is very important to secure as large available memory as possible. In Section III, we will describe our approach to solve this memory shortage problem on mobile devices.

III. Virtual nodes to avoid LMK operations

Figure 1 shows the overall architecture of the new memory partitioning technique for improving *process lifecycle* of the Android platform with the physically limited memory space. Our proposed memory partitioning technique mainly consists of three components as follows:

1) *vnode_setup_memblock*: sets up a memory node virtually from the start address to the end address.
2) *vnode_generation*: generates a physical memory configuration that maps between the virtual node and the physical memory address, and determines the size of the physical distance table.
    a) virtual node is a communication channel for logically partitioned memory access between the virtual memory and the physical memory.
    b) physical distance table is a map to get the physically separated memory block.
3) *vnode_set_cpumask*: allocates specific CPU masks for mapping between each CPU and each virtual node.

This work was supported by the IT R&D program of MKE/KEIT [10041244, SmartTV 2.0 Software Platform]. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology [2011-0025971].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

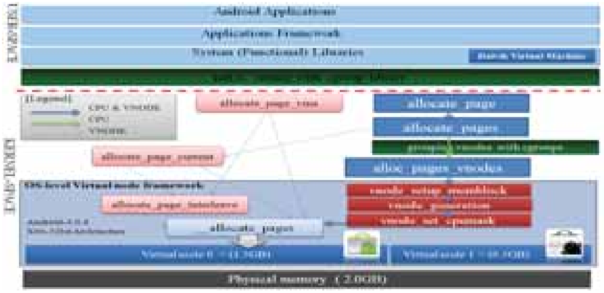
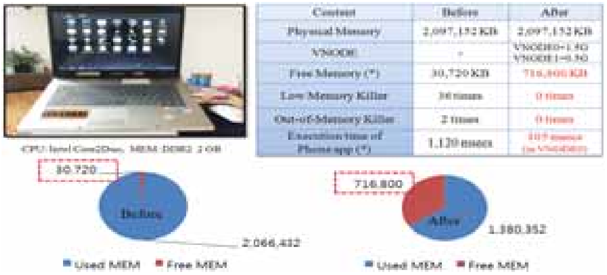


Fig. 1. The architecture of proposed memory partitioning



Content	Before	After
Physical Memory	2,097,152 KB	2,097,152 KB

VNODE	-	VNODE0=1.5G
		VNODE1=0.5G

Free Memory (*)	30,720 KB	716,800 KB
Low Memory Killer	36 times	0 times
Out-of-Memory Killer	2 times	0 times
Execution time of Phone app (*)	1,120 msecs	105 msecs (in VNODE0)

Fig. 2. The available memory result with virtual memory node

Our design has two advantages in Android-based mobile devices: (1) limiting the memory consumption of untrustworthy applications by partitioning the memory space into two areas, virtual node (VNODE) 0 for reliable applications (official market) and virtual node (VNODE) 1 for unreliable applications (black market), (2) avoiding LMK and OOMK operations which happen under physical memory shortage.

The arrows of Figure 1 represent the operations on the CPU and memory when an administrator sets the virtual memory nodes of the operating system from a physical memory on Android devices. For example, we run some critical applications only in VNODE 0. Also, we run other applications in VNODE 1. Through this approach, the operating system manages applications to avoid reaching the memory shortage even in a long running system.

Our memory partitioning scheme prevents an application from exhausting the entire memory by executing critical applications only in VNODE 0. Accordingly, these critical applications will stay in the memory of VNODE 0 continuously until a user terminates the critical application.

The key idea is that non-critical applications run in the physically partitioned specific memory area. This operation helps the system to avoid reaching no free memory. These non-critical applications only return their allocated memory with the page reclamation algorithm of Linux.

The proposed system completely offers virtual memory nodes at the operating system level for enhanced *process lifecycle* in Android devices. This equipment supports scalable system infrastructure as follows:

- Virtually separated memory space.
- Operating system level memory isolation.
- Advanced page reclamation based on virtual nodes.
- Memory controller interface at boot time.

IV. Results

We ported the latest *Android Ice Cream Sandwich 4.0.4* and *Busybox 1.18* to *Samsung SENS R60+ (CPU: Intel Core2Duo, MEM: DDR2 2G)* laptop to verify that our technical approach can be effective on the Android mobile platform. We also booted the *Android Ice Cream Sandwich* including our new memory partitioning scheme based on Linux kernel 3.0 as a test bed for the Android tablet platform. We configured the system by creating two virtual memory nodes in different size, *VNODE 0 of 1.5 GB and VNODE 1 of 0.5 GB*.

We evaluated and compared the memory consumption of the existing approach (*before*) and the proposed approach (*after*) when we executed the sequential file I/O operation with the raw contents of the size of 1.5 GB into VNODE 1 for 2 days. Figure 2 shows the available memory size, the result of LMK, and the status of the OOMK after running the sequential file I/O operations.

From our experiments, we gained the additional free memory of 670 MB and the reduced *Phone* application execution time of 1,015 milliseconds over the existing systems. Since the test workload on VNODE 1 can use only 0.5 GB memory, Linux kernel executes many page reclamation operations in VNODE 1. Through our approach, the proposed system does not meet the operation of LMK and/or OOMK which operates on free memory shortage, 335 MB in our experimental environment. The frequencies of the execution of memory killers, both LMK and OOMK, were improved dramatically after adjusting the virtual nodes based on the new memory partitioning scheme.

V. Conclusion

We proposed a virtual memory node technique for memory partitioning. It focuses on the page reclamation operation of non-critical applications and the non-page reclamation operation of critical applications. Also, our approach supports virtual memory isolation to separately run applications of black markets and applications of official markets in the Android platform based on discontiguous memory access model. These approaches prevent LMK and OOMK from killing processes because of the memory shortage of the system.

In conclusion, our approach innovatively overcomes the poor performance of applications incurred due to the operations of LMK and OOMK, without any physical memory extension.

References

[1] Google, “Android Application’s Life Cycle (Process, Activity),” in *http://developer.android.com/reference/android/app/Activity.html*, 2008.

[2] M. Gorman, “Understanding the Linux Virtual Memory Manager,” in *Prentice Hall Professional Technical Reference*, March 2004.

[3] N. Gupta, “Compcache; in-memory compressed swapping,” in *http://lwn.net/Articles/334649/*, May 2009.

[4] B. Singh, “Containers: Challenges with the memory resource controller and its performance,” in *Linux Symposium*, vol. 2, pp. 209–222, June 2007.

[5] D. Rientjes, “OOM Killer Rewrite; When the Kernel Runs Out of Memory,” in *LinuxCon Boston*, August 2010.

RAID-Optimal Data Placement in a Hybrid Solid-State Drive

Jungmin Seo, Jupyung Lee, Boncheol Gu, Hyun-Jung Shin, and Brian Myungjune Jung
Samsung Advanced Institute of Technology, Samsung Electronics

Abstract—**Flash-based solid-state drives (SSD) are widely accepted from mobile devices to enterprise storage arrays. They are best known for delivering low power consumption and high random I/O performance that far exceeds hard-disk drives (HDD). One configuration to enhance performance and reliability further is to connect multiple SSDs to a RAID controller. In this paper, we propose an intelligent, hybrid SSD that is optimized in distributed-parity RAID schemes by recognizing data-parity block sequence and placing parity blocks in higher endurance memory cells. The experimental results show that our predictive data placement extends SSD lifetime by 31% on average, compared to the history-based hot data placement.**

I. INTRODUCTION

Flash-based solid-state drives (SSD) operate differently from traditional magnetic disks such as hard disk drives (HDD). SSDs no longer have any moving mechanical components. They have complex processing logic software inside, called Flash Translation Layer (FTL), which significantly influences the SSD performance and lifetime. This layer carries out address mapping, garbage collection and wear-leveling strategies to manage physical flash memory.

One critical concern of flash-based SSDs is their limited lifetime. This limit originates from flash memory's finite number of program/erase (P/E) cycles. Flash memory does not allow in-place-updates and a P/E cycle is to be initiated before an update, which reduces lifetime of SSDs. Modern SSDs adopt log-structured design on their FTLs to accomplish out-of-place-updates and to hide the latency of a slow P/E cycle, but the lifetime limit still exists. NAND flash memory has two different types: single-level cell (SLC) and multi-level cell (MLC). SLC stores one bit per cell and MLC stores more than one bit per cell. Storing more bits per cell achieves a higher capacity, but it makes devices less reliable by suffering from higher bit-error-rate, and the MLC devices have less lifetime.

The Redundant Arrays of Inexpensive Disks (RAID) [1] aggregates multiple devices to improve reliability and performance. Many existing computing systems have adopted the RAID technology to protect device-level failures. Parity-based schemes, among others, are known to be cost-effective with redundancy of N+1 or N+2. When these are applied to SSDs, the faster random read latency of flash memory becomes an advantage since every data update requires reading the previously stored data and parity.

In parity-based RAID schemes, when a data block is updated, the corresponding parity block is also to be updated. It generates one random update operation to the SSD holding the parity block. Since the parity blocks are updated more frequently than the data blocks, these accumulated parity updates affect the overall lifetime of the SSD significantly.

In this paper, we introduce a hybrid SSD that recognizes the data-parity block sequence, and that places the parity blocks in high endurance memory cells and the data blocks in low endurance memory cells. Our RAID characteristic-aware data placement extends the lifetime of the low endurance memory, and that of the overall storage system. We also describe how to achieve the block sequence and show evidence on the effectiveness of our predictive data placement.

II. RAID-OPTIMAL HYBRID SSD

A. Parity Block Sequence Recognition

Our motivation comes from the input pattern to an SSD in a distributed parity-based RAID scheme such as RAID-5. One recognizes that there is a periodic pattern of parity blocks in such schemes. As shown in Fig. 1, a parity block is inserted every N blocks, where N is the number of total disks. The parity block sequence is dependent on the current RAID scheme, the total number of disk drives and the disk's location in the RAID group. We can calculate the parity block address with the modulo operation.

Other distributed parity schemes, such as RAID-6, have the similar sequence property, and our approach is also applicable.

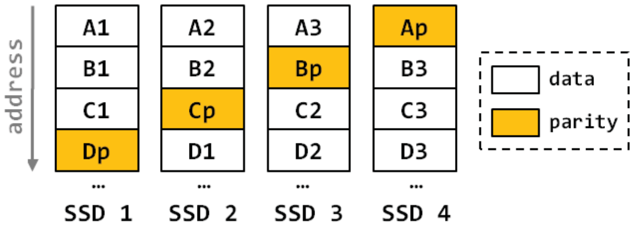


Fig. 1. RAID-5 Left-Symmetric Scheme. It shows that the block address corresponding to parity is written periodically in one SSD.

One method to recognize the sequence is by communicating with the RAID controller. The controller informs each device of the current RAID configuration for the device, e.g. the 3rd from total 5 drives in RAID-5 left-symmetric, or of each one's parity block sequence information directly.

The other method, more sophisticated, is by self-learning the parity block sequence. Each SSD figures out the sequence from data input access patterns. The advantage over the communication method is that this is fully compatible with legacy RAID controllers. No modification or code addition is required in the RAID controllers. In addition, a supervisor application program on the host can be used to accelerate the learning process.

B. Sequence-based Hybrid SSD

The parity block sequence information enables two key

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

features: configuration and data placement. It is used to configure the capacity ratio of high and low endurance memory cells. One simple configuration is to set the ratio to be the same as the ratio of parity and data block capacity. One better configuration takes the workload characteristics and its hybrid memory management strategies into consideration.

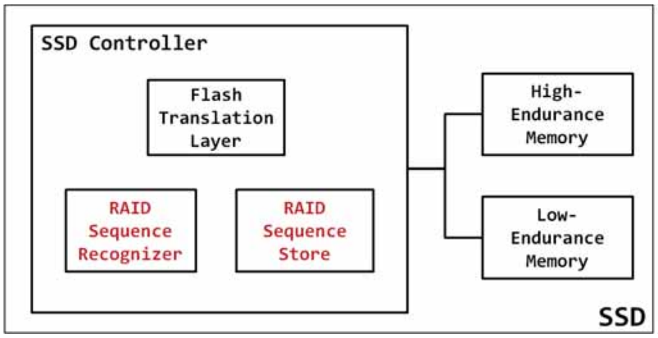


Fig. 2. RAID-Optimal Hybrid SSD Architecture. The sequence recognizer identifies the parity block sequence and stores it in the sequence store. The FTL places data based on the sequence information stored.

Based on this sequence, FTL stores the new parity blocks in the high endurance memory cells and the new data blocks in the low endurance memory cells. Update requests on both blocks are processed on the previously stored memory cells.

III. EVALUATION

The purpose of our experiment is to prove that our sequence-based placement approach improves the SSD lifetime. We compare it to the history-based data placement which stores more frequently-written data in the high endurance memory cells, regardless of parity and data. Since the rewrite frequency is not known at the beginning from the perspective of a device, it first stores all the data from high endurance to low endurance memory cells sequentially. Once the frequency of some data reaches a threshold, it stores the updated data into the high endurance memory.

We use the DiskSim simulator [2] and the SSD model [3] respectively developed by Carnegie Mellon University and Microsoft Research for our evaluation. Both have been widely used for research and development in HDDs and SSDs.

TABLE I
SSD MODEL PARAMETERS

Parameters	SLC	MLC
Total Capacity	8GB	32GB
No. of Packages	2	4
Pages per Block	64	128
Max. Erase Count	50,000	3,000

We extend the original simulators in two ways: by configuring hybrid models in the SSD model parameters and by implementing our own synthetic workload generator using the DiskSim's external interface. The workload generator implements the RAID-5 left-symmetric scheme and distributes requests to each DiskSim instance that represents a hybrid SSD. The SSD simulator is configured as shown in Table I. The default configurations from the MSR's SSD model are used, unless otherwise denoted. We simulate five hybrid SSDs. Each SSD has the ratio of SLC/MLC to be 1:4, which is the parity-data block ratio in an SSD.

We generate random workloads to the simulator. The rewrite requests are randomly distributed across the logical block addresses, so that after time all the data block addresses have about the same rewrite frequencies. The parity blocks are four times more frequently updated than the data blocks.

We insert the same 216M random writes to each approach. The workload corresponds to 15K random write IOPS running 4 hours a day, which is known as one of the de-facto methods to measure SSD lifetime.

As shown in Table II, our approach executes over 23% less P/E cycles on the MLC memory cells. It illustrates that our approach identifies hot data more effectively in the RAID-5 environment. Assuming that wear-leveling is well achieved, it extends the estimated lifetime by 31% on average.

TABLE II
EXPERIMENTAL RESULTS

	Sequence-based	History-based
Avg. Daily P/E Count (MLC)	12.89	16.95
Estimated Lifetime (in years, MLC)	0.64	0.48

The efficiency of the history-based approach is dependent heavily on the workload characteristics and the dynamic optimal threshold selection. It also requires large footprints. However, our sequence-based approach is more predictive on the hot data identification by leveraging the RAID characteristics. It has very small footprints. We plan to extend our work further by combining the two approaches.

IV. CONCLUSION

In this paper, we propose a new hybrid SSD that self-optimizes in distributed-parity RAID schemes. It recognizes parity block sequence and places parity blocks in high endurance memory cells. The ways to identify the parity block sequence and how to leverage it to build a RAID-optimal hybrid SSD are presented. We provide empirical evidence that our sequence-based data placement significantly improves SSD lifetime upon the history-based data placement.

REFERENCES

[1] D.A. Patterson, G. Gibson, R.H. Katz. "A Case for Redundant Arrays of Inexpensive Disks (RAID)", *SIGMOD '88 Proceedings of the 1988 ACM SIGMOD International Conference on the Management of Data*, vol. 17, no.3, pp. 106-116, 1988.

[2] J. Bucy, J.Schindler, S. Schlosser, G. Ganger, "The DiskSim Simulation Environment Version 4.0 Reference Manual", *Carnegie Mellon University Parallel Data Lab Technical Report CMU-PDL-08-101,* May 2008

[3] N. Agrawal, V. Prabhakaran, T. Wobber, J.D. Davis, M. Manasse, R. Panigrahy, "Design tradeoffs for SSD performance", *Proceedings of USENIX 2008 Annual Technical Conference,* pp. 57-70, 2008

S^3-RNC: A Novel V2V Transmission Scheme for Mobile Content Distribution

Woojin Ahn and Young Yong Kim
Yonsei University
Seodaemun-gu, Seoul, 120-749, Korea

Ronny Yongho Kim
Korea National University of Transportation
Uiwang, Gyeongki, 437-763 Korea

***Abstract*—In this paper, we propose a novel V2V transmission scheme, namely S^3-RNC (Shuffled Scattered Symbol-level Random Network Coding), for Mobile Content Distribution (MCD) between communication pairs traveling on opposite directions. To cope with the Doppler Effect caused by extremely high relative velocity, symbol-level random network coding (RNC) with shuffling (interleaving) and scattering is used in our proposed scheme. As the second step, receivers, moving together as a cluster, cooperatively relay their received S^3-RNC coded blocks, in order to maximize the probability of successful decoding. Our simulation results show that S^3-RNC improves throughput performance significantly in high mobility environment. By utilizing S^3-RNC, consumer vehicular communication devices are able to provide a lot of useful information to vehicular users.**

I. INTRODUCTION

With remarkable advances in consumer electronics and telecommunications technologies, development in the automobile industry has been progressed to telematics era, in which interaction between users and various platforms is very important [1]. The telematics market value is projected to reach $40.3 billion in 2016 at 5 year CAGR of 20.9% [2].

In mobile content distribution (MCD), which is one of the most promising platforms in telematics systems, multimedia contents of Area of Interest (AoI), including traffic information and local commercials, are distributed from fixed infrastructure devices to telematics devices installed inside vehicles (I2V) or relayed in vehicle to vehicle (V2V) manner [3]. Various protocols have been proposed in the literature to implement MCD and most of them have focused on transmission between I2V or V2V in the same lane. From the telematics users' point of view, generally AoI is where telematics users are heading and telematics devices can request the nearest infrastructure devices or neighboring vehicles to transmit contents of the AoI. In mobile content distribution, if the less infrastructure devices get involved, we can achieve the more cost effectiveness due to spectrum reusability and efficient power consumption. V2V contents distribution among vehicles in the same lane, contents of next AoI cannot be obtained efficiently, since contents are delivered through multi-hop relaying from a limited number of neighbors. On the other hand, if vehicles are able to directly communicate with vehicles traveling in the opposite lane, high spectrum reusability and power efficiency can be achieved due to short range communications of less than 20 meters. Moreover, as receivers have high probability of facing a large number of transmitters from various regions, mostly uncovered area for the transmitter, contents diversity can be achieved without multi-hop relaying. However, there are a couple of challenges in such opposite lane V2V communication scenario. Since relative speed of receivers and transmitters may reach up to 300 *km/h*, Doppler Effect will degrade system performance severely. There might not be enough time for packet retransmission in case of packet transmission failure due to fast moving speed. There have been several successful attempts to apply random network coding to wireless networks. Thanks to salient properties of random network coding, e.g., rateless and randomness, random network coding can provide efficient transmission schemes in wireless network.

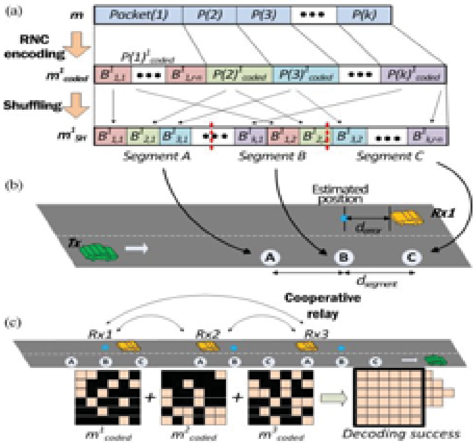


Fig. 1 Example of the proposed scheme

In this paper, we propose a novel transmission scheme for V2V MCD considering the aforementioned challenges, namely S3-RNC (Shuffled Scattered Symbol-level Random Network Coding). Using Symbol-level Random Network Coding (S-RNC), a very small unit of uncorrupted bits can be used for decoding that is highly efficient transmission can be obtained. In order to overcome the challenges of the opposite lane V2V communication scenario, two additional novel schemes: Shuffling and Scattering are combined with S-RNC

II. PROPOSED SCHEME: S^3-RNC

In Fig. 1, The transmitter node, *Tx*, and the receiver node, *Rx*, are traveling in opposite directions. Since vehicles in the same lane are typically moving together, *Rx* nodes could form a cluster and share their general mobility information of neighboring vehicles including velocity and location. A cluster header advertises such information [4], so that *Tx* could estimate when to start data transmission. As illustrated in Fig. 1, in order to explain the proposed scheme, S^3-RNC, we assume that each cluster node shares the content, *m* with *k* packets. S^3-RNC utilizes the following 3 salient schemes.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

i) Symbol-level Random Network Coding (S-RNC): To generate random-network-coded message for *Rx* (m^{Rx}_{coded}), *Tx* first divides each single packet into small blocks ($B^{Rx}_{k,i}$) with a certain batch-size (*n*). Then, each block is encoded using S-RNC with random coefficients generated in a given Galois field (GF) [5]. In order to generate redundancy for Forward Error Correction (FEC), when a code rate is *1/r*, *r·n* coded-blocks for each packet are generated.

ii) Shuffling (SH): After generation of the coded blocks, the generated coded blocks are shuffled using a simple shuffling rule. For example, as shown in Fig. 1 (a), the first blocks of each coded packet ($B^{Rx}_{k,1}$) are placed at the first position of the coded-message. Then, the same rule is applied for the next blocks. The shuffled message is denoted by m^{Rx}_{SH}.

iii) Scattering (SC): Before the message transmission, m^{Rx}_{SH} is divided into several segments. Each segment is transmitted at a time slot calculated with predefined distance ($d_{segment}$), as illustrated in Fig. 1(b). Number of segments and $d_{segment}$ may vary depending on number of cluster nodes, size of packets, mobility of cluster and distance between neighboring nodes.

By using RNC, any uncorrupted coded blocks can be used to perform decoding. Therefore, as long as *Rx* is able to collect uncorrupted coded blocks more than a certain batch size, *n*, *Rx* is able to decode m^{Rx}_{coded}. After reception of coded blocks, receivers in a same cluster may perform cooperative relaying (CR) in order to exchange their own received coded blocks. With CR, number of uncorrupted blocks can be further maximized. An example of CR is shown in Fig. 1(c).

Shuffling contributes to equalize clean block ratio of each packet. As phases of received symbols rotate continuously with small-scale fading, coded blocks of specific position might be corrupted severely compare to others, depending on the initial phase. Since shuffling distributes phase distortion effect across an entire message equally, random network coding gain and cooperative relaying gain can be maximized.

GPS error and velocity changes of communication pairs may lead imperfect estimation of transmission position and timing. Through scattering, potential burst loss caused by such an imperfect estimation can be minimized.

III. PERFERMANCE EVALUATION

In simulation, the same topology in Fig. 1 (c) is used with the distance between vehicles of 40 *m*, and $d_{segement}$ of 10 *m*. The original content of 4608 *bits* (512·9 *packets*) is RNC coded with code rate 1/3, and transmitted over Rayleigh fading channel. The transmission power and the noise power are set to be 10*dBm* and -95 *dBm*, respectively. The relative velocity of communication pairs is set to be 200 *km/h*. The estimation error of transmission position follows zero mean Gaussian distribution with standard deviation, σ. Fig. 2 shows the packet error rate (PER) of the proposed scheme as standard deviation of estimation error varies. In order to compare S^3-RNC with existing schemes, the PER result of Convolution Code (CC) using the same code rate is also depicted. Also, we can see that the proposed scheme outperforms the conventional CC in PER performance, and shuffling provides additional performance enhancement. As estimation error gets larger, the proposed scheme with scattering outperforms others.

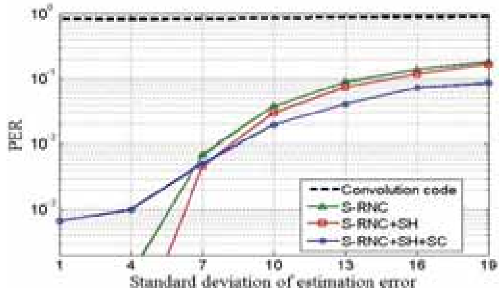


Fig. 2 PER performance versus standard deviation of estimation error

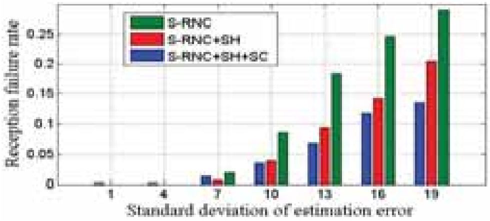


Fig. 3 Reception failure rate versus standard deviation of estimation error

As *Tx* is not able to retransmit corrupted packets, it is important to see whether the whole component of the target content can be delivered completely with limited chance of transmission. If there is more than one packet failed to decode, then the reception of target content is regarded as failure. Fig. 3 depicts the reception failure rate of S^3-RNC. And the proposed scheme shows a prominent performance, less than 0.15, both at high and low estimation error cases.

IV. CONCLUSION

We have proposed a novel transmission scheme, S^3-RNC, for V2V-MCD, especially focusing on the opposite lane case. Our simulation results prove that severe throughput degradation caused by extremely high relative velocity can be overcome by using S^3-RNC. Additional schemes (SH, SC) of the proposed algorithm further enhance the performance.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2012R1A1A1014610). R. Y. Kim is the corresponding author.

REFERENCES

[1] C. Lina, M. Hsiehb, G. Tzengb, "Evaluating vehicle telematics system by using a novel MCDM techniques with dependence and feedback," *Expert Systems with Applications*, vol. 37, pp. 6723-6736, Oct. 2010.

[2] L. Weisenbach, "Telematic Components: Technologies and Global Markets," *BCC research,* IFT062A. Apr. 2010.

[3] M. Li, Z. Yang, W. Lou, " CodeOn: Cooperative Popular Content Distribution for Vehicular Networks using Symbol Level Network Coding," *IEEE Journal on Selected Areas in Communications,* vol. 29, No. 1, pp.1-14, Jan. 2011.

[4] Y. Gunter, B. Weigel, "Cluster-based medium access scheme for vanets," in Proc. *ITSC*, pp.343-348 Oct. 2007.

[5] R. Y. Kim, J. Jin, B. Li, "Drizzle: Cooperative Symbol-Level Network Coding in Multichannel Wireless Networks," *IEEE Trans. Vehicular Technology,* vol. 59, no. 3, pp. 1415-1432, Mar. 2010

High-definition Video-based Multi-channel Top-view Vehicle Surrounding Monitoring System for Mobile Navigation Devices

SungRyull Sohn, Hansang Lee, Heechul Jung, and Junmo Kim, *Member, IEEE*
Dept. of Electrical Engineering, Korea Advanced Institute of Science and Technology, South Korea

***Abstract*—Providing visual information around automobile can make drivers blind-spot-free, which is helpful for several occasions such as parking, lane change and backward movement. In this paper, the vehicle surrounding monitoring system with multiple channel high-definition (HD) videos, which is embedded in portable navigation devices, is presented. The system collects three channel video inputs from rear, left, and right side of the vehicle, transforms each inputs into top-view image, respectively. Finally, the transformed outputs are aligned, and then displayed on the screen for assisting the driver. Implementation and its results showed that the system provides HD vehicle surrounding monitoring video which is helpful especially in parking assistance.**

***Index Terms*—Vehicle surrounding monitoring system, multi-channel video, top-view transformation, parking assistance**

I. INTRODUCTION

In automobiles, drivers usually focus on the frontal view while they are driving, and it causes blind spots including rear, left, and right side of the car. Though there are rear-view and two side mirrors which support drivers to cover these blind spots, since it can be confusing to watch three mirrors alternately while focusing mainly on the frontal view, drivers are still less aware of their blind spots.

To overcome this limitation, efforts to utilize the visual information of these blind spots and to provide them in a simple, unified interface has grown in the recent years. Ishii et al. [2], Kano et al. [3], and Lin and Wang [5] suggested the technique which transforms only the rear view camera images into top-view images. This rear-view-only technique is useful for parking assistance specifically, however, it doesn't cover all the blind spots including left and right side of a vehicle. On the other hand, Chen et al. [1] and Liu et al. [6] proposed all-around top-view monitoring system, which displays entire region surrounding vehicle including frontal view, with four and six fish-eye cameras, respectively. The all-around top-view technique provides more intuitive information to drivers than the rear-view-only technique does. However, since it deals with multiple sources of fish-eye-distorted images, its computation is complex to display as a video. Moreover, the frontal view information in this all-around top-view result is almost unnecessary since drivers are usually fully aware of their frontal view. As an alternate, Li and Hai [4] suggested three-channel top-view system including rear, left and right view images, which mainly focused on blind spots. In [4], the system incorporates three fish-eye cameras mounted on both sides and on the rear of the vehicle and shows promising results.

In this paper, we construct the blind-spot-free vehicle surrounding monitoring system using three-channel top-view videos. We embed this system into portable navigation devices, which are widely used in car environment currently, with three HD cameras mounted on both sides and on the rear side of the vehicle. Fig. 1 shows the locations of cameras mounted on the vehicle and an output image of our system.

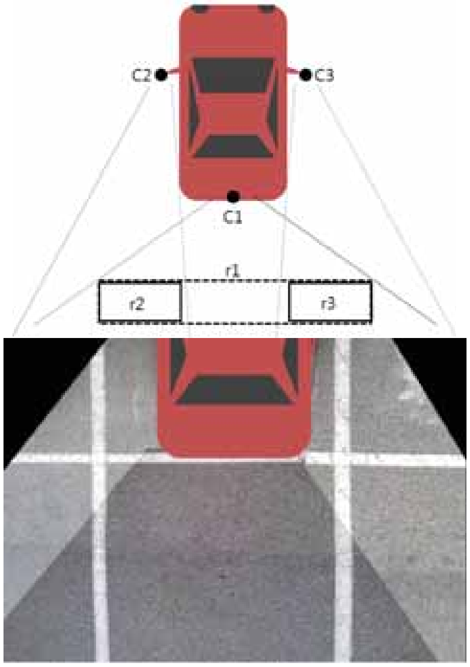


Fig. 1-(a): (top) Locations of cameras(C1, C2, C3) and reference rectangles (r1, r2, r3) (b): (bottom) Output image of our system.

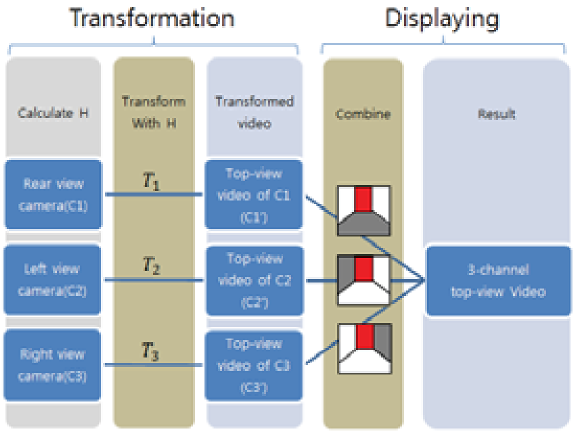


Fig. 2: Pipeline of the proposed system

II. PROPOSED SYSTEM

Fig. 2 shows the pipeline of the proposed system. In the system, three videos taken by HD cameras mounted on rear,

This work was supported by Leading Industry Development for Chung-cheong Economic Region. (A0022 00627)

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

left and right of the vehicle are put into the system as inputs. As shown in Fig. 2, our system consists of two parts: a top-view transformation part and a display part. In the top-view transformation part, three projective perspective sequential images are transformed into corresponding top-view images. Using these top-view transformed images, in the display step, the system combines these images to produce single output sequential image as Fig. 1 (b). As a result, the output images are displayed as a video on portable navigation devices.

A. *Top-view transformation*

With three input videos, the first part of the proposed system is to extract their top-view transformation parameters and to perform top-view transformation to input images. First, parameters for top-view transformation of each image are extracted according to its geometric position. Since the perspective projection of cameras is simply projective, its corresponding parameters can be computed with basic homography [2], [3]. To reduce computation time for top-view transformation of each image sequence, hash tables for each camera are created with corresponding top-view transformation parameters. With these hash tables, top-view images can be computed via simple coordinate substitution.

B. *Images combination and display*

With three top-view transformed images, the second part of the proposed system is to combine them by aligning each part. To combine these images, the transformed rear view image is set as a reference image. With the rear view image as a reference, the other two transformed side view images are adjusted to the reference image. To arrange these images accurately, the calibration technique using information of overlapping regions is applied. Finally, we create a combined hash table for output image with three hash tables computed in top-view transformation and combining adjustment of each image. For each pixel of output image, corresponding type of cameras and its coordinate are stored in the combined hash table. The output image is computed using this table.

III. IMPLEMENTATION

For experimental setting, three HD cameras were mounted on the surface of the vehicle, two on both side mirrors and one on the rear window. Three videos with the setting of 1280x720 and 24fps were acquired and put into the system as inputs. Fig. 3 shows three input sequential images and their computation results corresponding to each step. The computation time of entire steps is about 10.5ms on average. As a result, the output sequential image is with the setting of 900x630 and 24fps. Fig. 3 shows the output image displayed on the screen with the vehicle diagram.

The detailed algorithm can be described by the pseudo-algorithm below.

1. Get the transformation matrices for each camera from the reference rectangles (r1, r2, r3).
2. Create hash table from the matrices obtained in step. 1.
3. Transform each image to form output image with the hash table obtained in step.2.

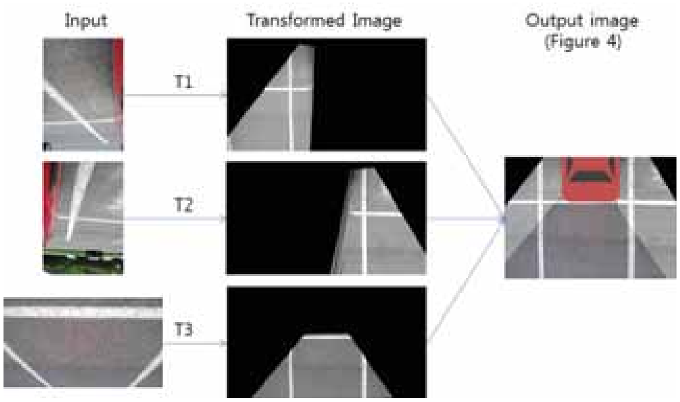


Fig. 3: Three sequential images on each step and output image.

IV. CONCLUSION

In this paper, we proposed real-time multi-channel top-view vehicle surrounding monitoring system with HD videos. Three cameras are mounted on the rear, left and right side of the vehicle and the system is embedded into portable navigation devices. The output video provides intuitive information of blind spots around vehicles. The implementation results show that the proposed system is practical and helpful in several driving situations like parking assistance.

ACKNOWLEDGMENT

The authors thank all the colleagues at KAIST, TeleWorks, Finedigital Inc., and Crewave who worked on the project.

REFERENCES

[1] Y-Y. Chen, Y-Y. Tu, C-H. Chiu, and Y-S. Chen, “An Embedded System for Vehicle Surrounding Monitoring,” *Proc. 2nd International Conference on Power Electronics and Intelligent Transportation System*, 2:92-95, 2009.

[2] Y. Ishii, K. Asari, H. Hongo, and H. Kano, “A Practical Calibration Method for Top View Image Generation,” *International Conference on Consumer Electronics*, 1-2, 2008.

[3] H. Kano, K. Asari, Y. Ishii, and H. Hongo, “Precise Top View Image Generation without Global Metric Information,” *IEICE Transactions on Information and Systems*, E91-D(7):1893-1898, July 2008.

[4] S. Li and Y. Hai, “Easy Calibration of a Blind-Spot-Free Fisheye Camera System Using a Scene of a Parking Space,” *IEEE Transactions on Intelligent Transportation Systems*, 12(1):232-242, March 2011.

[5] C-C. Lin and M-S. Wang, “A Vision Based Top-View Transformation Model for a Vehicle Parking Assistant,” Sensors, 12:4431-4446, 2012.

[6] Y-C. Liu, K-Y. Lin, and Y-S. Chen, “Bird’s-Eye View Vision System for Vehicle Surrounding Monitoring,” *Proc. 2nd International Conference on Robot Vision*, 207-218, 2008.

Guidance Protocol via personal ITS station for advisory safety systems

Jeong-Dan Choi, Kyoung-Wook Min
Electronics and Telecommunications Research Institute

***Abstract*-- We propose the guidance information protocol to provide the real-time decision support system to drivers or pedestrians using personal ITS station. The reference architecture provides a general structure for the real-time decision support system and the method of message exchange between the personal ITS station and the roadside ITS station. The proposed method is a flexible application protocol for safety warning and parking guidance services. This protocol makes the client part independent of use cases for supporting light-weighted devices. We aim to design the protocol to be flexible for accommodating various use-cases as well as suitable for being implemented on light-weighted personal devices.**

I. INTRODUCTION

In general, application level protocols for intelligent transportation systems (ITS) define some important applications, their messages and message transmission sequences. These predefined applications are generally called use cases. The messages and message sequences of the ITS application protocol which are fixed and applications, should implement rigid-formatted message set and message sequences for each use-case. Global Telematics Protocol (GTP) [1] and Mobile Location Protocol (MLP) [2] are typical examples of the use case based protocol.

However, there are two important issues on existing protocols for ITS applications [3][4]. The first is that message formats and message sequences in an application protocol are fixed. This can cause a problem because use-cases in an application protocol can be frequently inserted, modified and deleted. The second is, more importantly, that a user device should implement all the use-cases. And the Personal ITS stations, by nature, requires light-weighted applications due to the limitation of resources.

In this letter, we propose an application protocol for safety warning and parking guidance considering aforementioned issues. Although we present a protocol for a specific purpose, our scheme for making a protocol independent of use-cases can be widely used in ITS and Telematics domains.

II. APPLICATION PROTOCOL DESIGN

In this section, we present an overall architecture for safety warning and parking assistance applications that are composed of P-ITS-S(personal ITS station), R-ITS-S(road-side ITS station) and C-ITS-S(central ITS station). We develop a novel technique for designing application protocols for light-weighted personal devices. The application can be described by the use case description format (UDF). And also we propose primitive data elements for safety warning and parking guidance services.

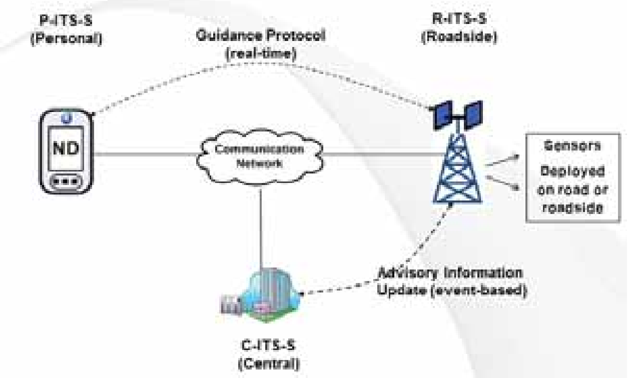


Fig. 1. Guidance Protocol and Advisory Information update flow

The guidance protocol and advisory information update flows are shown in Fig. 1. We describe our protocol designing scheme for supporting light-weighted personal devices in Section III. We present UDF in Section VI and we conclude this letter in Section V.

III. DESIGNING SCHEME

The environments of road and parking lots are subject to change with time and place while driving, the use cases are frequently added, modified, and deleted on personal device. In our protocol, R-ITS-S transmits a use case description data instead of a software module. In this letter, we use the term of use case instance to refer to such a use case description data. A use case instance specifies a set of message templates and a sequence of message exchanges. The client program in a terminal device exchange messages with a local server according to the sequence specified in the use case instance. A message transmitted to the terminal device is in a binary format. It is instantiated using the corresponding message template.

A use case instance is made by the XML document and should satisfy the description format.

To express the use case in a machine readable format for both the server and nomadic device, the means are needed that include rules for the description of the use case. In our protocol, these rules are defined using a document type definition (DTD) and the use case description format (UDF). The overall scheme is depicted in Fig. 2. A use case designer make a use case instance following the UDF. A local server implements the use case instance and transmits the instance to a terminal device when the terminal device enters into the

This work was supported by the Industrial Strategic Technology Development Program (10035250, Development of Spatial Awareness and Autonomous Driving Technology for Automatic Valet Parking) funded by the Ministry of Knowledge Economy(MKE, Korea).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

server's boundary. The client program can interpret the use case instance, construct necessary messages, and exchange messages according to the protocol.

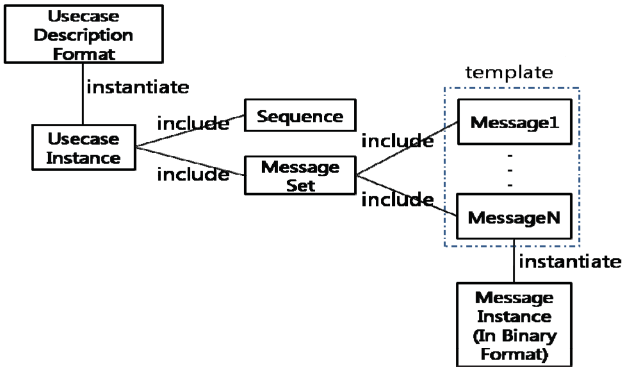
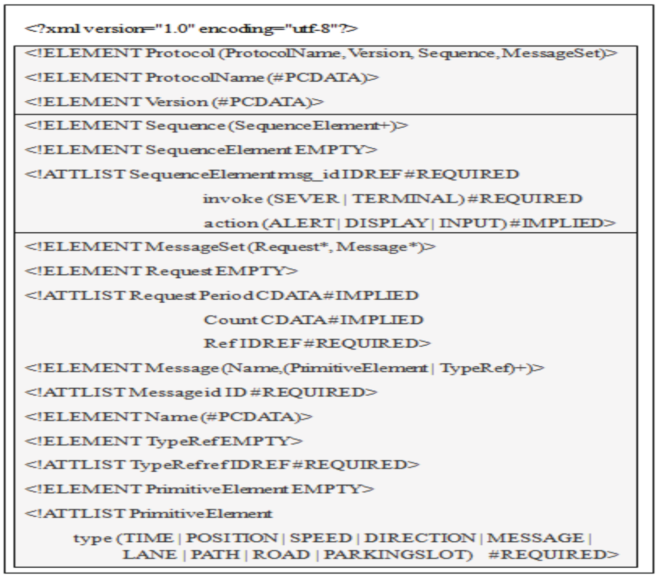


Fig. 2. Protocol Design Scheme

IV. USE CASE DESCRIPTION FORMAT

In this section, we present the use case description format (UDF) in detail. We demonstrate how the client program is able to assemble a message instance by introducing primitive data elements. Fig. 3 shows simplified version of the UDF in our protocol. The UDF has three sections. The first specifies the information about the protocol; the second provides the message sequence format; and the third is the specification of the message set.



```
<?xml version="1.0" encoding="utf-8"?>
<!ELEMENT Protocol (ProtocolName, Version, Sequence, MessageSet)>
<!ELEMENT ProtocolName (#PCDATA)>
<!ELEMENT Version (#PCDATA)>
<!ELEMENT Sequence (SequenceElement+)>
<!ELEMENT SequenceElement EMPTY>
<!ATTLIST SequenceElement msg_id IDREF #REQUIRED
                    invoke (SEVER | TERMINAL) #REQUIRED
                    action (ALERT | DISPLAY | INPUT) #IMPLIED>
<!ELEMENT MessageSet (Request*, Message*)>
<!ELEMENT Request EMPTY>
<!ATTLIST Request Period CDATA #IMPLIED
                    Count CDATA #IMPLIED
                    Ref IDREF #REQUIRED>
<!ELEMENT Message (Name,(PrimitiveElement | TypeRef)+)>
<!ATTLIST Messageid ID #REQUIRED>
<!ELEMENT Name (#PCDATA)>
<!ELEMENT TypeRef EMPTY>
<!ATTLIST TypeRefref IDREF #REQUIRED>
<!ELEMENT PrimitiveElement EMPTY>
<!ATTLIST PrimitiveElement
      type (TIME | POSITION | SPEED | DIRECTION | MESSAGE |
           LANE | PATH | ROAD | PARKINGSLOT)   #REQUIRED>
```

Fig. 3. Use case Description Format

A sequence of messages used in a use case consists of sequence elements. A sequence element has a message to be transmitted, an invoker who is responsible for sending the message, and an action which is optionally specified when the invoker is the server. There are three types of actions that a client plays. ALERT denotes to provide a user with a safety warning. DISPLAY denotes to display road segments and/or parking slots on the screen. INPUT denotes to get an input data from user.

There are two sorts of messages in the message set. One is a message for a request and the other is a message for data items. A request message is transmitted by a R-ITS-S. When a terminal receives a request message, the terminal should fill out contents of the message specified by the 'Ref attribute' and transmit the message to the server periodically. The period of message transmission is specified by the 'Period' attribute and the number of message transmission is specified by the 'Count' attribute of a request message. A message for data item can be transmitted by either a server or a terminal. A message can contain primitive elements and references to other messages. Primitive elements are commonly used data item for configuring a safety warning, a parking guidance messages. Primitive elements include TIME, POSITION, SPEED, DIRECTION, MESSAGE, LANE, PATH, ROAD, and PARKINGSLOT. Formats of primitive elements are pre-defined in our protocol.

Each message specification in a use case instance is a template for constructing a message. The terminal should fill out position and speed information in a binary format and send it to the server. Our protocol includes a pre-defined binary data format for each primitive element and combination rule for them. We implemented our protocol and simulated the protocol with various configurations. (Fig. 4)

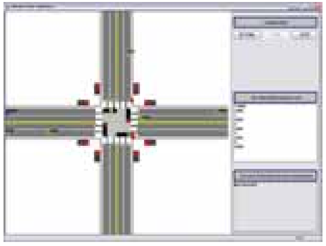


Fig. 4. Protocol Simulation Program

V. CONCLUSIONS

In this letter, we proposed a flexible application protocol for safety warning and parking guidance services. Unlike many other application protocols in ITS and Telematics domains, our protocol makes the client part independent of use cases for supporting light-weighted nomadic devices. For this purpose, we introduce the UDF which enables P-ITS-S to interpret any use case instance formatted in the UDF. We plan to apply our protocol to the real road and parking lot environments.

REFERENCES

[1] Joingik Kim, Oh-Cheon Kwon, and Hyunsuk Kim, "Development of an Event Stream Porcessing System for the Vehicle Telematics Environment," *ETRI Journal*, vol.31, no.4, pp. 423-425, August 2009.

[2] Open Mobile Alliance, "Mobile Location Protocol (MLP)", Enabler Release Definition for Mobile Location Protocol (MLP) Candidate Version 3.1, March 2004.

[3] U.Manni, "Smart sensing and time of arrival based location detection in parking management services," *Proc. Of IEEE Electronics Conference*, pp. 1736-3705, 2010.

[4] M.Wada, K. S. Yoon, and H. Hashimoto, "Development of Advanced Parking Assistance System," *IEEE Transactions on Industrial Electronics*, vol.50, pp. 4-17, 2003.

Received Signal Strength Ratio Based Optical Wireless Indoor Localization Using Light Emitting Diodes for Illumination

Soo-Yong Jung, Chang-Kuk Choi, Sang Hu Heo, Seong Ro Lee, and Chang-Soo Park, *Member, IEEE*

***Abstract*—We propose optical wireless indoor localization using light emitting diodes (LEDs). In the proposed method, four LED lamps for illumination are employed and strength ratio between received LED light signals is utilized for location estimation.**

I. Introduction

Indoor localization has numerous potential applications in indoor robotics, automation system, personal tracking, and location based service (LBS). So far, lots of techniques have been proposed and studied for indoor location sensing [1], [2]. Recently, LED based positioning systems have been proposed keeping pace with the growth of LED illumination industry [3]-[6]. They use existing infrastructure of LED ceiling lamps for affording cost effective indoor localization.

We propose received signal strength ratio based indoor localization under LED ceiling lamps environment. This proposed system uses strength ratio between received signals to obtain distance ratio, and provides accurate location information with low complexity.

II. Location Estimation

The overview of the proposed indoor localization system under LED ceiling lamps environment are depicted in Fig. 1. Because LED can provide not only lighting but also high speed switching, four LED lamps radiate their light in each assigned time slot without human recognition. Using the power ratio between the received signals from each LED lamp, the location of the object can be estimated. The received power can be expressed as [7],

$$
\begin{aligned}
P_R &= H(0)\times P_T \\
&= \frac{n+1}{2\pi d^2} A_R \cos(\theta)\cos^n(\phi) rect\left(\frac{\theta}{FOV}\right) P_T , \qquad (1)
\end{aligned}
$$

where $H(0)$ is channel DC gain, A_R is detector effective area, d is distance between LED and a receiver, n is the mode number of the radiation lobe, P_T is the source power, □ is the angle of irradiance with respect to the transmitter perpendicular axis, θ is the angle of incidence with respect to the receiver perpendicular axis, and FOV is field of view of the receiver.

This research was supported by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency(KOCCA) in the Culture Technology Research & Development Program 2012 and by Priority Research Centers Program through the NRF funded by the Ministry of Education, Science and Technology(2012-0006688).

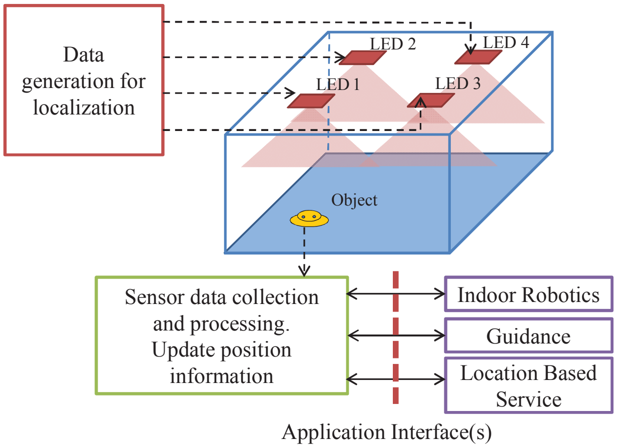

Data generation for localization
LED 1
LED 2
LED 3
LED 4
Object
Sensor data collection and processing. Update position information
Indoor Robotics
Guidance
Location Based Service
Application Interface(s)

Fig. 1. Overview of the proposed indoor localization system.

In this system, we assume perpendicular axes of the LED lamps and the receiver are parallel, and θ is equal to □. Also, the receiver angle, θ, is smaller than FOV at every point of the system. Then, we can rewrite the received power followed by

$$
\begin{aligned}
P_R &\cong \frac{n+1}{2\pi d^2} A_R \cos^{n+1}(\theta) P_T \\
&= \frac{n+1}{2\pi} A_R \frac{h^{n+1}}{d^{n+3}} P_T \qquad (2) \\
&= K \frac{1}{d^{n+3}},
\end{aligned}
$$

where h is height of the room, and K is common value for each LED, expressed as $K = (n+1)A_R h^{n+1} P_T / 2\pi$. As we can see in (2), received power can be expressed as a function of distance. In the received data, each time slot contains the received signal from each LED lamp, and we can easily distinguish the received signals and get strength ratio between each other. We call the received signal strength ratio to be RSSR in this paper. Based on (2), the information of the RSSR from each LED lamp can give the distance ratio. If the distance ratio between two points is given, an equation of a circle or a straight line is obtained. In the proposed system, there are four LED lamps and several equations of a circle or a straight line can be obtained using RSSR.

Consequently, using the relations between LED1 and LED2, between LED1 and LED3, and between LED1 and LED4, three equations can be obtained. The location of the object is a solution of the simultaneous equations.

III. Performance Evaluation

We evaluated the proposed method using computer simulation. The dimension of the system model is 5.0 m 5.0 m 3.0 m, and four LED lamps are located at (1.5, 1.5), (1.5, 3.5), (3.5, 1.5), and (3.5, 3.5) in the ceiling, respectively. Each LED lamp is composed of 100 LED chips, and LED chip interval is 3 cm. To distinguish the light signal power from each LED lamp, time division multiplexing (TDM) technique is used. In case that there exist a number of LEDs, the brightest four LEDs are selected. The LED lamps radiate their light in assigned time slot, that is, [1 0 0 0], [0 1 0 0], [0 0 1 0], and [0 0 0 1] are transmitted by LED1, LED2, LED3, and LED4, respectively, and white Gaussian noise is added to the signal. Fig. 2 shows the graph of three circles obtained at (4 m, 3 m). As we can see in Fig. 2, the crossing point of the circles indicates (4 m, 3 m), exactly same with the location of the object. Each circle was obtained using RSSR between LED1 and LED2, between LED1 and LED3, and between LED1 and LED4, respectively.

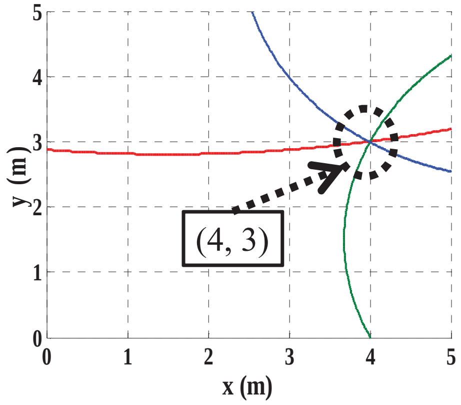


Fig. 2. Obtained circles at (4 m, 3 m) on the floor.

We evaluated the performance of location error at several points on the floor. Fig. 3 shows the result of location error. Among 81 points, the maximum and mean location error was 3.65 cm and 1.12 cm, respectively. The location error is mainly caused by additive white Gaussian noise. Especially, at the corner, the location error is a growing trend. It is caused by a weakening of the signal from a LED lamp in a diagonal position resulting in lower signal-to-noise ratio.

In this paper, we focused on the realization of localization. For final product, additional efforts to calibrate initial position and remove the reflected light from walls are needed. For calibration, a method of identifying the brightest light and then next can be tried but is beyond our current scope. In general, the lights for illumination are located a little far away from walls and thereby the noise from the reflected light can be lowered. Especially, in the case of requiring many lights, selection of four lights far away from walls is good enough to identify the position.

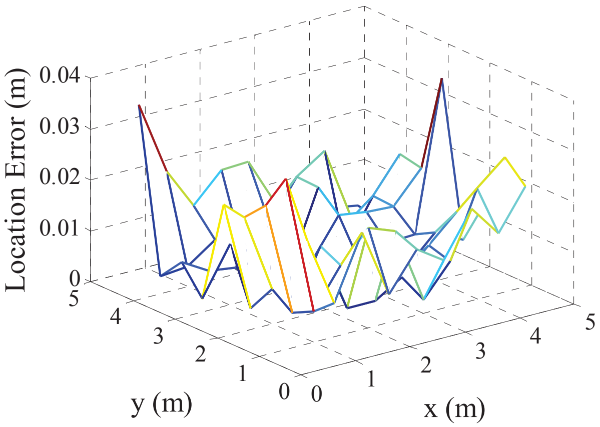


Fig. 3. Location error on the floor *xy*-plane.

IV. Conclusions

We proposed an indoor localization method based on RSSR using LED ceiling lamps. Using LED properties such as lighting and switching, the LED lamps transmit their light signal in assigned time slot while they were utilized for illumination. RSSR was used to get distance ratios and three equations could be obtained using the distance ratios. The target position is a solution of the simultaneous equations. The localization method was successfully performed via computer simulation. This proposed positioning system provides high accuracy, low cost, and low complexity. Thus it possibly represents an effective candidate for future indoor localization under LED ceiling light environments, and can be applied to indoor robotics, smart mobile networking, and LBS.

References

[1] H. Liu, H. Darabi, P. Banergee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybernet.*, vol. 37, no. 6, Nov. 2007.

[2] K. Pahlavan, X. Li, and J. Makela, "Indoor geolocation science and technology," *IEEE Commun. Mag.*, vol. 40, no. 2, Feb. 2002, pp. 112-118.

[3] C. Sertthin, E. Tsuji, M. Nakagawa, S. Kuwano, and K. Watanabe, “A switching estimated receiver position scheme for visible light based indoor positioning system,” *Proc. 4th Int. Conf. Wireless Pervasive Computing*, Melbourne, Australia, Feb. 2009, pp. 64-68.

[4] S.-Y. Jung, S. Hann, and C.-S. Park, “TDOA-based optical wireless indoor localization using LED ceiling lamps,” *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, Nov. 2011, pp. 1592-1596.

[5] S.-Y. Jung, S. Hann, S. Park, and C.-S. Park, “Optical wireless indoor positioning system using light emitting diode ceiling lights,” *Microw. Opt. Technol. Lett.*, vol. 54, no. 7, July 2012, pp. 1622-1626.

[6] J. Vongkulbhisal, B. Chantaramolee, Y. Zhao, and W. S. Mohammed, “A fingerprinting-based indoor localization system using intensity modulation of light emitting diodes,” *Microw. Opt. Technol. Lett.*, vol. 54, no. 5, May 2012, pp. 1218-1227.

[7] J. M. Kahn and J. R. Barry, “Wireless infrared communication,” *Proc. IEEE,* vol. 85, 1997, pp. 265-298.

Wireless Access Control System based on IEEE 802.15.4

G Dhivya, C Sethukkarasi and R Pitchiah
Centre for Development of Advanced Computing (C-DAC), Chennai, India

Abstract-- **Access control systems are the main security mechanisms to control the access of environments. This paper describes about the implementation and deployment of wireless access control system for providing authorized access in a smart home environment. It makes use of ZigBee and image processing technique to control the door lock. ZigBee enabled door lock module has been designed and developed. The image transfer over ZigBee network has been analyzed for different image size and the challenges involved in the face recognition module are discussed.**

I. Introduction

Wireless technologies like RFID (Radio Frequency Identification), UWB (Ultra Wide Band), and ZigBee [3] etc. are used in access control systems. The proposed system (Fig. 1) is a wireless access control system designed and developed for smart home environment. It identifies the user's presence, capture and transfers the image wirelessly, recognizes the user and provides access. Our contributions differ from [3], [4] are as follows.

- Developed a wireless access control application and face recognition module for user authentication
- Implementation of ZigBee enabled door lock module
- Deployment of the proposed system and analysis of image transmission time over ZigBee

II. System Architecture

The system comprises of door and lock module installed on door and a central server that resides inside the home. Two types of users are considered namely 'defined users' (inhabitants of the home) and 'other users' (e.g., guest, friends and neighbors). Other users are those who are all authorized by the defined users. A ZigBee based star network with a ZigBee Coordinator (ZC) and two ZigBee End Device (ZED) has been established. ZC has been connected via USB to the central server. The ZED's are connected to door and lock module.

A. Door Module

It consists of a speaker and a ZED. The ZED comprises of a Video Graphics Array (VGA) camera and a PIR sensor of range 5m. The PIR sensor detects the presence of the moving object and triggers the camera. The captured image is transferred to the ZC. The speaker module interacts with the user through an audio message.

B. Central Server

The central server comprises of a face recognition module, wireless access control module and databases. Database holds the details of defined users, other users and also keeps the record of unauthorized users. The ZC receives the image transmitted by the ZED of the door module and routes it to the face recognition module.

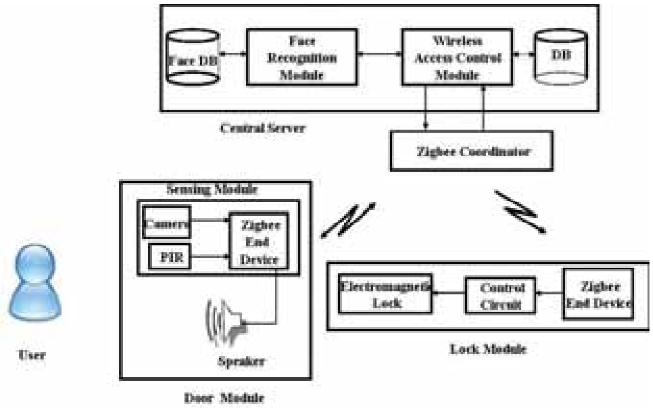


Fig 1. Overall Architecture

1) Face Recognition Module

The steps of face recognition module [5] are shown in Fig.2. Face recognition module uses OpenCV image processing library. The captured image is to be pre-processed in order to improve the image contrast. Next step is face detection for which Viola and Jones algorithm [1] had been used. The captured image is scanned from top left to bottom right by a small window for the presence of HAAR features. The window with HAAR features is sent through a chain of filters called as classifiers. The filters are trained with positive (face images) and negative (non-face image) samples. The window which passes through all the filters is classified as "Face" image else it is a "Non-Face" image. The detected face image has to be cropped and resized. "Eigen Face" Recognition method [2] is used for face recognition. It is difficult to compare the images pixel by pixel; hence Principal Component Analysis (PCA) technique has been used for dimension reduction. The algorithm has been trained with the standard face database images which have been created using the camera in the ZED. The Eigen values and Eigen vectors of images are calculated and projected over a lower dimension space. Mahalanobis distance method is used to find out the

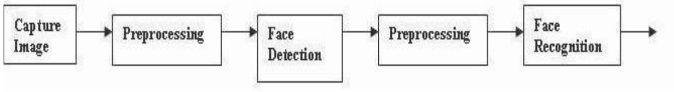


Fig. 2 Steps involved in face recognition module

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

distance between the test image and the training images. The threshold values have to be calculated experimentally and set for the minimum distance measured to classify the known, unknown and non face image. The face recognition module output is sent to the wireless access control module.

2) Wireless Access Control Module

To use the system for the first time, the defined user has to register his/her personal details and upload the face image in the face database through a GUI. Other users face images are also captured and uploaded to the face database. In the absence of the defined users, the system can be informed about the other users who can have access to their home. If new person is trying to access the home, in the presence of defined user, the image of the new person will be displayed in the display device for authentication. The defined user can approve/reject the request. In the absence of defined user, an SMS will be sent to his/her handheld device about the presence of a new person. The captured image of new person with captured time is saved in the database. Based on the output from face recognition module, the wireless access control module generates a command to the ZED of the lock module to open/close the door.

C. Lock Module

Lock module includes a ZED, a control circuit and an electromagnetic lock. A 5v DC circuit has been designed to operate the ZED. The electromagnetic lock has been controlled by the ZED through the control circuit. When the command received is “open”, the control circuit will de-energize the lock to open the door. Else the lock gets energized to close the door. To leave the home, the user has to press the exit button in door frame.

III. DEPLOYMENT RESULTS

The system had been deployed in Ubicomp lab at C-DAC, Chennai (Fig 3). Image transmission time over USB and ZigBee had been measured for different image sizes (Table I). To avoid packet loss during transmission, 20 ms delay was provided between packets. For color QVGA image without compression, the transmission time of 69.546 sec over ZigBee was not acceptable for an access control application. In order to reduce the transmission time, the image needs to be compressed before transmission. With compression, the transmission time was found to be approximately 0.4 sec for a Jpeg color QVGA image. The camera had been mounted at a height of 150 cm to cover the persons of different heights (150 - 175 cm). The distance between user and door module was kept constant as 150 cm to avoid scaling variation. The challenges involved in face recognition algorithm such as illumination changes, scaling variation, face orientation, with/without spectacles and facial expressions are being addressed.

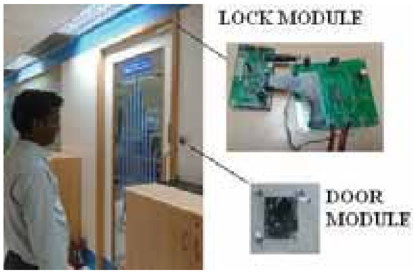


Fig. 3 Deployment in Ubicomp lab at C-DAC Chennai

TABLE I COMPARISON OF IMAGE TRANSMISSION TIME OVER ZIGBEE NETWORK AND USB FOR DIFFERENT IMAGE TYPES

Image Type	Image Size (bytes)	Image Transmission Time (sec)	
		USB	ZigBee
Color QVGA	153600	19.547	69.546
VGA	307200	39.421	140.851
Color VGA	614400	79.359	277.201
JPEG Color QVGA	4032	0.125	0.4265

Training and Recognition time of face recognition algorithm were found to be 5 and 2.3 seconds respectively for 140 training images of 14 different persons. The face detection algorithm had been tested with the CMU face database. False detection (false positive) rate and missing face (false negative) rate were found to be 7.56% and 9.18% respectively.

IV. CONCLUSION AND FUTURE WORK

Proposed system is an automated system that avoids the need of carrying any access card along with, thereby increasing the user’s comfort and convenience. The system had been deployed and image transmission results are discussed. The analysis of transmission time, power consumption, date error rate etc., for different distances, obstacles and other interference sources (Wi-Fi, Bluetooth etc) has been planned for future work. The accuracy of the face recognition algorithm needs to be tested in the deployed environment.

REFERENCES

[1] Paul Viola, and Michael J. Jones, “Rapid object detection using a boosted cascade of simple features”, IEEE Conference on Computer Vision and Pattern Recognition, pp. I-511-I-518, 2001.

[2] Matthew A. Turk, and Alex P. Pentland, “Face recognition using Eigen faces”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586-591, 1991.

[3] Il-Kyu Hwang, and Jin-Wook Baek, “Wireless access monitoring and control System based on digital door lock”, IEEE Transactions on Consumer Electronics, Vol. 53, pp. 1724-1730, Nov 2007

[4] Georgiy Pekhteryev, Zafer Sahinoglu, Philip Orlik and Ghulam Bhatti, “Image transmission over IEEE 802.15.4 and ZigBee networks”, IEEE ISCAS, May 2005, Kobe Japan, vol. 4, pp. 3539-3542

[5] Chidambaram Sethukkarasi, Vijayadharan SuseelaKumari HariKrishnan, Raja Pitchiah, “Design and development of interactive mirror for aware home”, First International Conference on Smart Systems, Devices and Technologies, SMART 2012, IARIA, Stuttgart, pp. 1-8.

Resource Allocation for Cyclic Prefixed Single-Carrier Cognitive Two-Way Relay Networks

Hongwu Liu and Kyung Sup Kwak
School of Information and Communication Engineering, Inha University, Incheon, Korea

***Abstract*—This paper proposes a joint optimization of power allocation and subcarrier pairing for a cyclic prefixed single-carrier cognitive two-way relay network. The optimal frequency-domain linear equalization receiver and the decision-feedback equalization receiver are considered. Karush-Kunh-Tucker conditions and a dynamic ordering greedy algorithm are applied to efficiently solve the joint optimization problem.**

I. Introduction

Cyclic prefixed single-carrier (CP-SC) based cooperative transmission [1], [2] is an attractive candidate for the cognitive radio (CR) relay communications [3], [4], such as those in wireless personal area networks [5] and wireless regional area networks [6]. While the resource allocation for the CP-SC cognitive relay networks is in infancy, the power allocation across subcarriers has never been developed in such networks, neither does the subcarrier pairing (SP). This paper proposes a new resource allocation scheme for a CP-SC cognitive two-way relay network. The optimal frequency-domain linear equalization (FD-LE) and the frequency-domain decision-feedback equalization (FD-DFE) receivers are derived and the minimization of the sum mean squared error (MSE) is adopted to jointly optimize power allocation and SP, subject to a pre-specified limited interference to a licensed primary user (PU).

II. Resource Allocation

The considered CP-SC cognitive two-way relay network consists of two secondary transceivers (STs) T_1, T_2, and one secondary relay (SR) T_3 as depicted in Fig. 1. Two STs accomplish one time of information exchange through the help of the SR within two time phases: the multiple access (MA) phase and the broadcast (BC) phase. In the MA phase, we assume that no channel state information (CSI) is available at two STs, thus uniform power allocation (UPA) is employed at two STs. Whereas the SR extracts CSI from the ST pilot signal to realize the resource allocation and feed back CSI to ST to facilitate the signal detection. In order to maintain a simple operation, no complex equalization or beamforming is employed at the SR except SP and optimal power allocation (OPA) across subcarriers.

In the MA phase, each symbol block to be transmitted by T_i $(i = 1, 2)$ is first scaled with the square root of the UPA factor p_i, which is less than or equal to the peak power P_i. Then, the symbol block will be transmitted out after appending a CP in its front. To prevent inter-block-symbol interference, the length of the CP is assumed to comprise the maximum path delay. The sum interferences introduced by two STs at the PU should not beyond the pre-specified interference threshold.

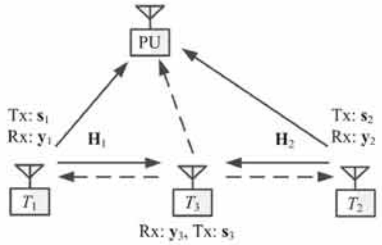


Fig. 1. CP-SC Cognitive two-way relay network.

The received signal at the SR is the composition of the faded signals of T_1 and T_2 plus the additive noise, which will be CP-removed and transformed into the frequency-domain by using IFFT. An normalization coefficient is applied to each subcarrier signal to limit the power to be unit. Then the SP is performed by a row permutation matrix such that the received signal over a particular subcarrier in the MA phase is broadcasted out on an suitable subcarrier in the BC phase. After subcarrier paring, OPA across subcarriers is carried out at the SR subject to the peak power. Also, the interference introduced by the SR at the PU should not beyond the pre-specified interference threshold.

At the end of the BC phase, the CP related signal will be first removed from the received signal at two transceivers, respectively. Then, the self interference can be eliminated from the received CP-removed signal due to the channel reciprocal. The FD-LE and the FD-DFE receivers are applied at two transceivers to detect the information-bearing signal. Since the performance of the FD-LE and the FD-DFE receivers are directly influenced by the MSE at T_1 and T_2, the goal of the resource allocation scheme is to minimize the sum MSE of two transceivers through power allocation and SP, subject to the peak power constraints at two STs and the SR, respectively, and the pre-specified interference thresholds at the PU in the MA phase and the BC phase, respectively.

For the given quality of service (Qos) coefficients at two STs, the optimal UPA factors can be easily determined respect to the peak power constraints at two STs and the pre-specified threshold at the PU in the MA phase. The joint optimization of OPA and SP is a mixed integer programming problem, which is computationally prohibitive even for a moderate number of subcarriers. Since the duality gap between the optimal solution of the mixed integer programming and that of its

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

Algorithm 1 Modified Greedy Algorithm

```
Set the SP matrix $M = 0$, set $\mathcal{P} = \{1, ..., N\}$ and $\mathcal{Q} = \{1, ..., N\}$,
where $\mathcal{P}$ and $\mathcal{Q}$ are the subcarrier sets of the MA phase and the BC
phase, respectively.
Calculate $\mathcal{S} = \{S_k | S_k = \sum_{l=1}^{N} \tilde{f}_{l,k,\mathrm{Rx}}, \forall k \in \mathcal{P}\}$
while $\mathcal{P} \neq \emptyset$ do
    $k^{\circ} = \arg\max_{k \in \mathcal{P}} \{S_k\}$ and $l^{\circ} = \arg\min_{l \in \mathcal{Q}} \{\tilde{f}_{l,k^{\circ},\mathrm{Rx}}\}$
    Set $[M]_{l^{\circ},k^{\circ}} = 1$.
    Remove $k^{\circ}$ and $l^{\circ}$ from $\mathcal{P}$ and $\mathcal{Q}$, respectively.
    if $|\mathcal{P}| \geqslant 2$ then
        Calculate $\mathcal{S} = \{S_k | S_k = S_k - \tilde{f}_{l^{\circ},k,\mathrm{Rx}}, \forall k \in \mathcal{P}\}$
    end if
end while
```

corresponding dual problem approaches zero for sufficiently large number of subcarriers, the solution of the dual problem is asymptotically optimal. By using the Lagrange dual multiplier method, we separate the dual problem into two sub-problems: 1) Optimal power allocation for given subcarrier pair. 2) SP for known power allocation. For any given subcarrier pair, the optimization of OPA becomes a concave problem, the Karush-Kunh-Tucker conditions provides the optimal solutions for FD-LE and FD-DFE receivers. Once an OPA solution is given for each and every subcarrier pair, the optimal SP can be found by the Hungarian algorithm with a complexity of $\mathcal{O}(N^3)$, where N is the number of subcarriers. To implement the SP for a real-time system, a modified greedy algorithm with the complexity of $\mathcal{O}(N^2)$ is devised. After solving two sub-problems, the optimal dual problem is solved with sub-gradient method.

III. Simulation Results

In simulations, the quasi-static channel is adopted and the perfect CSI is assumed to be known at the SR, while the SR feedback the CSI to two STs to detect signal. The number of subcarriers is $N = 32$, while the peak power is set to be $P_i = 32$ for $i = 1, 2, 3$. The average bit error ratio (BER) performance of the proposed resource scheme is investigated with respect to the signal-to-noise ratio (SNR) and the pre-specified interference threshold to the PU, respectively. For simplicity, we assume the SNRs of two hops are same. From the viewpoint of the implementation in a real-time system, only the modified greedy algorithm is considered for the SP. Fig.2, the average BER vs. the SNR with $I_{\mathrm{th}} = 3\mathrm{dB}$ is given out. In Fig. 3, the average BER vs. I_{th} with the $\mathrm{SNR} = 22\mathrm{dB}$ is given out. The simulation results prove that the great average BER performance can be achieved with our proposed scheme.

IV. Conclusion

A resource allocation scheme is proposed for the CP-SC cognitive two-way relay networks. The FD-LD and FD-DFE receiver are devised. The power allocation and the SP are jointly optimized to minimize the sum MSE subject to a pre-specified limited interference to a licensed PU. The mixed integer programming problem is solved with its equivalent Lagrange dual problem, which has been separated into two sub-problems of optimizing power allocation and optimizing SP, respectively. Then the sub-gradient method is applied to solve the optimal dual problem, while a modified greedy algorithms is proposed to efficiently realize the SP.

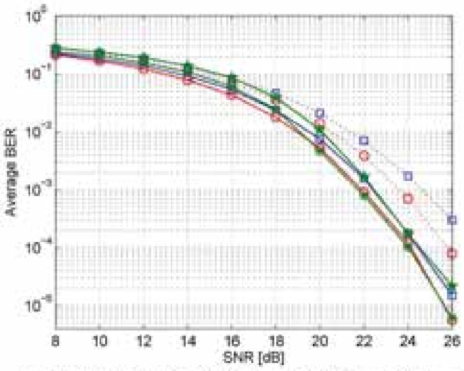


Fig. 2. Average BER performance vs. the SNR with $I_{\mathrm{th}} = -3$ dB.

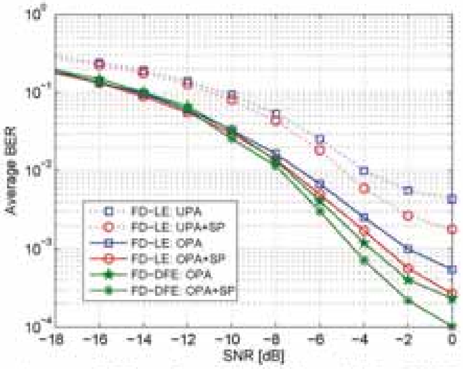


Fig. 3. Average BER performance vs. I_{th} with the SINR = 22 dB.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No.2010-0018116).

References

[1] K. J. Kim, T. A. Tsiftsis, and H. V. Poor, "Power allocation in cyclic prefixed single-carrier relaying systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2297-2305, July 2011.

[2] H. Miyazaki, M. Nakada, T. Obara, F. Adachi, "Adaptive power allocation for bi-directional single-carrier relay using analog network coding," *Proc. International Conference on Information, Communications and Signal Processing*, pp. 1-5, Singapore, Dec. 13-16, 2011.

[3] K. J. Kim, T. Q. Duong, and H. V. Poor, "Performance analysis of cyclic prefixed single-carrier cognitive relay systems," *IEEE Trans. Wireless Commun.*, accepted and to be published.

[4] K. J. Kim, T. Q. Duong, and X. Tran, "Performance analysis of single-carrier systems in cooperative spectrum sharing environment with multiple DF relayings" *IEEE Trans. Signal Prcess.*, accepted and to be published.

[5] IEEE P802.15.3c/D00, "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs): Amendment 2: Millimeter-wave based Alternative Physical Layer Extension," 2008.

[6] IEEE P802.22-D3, "Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 22: Cognitive Wireless RAN Mediu Access Control (MAC) and Physical Layer (PHY) Specification: Policies and Procedures for Operation in TV Bands," March 2011.

Software Design of Giga-bit WLAN on Coarse Grained Reconfigurable Array Processors

Kitaek Bae, Peng Xue, Navneet Basutkar, and Ho Yang
Samsung Advanced Institute of Technology, Samsung Electronics Co., Ltd.
Yongin-si, Gyeonggi-do, 446-712 Korea
{kitaek.bae, peng.xue, navneet.basutkar, and hoyang}@samsung.com

Abstract—**In this paper, we present a software-defined radio (SDR) implementation of the 4x4 MIMO-OFDM baseband receivers on a coarse-grained reconfigurable array (CGRA) processor operating at 1 GHz clock for IEEE 802.11ac, which can support over 1Gbps data rate. However, software implementation of 802.11ac is very challenging because of the increasing computational complexity supporting giga-bit data transmission up to 6.9Gbps. For the software implementation, we focus on two major design issues: the software optimization for CGRA processors and the solution design for the preamble latency requirement. By measuring the computational cycles on the CGRA processor, we show the feasibility of SDR implementation for the 4x4 MIMO receiver of the 802.11ac. The BER is also evaluated to confirm the robustness of fixed point implementation.**

I. Introduction

There have continually been high demands for high data rate and a variety of multimedia services in the wireless environment. To accommodate new radio standards, software-defined radio (SDR) is a promising solution because of its flexility supporting fast prototyping and easy software upgrade. IEEE 802.11ac provides the high data rate up to 6.9Gbps by increasing the bandwidth, the number of antenna, and the order of mudulation [1]. However, this increases the complexity of the system linearly or exponentially, which makes the SDR implementation for 802.11ac challenging. Thus, to develop the SDR for 802.11ac, we need both high performance DSP processor and low complexity DSP software design. There has been the hardware implementation for 802.11ac in [2]; however, SDR implementation for 802.11ac has not been reported yet.

In this paper, we use a coarse grained reconfigurable array (CGRA) processor operating at 1 GHz clock. High performance of computing is achieved by exploiting the data level parallelism (DLP) of 512-bit single instruction multiple data (SIMD) architecture and the instruction level parallelism (ILP). ILP executes multiple instructions simultaneously exploiting the software pipelining of the array architecture.

For the software implementation, we focus on two major design issues; the software optimization for CGRA processors and the solution design of the preamble to meet the latency requirement for the 4x4 MIMO channels. In this work, software for CGRA processor is mainly optimized based on the profiler results, which allows us to easily check the allocation of computational resources in time and space, to resolve the bottleneck of the software design. The preamble preprocessing of the 802.11ac is very tight because it requires intensive computations for the MIMO channel equalization at the last training symbol. Progressive computing for MIMO equalizer matrix is proposed to meet the processing-time latency constraints. To evaluate the proposed approaches, we measure the cycles of the 802.11ac codes mapped on the CGRA processor.

II. Software Design and Optimization

We consider a 4x4 MIMO OFDM system with 80MHz to support 1Gbps data rate. Figure 1 shows the block diagram of the receiver block diagram for 11ac. It mainly consists of three parts: digital front end (DFE), inner modem (IMD), and outer modem (OMD).

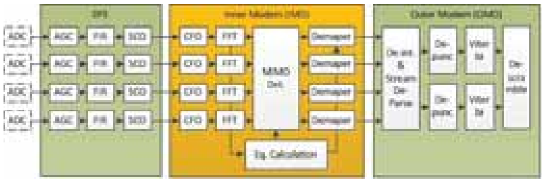


Fig. 1: IEEE 802.11ac 4x4 MIMO receiver block diagram.

Among them, we focus on the SW implementation of IMD in this paper. 802.11ac can achieve high data rate of over 1Gbps by increasing the number of antenna, the order of constellation size, and the channel bandwidth, which leads to high computational complexity. This high computational complexity is a big burden on SW implementation using SDR, because the severe latency constraint in WLAN requires the short processing time for each OFDM symbol. Thus, we propose the systematic approach for software design and optimization to resolve the two major challenging issues: low complexity and tight latency.

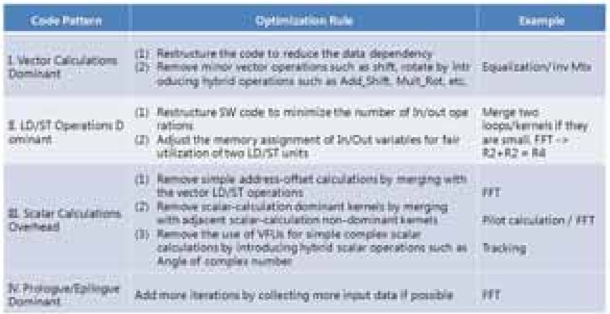
Fig. 2: An example of a profiling result for FFT kernel.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

To resolve the low complexity issue, we systematically design and optimize the major functions. First, the floating point functions are converted to fixed point functions manually and further mapped on the CGRA processor. As shown in figure 2, the profiler tools developed together with CGRA processor show the usage of the instructions in time and space(or location of function units).

For SW optimization, we use systematic approach based on the code pattern, which is summarized in TABLE I. The code optimization is done iteratively to achieve the best possible instruction per cycle (IPC). For some functions such as FFT and matrix inverse, where the best IPC with certain algorithms cannot satisfy the expected throughput, low complexity algorithms should be considered. For example, specific FFT algorithm based on $2^3, 2^2$, and 2^3-mixed radix algorithm [3] is used for 256 point to reduce the traffic of memory access. Contrary to the conventional radix 2 FFT algorithm, the output is calculated in sequential order by using special interleaving intrinsics. Furthermore, the FFT average processing time can be reduced by processing 4 symbols at once, which enhance the software pipelining thus increasing the ILP of CGRA processors. For other functions, we propose the systematic approach for SW optimization based on the code pattern.

TABLE I: Proposed SW optimization rule for major kernels.



Code Pattern	Optimization Rule	Example
I. Vector Calculations Dominant	(1) Restructure the code to reduce the data dependency (2) Remove minor vector operations such as shift, rotate by intr oducing hybrid operations such as Add_Shift, Mult_Rot, etc.	Equalization/ Inv Mtx
II. LD/ST Operations D ominant	(1) Restructure SW code to minimize the number of In/out ope rations (2) Adjust the memory assignment of In/Out variables for fair utilization of two LD/ST units	Merge two loops/kernels if they are small, FFT -> R2+R2 = R4
III. Scalar Calculations Overhead	(1) Remove simple address-offset calculations by merging with the vector LD/ST operations (2) Remove scalar-calculation dominant kernels by merging with adjacent scalar-calculation non-dominant kernels (3) Remove the use of VFUs for simple complex scalar calculations by introducing hybrid scalar operations such as Angle of complex number	FFT Pilot calculation / FFT Tracking
IV. Prologue/Epilogue Dominant	Add more iterations by collecting more input data if possible	FFT

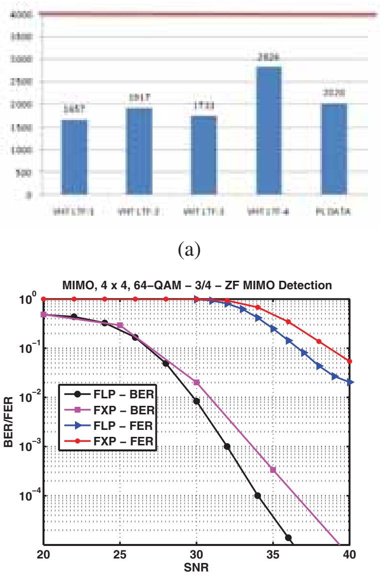
For the latency of the processing time, MIMO equalization calculation requires latency optimization. Channel is generally estimated and then inverted to calculate the equalizer matrix. This leads to computational bottleneck in the last training symbol, which is critical to satisfy the latency constraint. In our implementation, the training symbol matrix is directly inverted, and the operation is distributed to each training symbol as much as possible [4]. This reduces the computations required for the equalizer matrix in the last training symbol. The approach can significantly reduce the pressure on the latency constraint.

III. RESULTS AND CONCLUSION

In this section, we describe the mapping result of 1 Gbps (4x4 MIMO at 80MHz bandwidth) in the 802.11ac WLAN receiver in order to show the feasibility of SW implementation on a CGRA with 1GHz clock frequency. Since 802.11ac has a preamble based frame structure, we consider two main processing: preamble processing and payload processing on a CGRA for mapping. The preamble processing is to calculate the equalization coefficient and LLR coefficient, while the payload processing is to calculate the soft-output to be sent to the channel decoder. In WLAN, the latency of symbol processing time is limited to 4usec. The cycle performance of our implementation for preamble and payload processing is shown in figure 3a, where VHT-LTF 1∼4 is for preamble processing and PL DATA is for payload processing. It is clearly observed that all symbols require the cycle number less than 4000 cycles which corresponds to 4usec to meet the real-time processing. To check the performance degradation from the 16-bit fixed point implementation, we compared the bit error rate (BER) and frame error rate (FER) performance with the results of floating point code.

Figure 3b shows the FER/BER performance under the multipath channel. We observe about 2 dB degradation at 10^{-4} BER and 10^{-1} FER. The performance degradation is mainly due to the 16 bit precision, which is not enough for 4x4 matrix inverse due to having large dynamic data range.

Conclusively, we have demonstrated the feasibility of the software implementation for a 4x4 MIMO OFDM system with a 80MHz bandwidth, which has maximum 1 Gbps data rate. Our software design based on a CGRA processor having 1 GHz clock frequency.



(a)

(b)

Fig. 3: (a) Cycles and (b) BER/FER performance of major functions for 802.11ac receiver.

REFERENCES

[1] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 5: Enhancements for Very High Throughput for Operation in Bands below 6 GHz*, IEEE Std. IEEE P802.11ac/D1.0, 2011.

[2] S. Yoshizawa, D. Nakagawa, N. Miyazaki, T. Kaji, and Y. Miyanaga, “LSI development of 8x8 single-user MIMO-OFDM for IEEE 802.11ac WLANs,” in *International Symposium on Communications and Information Technologies (ISCIT)*, Oct. 2011, pp. 585–588.

[3] P. Westermann and H. Schröder, “On the scalability of SIMD processing for software defined radio algorithms,” in *International Conference on Embedded Computer Systems (SAMOS)*, July 2010, pp. 309–317.

[4] P. Xue, K. Bae, K. Kim, and H. Yang, “Progressive Equalizer Matrix Calculation using QR Decomposition in MIMO-OFDM Systems,” in *IEEE CCNC*, Jan. 2013.

Rotation Multiple-Channel Allocation Scheme for Seamless Handoff in IEEE 802.11 WLANs

Youchan Jeon, Myeongyu Kim, Sangwon Park, and Jinwoo Park

***Abstract*--IEEE 802.11 Wireless LANs can provide broadband wireless data services without dealing with a cable in a limited area. However, when a MS moves to a neighboring AP, it has significant constraints on seamless mobile services. In this paper, we propose a rotational multiple-channel allocation scheme to manage mobility between APs that are equipped with multiple wireless interfaces, prohibiting from using the same channel as neighboring APs simultaneously. Performance evaluations show that the proposed scheme can achieve low handoff delay.**

I. INTRODUCTION

Wireless LANs based on the IEEE 802.11 standards have watched remarkable growth over the last few years [1]. However, none of 802.11 standards have not considered seamless mobile services in medium or large-scale Wi-Fi networks for the service areas within less than a few kilometers in radius, such as a company, campus, or small residential areas which are in general managed by a private network operator. To cover the areas, plural access points (APs) need to be deployed. When a mobile station (MS) moves between APs, it needs to perform the handoff process [2]. The handoff latency results in packet losses or disconnection due to sequential phases such as detection, channel scanning, authentication and reassociation. Consequently, it is difficult to accept delay-sensitive applications such as Voice over IP and real-time video conferences in IEEE 802.11 WLANs.

We propose a rotational multiple-channel allocation scheme to support seamless mobility by modifying APs, by not MSs. In the proposed scheme, APs have multiple wireless interfaces. APs are responsible for mobility of MSs to provide fast handoff. In other words, unlike virtualized wireless LAN proposed by Meru networks Inc., each AP is directly in charge of packet transmission in its service area, buffering packets like the conventional WLAN, and a representative AP called Master AP (MAP) manages authentication, association, and transmission duration (TD) for each AP [3]. Therefore, the proposed scheme does not require a powerful controller and all APs communicate with MSs independently of each other after TDs for APs are set by MAP. It may be effective for especially very high throughput networks like 802.11n or 802.11ac because high amounts of traffic can be exchanged. Thus, the MSs never perform any handoff process specified in IEEE 802.11 standards when MSs move between APs.

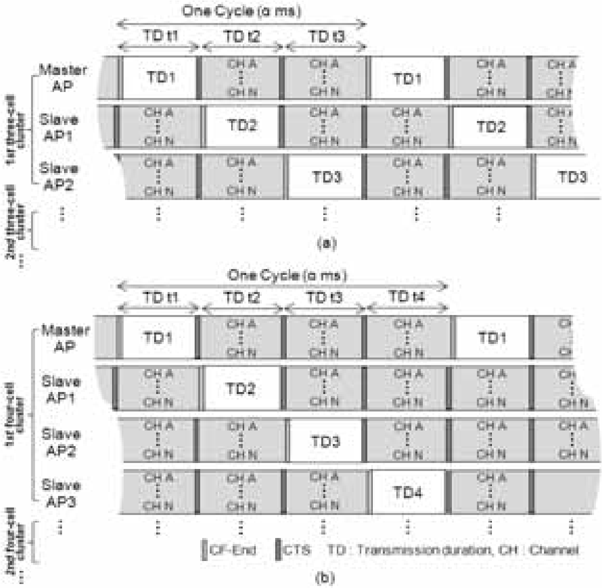


Fig. 1. Rotational multiple-channel allocation between APs in (a) three-cell clusters and (b) four-cell clusters

II. PROPOSED SCHEME

A. Operation

In the proposed scheme, APs use *N* non-overlapping channels simultaneously, having *N* wireless interfaces. All APs and interfaces are set to the same medium access control (MAC) address and basic service set identity (BSSID). By setting the same BSSID, a MS believes to interact with the same AP although the MS moves to different APs. The proposed scheme is comprised of a Master AP (MAP) and plural slave APs (SAPs). MAP manages authentication, association, and transmission duration (TD) for slave APs (SAP). Additionally, both MAP and SAPs are directly in charge of packet transmission in their own service areas during the allocated TD. SAPs are synchronized by the MAP, using precision time protocol for high accuracy [4]. SAPs have list1 and list2 which include MSs in their own areas and MSs in list1 of neighboring APs which do not communicate over the same TD, respectively, and MAP has an extra list for all associated MSs as well as list1 and list2.

Fig. 1 shows a rotational multiple-channel allocation scheme between APs in three-cell clusters and four-cell clusters. When an AP simultaneously uses *N* non-overlapping channels in rotation, neighboring APs stop communicating to avoid interference. One cycle can have three or four TDs. According to the number of available channels and interfaces, the number of TDs should be decided. For example, when each AP has three interfaces, a three-cell cluster is configured having three

This research was partly supported by the KCC (Korea Communications Commission), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2012-08-911-05-001) and partly by the BLS project from the Seoul Metropolitan City [WR080951, Seoul R&BD Program].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

TABLE I. SIMULATION PARAMETERS

Parameter	Values
Mac overhead	224 bits
PHY header	192 bits
ACK	112 bits + PHY header
CTS	112 bits + PHY header
CF-End	160 bits + PHY header
Probe request	544 bits + PHY header
Probe response	656 bits + PHY header
SIFS	10 μs (11b), 16 μs (11a)
DIFS	50 μs (11b), 34 μs (11a)
Slot time	20 μs (11b), 9 μs (11a)
Propagation delay	2 μs
CWmin	32 (11b), 16 (11a)
Basic rate for PHY header	1 Mbps (11b), 6 Mbps (11a)

TDs. A contention free (CF)-End frame announces the beginning of TD and indicates the termination of TD by including the MAC address of MAP in receiver address (RA). When MSs in an AP receive a CF-End frame, they start to communicate by setting NAV to 0. On the other hand, when the MSs receive a CTS frame, they set NAV to maximum value. The size of TD can be adaptively adjusted by the MAP that collects the distribution of traffic loads and the number of MSs from SAPs.

B. *Handoff*

To describe the handoff process for the proposed scheme, we assume that a MS is interacting with SAP1 through channel A. If the MS moves into the overlapped area between SAP1 and SAP2 in Fig. 1(a), SAP2 will also receive the data or ACK from the MS through an interface using the same channel although SAP2 is idle over TD t2. Then, SAP2 checks whether the MS is in its own list1 or list2. If SAP2 recognizes that the MS is in its list2, it reports to MAP. MAP checks if the MS is included in the extra list, and then requests the received signal strength indicator (RSSI) of the MS to SAP1 and SAP2. Since then, SAP1 and SAP2 send RSSI of the MS whenever they receive data or ACK from the MS. Over TD t3, SAP2 does not send any frames to the MS, but can only send ACK after receiving data frames from the MS because the MS is in list2. At this time, SAP1 is idle. Consequently, the MS can communicate with either SAP1 in uplink and downlink over TD t2, and SAP2 in only uplink over TD t3 through channel A. Over TD t1, the MS communicates with neither SAP1 nor SAP2, and repeats the same procedure described above in the following TDs. When RSSI value from SAP2 is enough, MAP sends list update massages to SAP2, SAP1 and neighboring APs of SAP2. Then, SAP2, SAP1 and new neighboring APs involve the MS in list1, list2 and list2, respectively, and old neighboring APs remove the MS from list2.

III. PERFORMANCE EVALUATION

For performance evaluation, we compare the proposed scheme with DeuceScan and FastScan proposed in [5] and [6]. Performance comparisons are based on the saturation throughput. We assume that MSs are uniformly distributed over every channel and AP, one cycle is set to 60 ms. The other operational parameters follow table I.

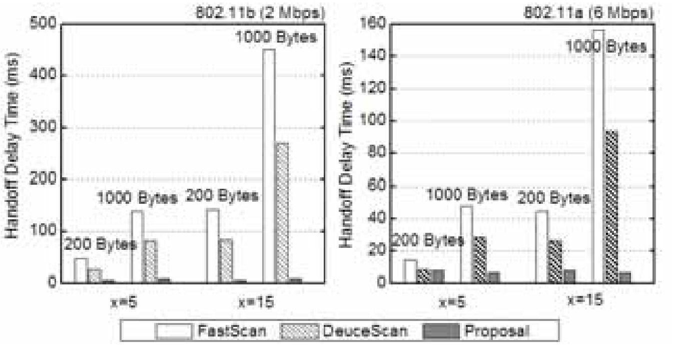


Fig. 2. Comparison of handoff delay time

Fig. 2 shows the handoff delay time. Data rate is 2 Mbps for 802.11b and 6 Mbps for 802.11a, payload size is 200 bytes and 1000 bytes, and the number of MSs (x) per an AP is 5 and 15, respectively. In both DeuceScan and FastScan, the handoff time increases according to the number of MSs and payload size because a MS transmits frames required for channel scanning, authentication and reassociation through competition with other MSs. DeuceScan for prescan is superior to FastScan which needs to scan at least two channels. However, the proposed scheme can acheive the low handoff delay time. It is because both the number of MSs and payload size do not impact greatly on the handoff delay time.

IV. CONCLUSION

This paper proposes a rotational multiple-channel allocation scheme to support seamless mobility without imposing any requirements on the MS side. Based on the transmission duration concept, APs use the wireless channels in rotation to avoid interference with neighboring ones. Consequently, MSs does not need to perform the conventional handoff process. Performance evaluations show that the proposed scheme can provide low handoff delay almost irrespective of the number of MSs, payload sizes and data rates.

REFERENCES

[1] E. Perahia, and R. Stacey, *Next Generation Wireless LANs*, Cambridge University Press, 2008.

[2] A. Mishra, M. Shin, and W. A. Arbaugh, “An empirical analysis of the IEEE 802.11 MAC layer handoff process,” *ACM Computer Communications Review*, Apr. 2003.

[3] Meru Networks. Available: http://www.merunetworks.com

[4] IEEE 1588, “IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems,” *IEEE Std. 1588*, July 2008.

[5] Y. S. Chen, M. C. Chuang, and C. K. Chen, “DeuceScan: Deuce-Based Fast Handoff Scheduling in IEEE 802.11 Wireless Networks,” *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 1126-1141, Mar. 2008.

[6] I. Purushothaman and S. Roy, “FastScan: A Handoff Scheme for Voice over IEEE 802.11 WLANs,” *Wirel. Netw.*, vol. 16, no. 7, pp. 2049-2063, Mar. 2010.

TV-centric Gaming Applications for Android OS: Architecture and a Framework

Milan Z. Bjelica[1], *Member, IEEE,* Vladan Zdravkovic[2], Marija Punt[3] and Nikola Teslic[4], *Member, IEEE*
[1]Faculty of Technical Sciences, University of Novi Sad, Serbia, [2]Sheffield Hallam University, Sheffield, United Kingdom, [3]School of Electrical Engineering, University of Belgrade, Serbia, [4]RT-RK Institute for Computer-Based Systems, Novi Sad, Serbia

***Abstract*—The proliferation of devices with Android OS recently facilitated the integration of various consumer electronic devices running Android for novel applications. In this paper the architecture and a framework for the development of TV-centric games is presented. These games involve set-top boxes or TV receivers, mobile devices and the Internet towards the creation of an innovative gameplay. The paper discusses the benefits of TV-centric gaming concept, presents several developed game prototypes and gives the first results of user survey with regard to the usability of the concept.**

I. INTRODUCTION

Integration among consumer electronics devices is an always growing trend. With the increase of the overall devices' performance, emerged the possibilities for virtualization of the software stacks used for applications development. Android OS is an open platform for the development of applications running on a vast variety of today's smartphones, tablet PC-s and most recently, set-top boxes and TV receivers. Gaming industry, as one of the most prospective for consumer electronics, may benefit from this OS unification. This paper gives details of a novel gaming concept, in which games are designed to run on top of a platform consisting of a TV receiver and various mobile devices.

Well-known gaming concepts consider the use of computer or a gaming console attached to a screen or projector, connected to a local network or the Internet. Multiplayer mode involves multiple players each residing in his/her own home, connected via Internet. Alternatively, if players are located in the same room playing via a local network, the physical gaming infrastructure gets increasingly complex in order to get sufficient usability. Mobile gaming facilitates the connection efforts, yet, gameplay suffers from limitations such as an inadequate screen size and burdensome social interactivity given the player's eyes and hands are locked to a single I/O medium. Overall, digital games of today suffer from the lack of face-to-face social momentum that even the old-fashioned tabletop games used to have.

This paper presents a concept in which the nature of tabletop gaming is brought back to users in digital form. The concept is extending the typical living room scenario, in which people gather around a big-screen TV. Assuming that each TV viewer in the room possesses a mobile device, the gaming concept considers a scenario in which the main game content is shown on the TV (racing track, card decks, board etc), whereas mobile devices are used as controllers (using touch screen and gyro) and outlets for private portion of the game (e.g. fuel level and gauges, private item inventory, cards in the hand). Android OS, as a widespread platform, now spanning from mobile devices to broadband TVs (based on GoogleTV), appears ideal for the aforementioned concept realization. Recent advances in equipping standard broadcast STBs with Android [1] allow the games to benefit from broadcast-related data, such as the EPG. Access to Internet allows the game to integrate the social circle of the living room to the society online, by interacting with social networks (Facebook, Twitter etc). This way a TV-centric gaming application is given birth, with the concept given in Figure 1.

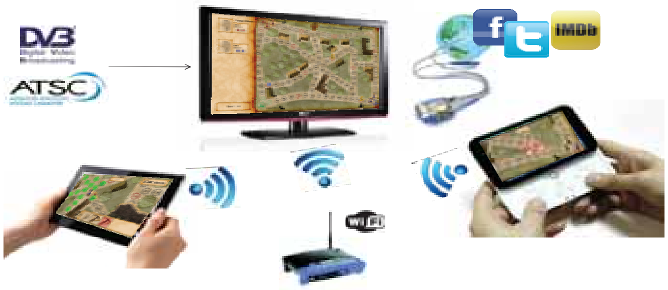


Fig. 1. TV-centric gaming architecture, involving the integration of mobile devices and TV via the local network, using broadcast-related data and communication with the online society.

II. RELATED WORK

Recent researches, most similar to the proposed concepts are mostly related to tangible, mixed-reality gaming experiences [2]. For example, first electronic chess boards were combination of "ordinary" chess table that captured physical motion of the player and responded with position that the player moved (or in the extreme case figures moved automatically). Weathergods is a tabletop game with the digital on-screen board on top of which tangible wooden pawns are placed [3]. Providing private content in such games by the addition of mobile electronic devices is also proposed [4]. Poppet system, as defined in [5], allows the integration of mobile phones as game controllers for the games running on large public displays. Social integration and TV viewing were evaluated with positive user feedback to this practice [6]. Social TV application for Android, allowing TV viewers to communicate to their TV watching peers is presented in [7]. Although many concepts are investigated, to the best of our knowledge there are no works which consider an integrated environment of mobile devices and TVs running Android, with the access to both broadcast and broadband services.

This work was partially supported by the Ministry of Education and Science of the Republic of Serbia, project No. 32041, year 2011.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

III. TV-CENTRIC GAME DEVELOPMENT FRAMEWORK

The developed framework is an Android Java package that encapsulates TV-centric concept providing the following classes: (1) multiplayer networking (*GameHost*, *GameClient*); (2) tabletop games rapid development (*TabletopGame*, *CardGame*, *BoardGame*); (3) Broadcast (*Comedia* DTV API); (4) Social networks integration (*Twitter*, *Facebook*, *IMDB*). Usage of the framework allows an express development of TV-centric games that use data from broadcast streams, post info to social networks and provide an immersive multiplayer experience. Rapid development of tabletop games is fully supported, being that most of the logic and even UI blocks are already provided within the framework, making it possible to develop a simple TV-centric tabletop game in a couple of days. Block diagram of the framework and its position within the Android software stack is shown in Figure 2.

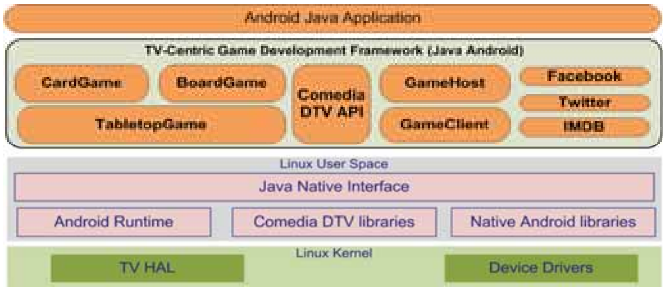


Fig. 2. TV-centric game development framework block diagram

IV. DEVELOPED GAMES AND USER SURVEY

Three games were developed using the TV-centric game development framework and a user survey was conducted with the goal to assess usability of the concept as opposed to other gaming concepts and with the traditional tabletop gaming.

The first game was a tabletop board game with a medieval thematic. Up to four players act as enemies moving their pawns across the board with the goal to destroy one another. The board is shown on a TV screen, whereas mobile devices are used for pawn control, rolling the dice and deploying items (sward, crossbow, land mine etc) which were gathered by users and stored in their private inventory, visible only from a mobile device (Figure 3).

The second game is a traditional Crazy Eights card game. Contents of the table with a deck of cards and player scores are shown on the TV screen, whereas cards in each of the players' hands are placed on the mobile device. Users swipe out cards to the "table" while they are playing with the goal to use up all their cards as soon as possible.

The third game is a micro game intended for playing during the course of TV broadcast. For example, users might get bored by a TV show or a set of commercials. They start the apps on their mobile phones and the game starts – flies start buzzing across the screen. Players use their phones as swatters to kill as much flies as possible within a given time. Accelerometer is used to detect swatting gestures. When the game is complete, the winner is announced to the social networking page of the game, stating names of players, scores and the TV context (for example, *Joe Humiliated Mike (12:2) while watching Show A on TV Station B*).

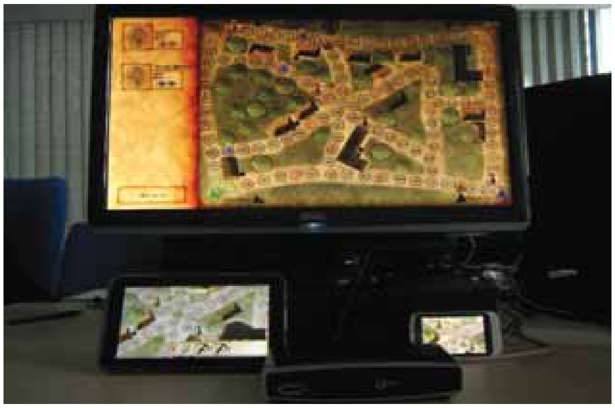


Fig. 3. Developed medieval board game (experimental setup)

Results of user survey are given in Table I. Questions answered were used to assess the experience while playing our games, equivalent tabletop games played online and traditional tabletop gaming, on the Likert scale 1-10.

TABLE I
RESULTS OF USER SURVEY

Game	TV-centric	Online	Traditional
Medieval board game	8.13	5.77	8.69
Crazy Eights	7.9	5.3	8.1
Fly Swatter	9.2	-	-

V. CONCLUSION

This paper presented a concept of TV-centric applications for Android OS, a framework for the development of such applications, three games developed with the framework and a user survey with the goal to assess the usability of a novel concept. Scores given by users testify that the approach taken is refreshing and promising, and that the social face-to-face dimension was brought back to digital games. High rating was also given to a micro game which utilizes broadcast information. This is an indication of potential application of these games for broadcasters' advertising purposes.

REFERENCES

[1] M. Vidakovic, N. Teslic, T. Maruna and V. Mihic, "Android4TV: a Proposition for Integration of DTV in Android Devices, " *IEEE International conference on Consumer Electronics,* 2012.

[2] Regan L. Mandryk and Diego S. Maranan, "False prophets: exploring hybrid board/video games", *CHI EA '02*, pp. 640-641, April 2002.

[3] S. Bakker, D. Vorstenbosch, and E. van den Hoven, "Weathergods: tangible interaction in a digital tabletop game," *Proc. of Tangible and Embedded Interaction 07,* 2007, pp. 151-152.

[4] S. Masanori, H. Kazuhiro, H. Hiromichi, "Caretta: a system for supporting face-to-face collaboration by integrating personal spaces", *CHI EA '04*, pp. 41-48, April 2004.

[5] T. Vajk, P. Coulton, W. Bamford and R. Edwards, "Using a Mobile Phone as a 'Wii-like' Controller for Playing Games on a Large Public Display," *Int. Journal of Comp. Games Tech.,* vol. 2008, 2008, pp. 1-6.

[6] N. Mukhesh, C. Harrison, S. Yarosh, L. Terveen, L. Stead and B. Amento, "CollaboraTV: making television viewing social again, " *Proc. of the UXTV'08,* 2008*,* pp. 85-94.

[7] P. N. Akmar, R. S. Ganesh and R. Sane, "Mobile Based Social TV Application on Android Operating System," *Advances in Computer Science and its Applications,* vol. 1, no. 2, 2012, pp. 97-103.

Sound-based Real-Time Context Recognition on Smartphone

Heeyoul Choi, Sunjae Lee, Jaemo Sung, Sangdo Park
Intelligent Computing Lab
Samsung Advanced Institute of Technology
San 14, Nongseo, Yongin, Gyeonggi, Korea 446-712
{heeyoul.choi, sunjae79.lee, jaemo.sung, sdpark}@samsung.com

***Abstract*—Recently, many people bring their smartphone almost all the time, so the smartphone is considered as a proper device for context recognition. However, the computing power of smartphone is relatively limited to analyze rich contextual cues. In this paper, we present a simple method for smartphone to recognize the context based on sound signals.**

I. Introduction

Due to the advance of technology, smartphone is getting more popular, and many people keep their smartphone with them all the time. So, smartphone is considered as a good device for context recognition. Most of context recognition methods with smartphone have been based on simple cues such as location, user-identity and time information.

However, as people use sound signals to understand the context around them, environmental sound signals are also rich contextual cues [1]. Recently, sound-based context recognition has been proposed [2], [3]. They extract computationally complex features such as Mel-frequency cepstral coefficients (MFCCs), linear predictive coefficients (LPC), RASTA analysis, and power spectral density (PSD). For classification, they adopted mainly hidden Markov models (HMMs). But, those methods run on server, not on smartphone at a real-tiime as a background program.

In this paper, we present a simple approach for sound-based context recognition so that it can be run on smartphone at a real-time as a background program. On a smartphone, sound signals are transformed into frequency domain, where features are frequency bins, by fast Fourier transform (FFT). Then, the dimension of features are reduced by linear methods such as linear discriminant analysis (LDA) and nonnegative matrix factorization (NMF), and the features are classified by a simple classifier like support vector machines (SVMs), and decision tree (DT). We run this process at a real-time on smartphone for 5 contexs: ‘bus’, ‘subway’, ‘train’, ‘street’, and ‘office’. Our experimental results confirm our approach.

II. Data Acquisition

The sound signals were recorded in several places in and around Seoul city with smartphones[1], and the sound format was raw 11025 Hz 16 bit mono PCM. Figure 1 shows example scenes of 5 contexts: ‘bus’, ‘subway’, ‘train’, ‘street’ and ‘office’. Note that the data was gethered while the user was walking around the places holding the smartphone. The data is consist of 3748 signals for training process, each of which contains a sound signal for the time unit, around 3 seconds. Each context has around 700 signals.

Fig. 1. Example photos for ‘bus’, ‘subway’, ‘train’, ‘street’ and ‘office’ contexts where we recorded and tested the sound signals.

III. Method

Figure 2 provides an overview of the approach. After gethering sound signals with smartphone microphones for training, the training process is done on a PC server. The actual context recognition can be conducted on a smartphone at a real-time, using the models obtained from the learning process.

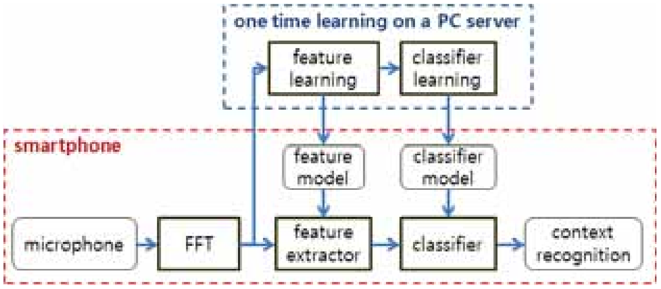


Fig. 2. Overview of the approach. Once the training process is done in MATLAB on a PC server with a training data set, the actual context recognition is conducted in Java on a smartphone at a real-time.

First, sound signals are transformed to a frequency domain via FFT, which can capture some temporal information. We assume that in environmental noise sound, most temporal patterns can be detected within the time unit. Since FFT takes the temporal information for the time unit, the effect of classifiers based on temporal models like HMM is not significant, compared to simple classifiers like SVMs. In our model, the original frequency bin number is 16384 to capture

[1] Samsung Galaxy S2 and Samsung Nexus S

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

all the details in sound. We normalized the result so that the norm of frequency bins is 1.

Since the FFT results are too high dimensional for a classifier to learn on it, we reduce the dimension to find a few dimensional feature representation, using LDA on the frequency bins with training data samples. Before LDA, we combined the close 64 bins which reduced 16384 into 256 bins. Figure 3 shows the averaged FFT results for each context. Note that the contexts are distinguishable when averaged.

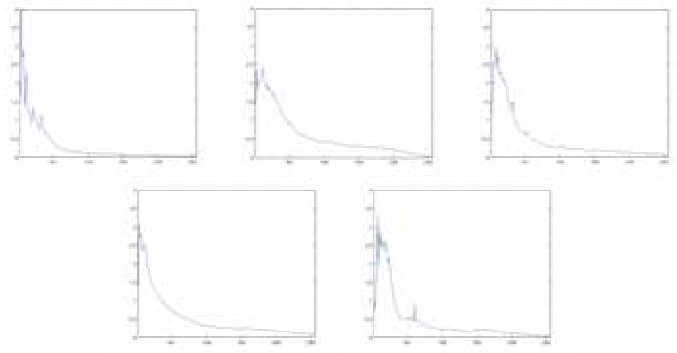


Fig. 3. Averages of FFT results for each context. From the top left: 'bus', 'subway', 'train', 'street' and 'office' contexts, with x and y axes for frequency and power, respectively.

Using the class information, LDA finds the most separable subspace by minimizing the within-scatter and maximizing the between-class on the projected space. Figure 4 shows 4 features on the projected spaces for training data samples in a sequential way. We can see that each context has different sound features from others. Note that since we have 5 contexts, LDA finds 4 features at maximum. Other (semi-) supervised linear projection methods like orthogonal semi-supervised NMF (OSSNMF), would be alternative choises.

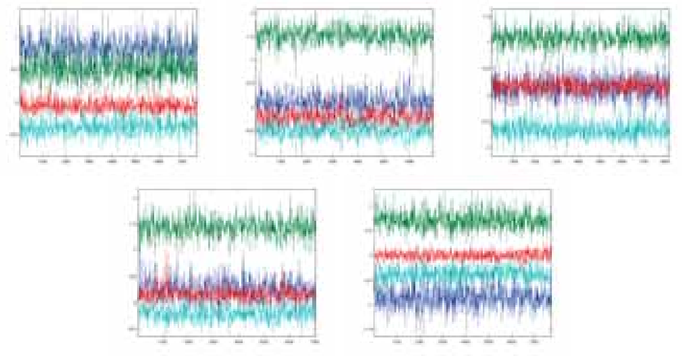


Fig. 4. LDA results for each context. From the top left: 'bus', 'subway', 'train', 'street' and 'office' contexts. The same color across the figures means the same feature.

Figure 5 shows the projected features of the 5 contexts on the same 4 dimensional space in 2 sub-figures. Note that contexts are well clustered, especially the 'bus' and 'office' contexts. Also note that 'street' and 'subway' overlap each other more than others.

As a classifier, SVM and DT were tested. They need a model which can be learned on server and saved on the smartphone. Although SVM provides more accurate performance, there is no significant difference on the performance and runing time.

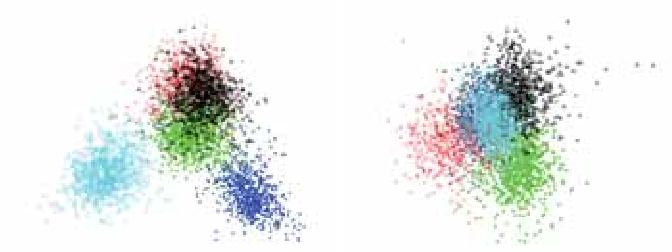


Fig. 5. LDA Projections of training data. (Left) the first 2 features and (Right) the other 2 features. The blue dots are for 'bus', red crosses for 'subway', green stars for 'train', black triangles for 'street', and cyan circles for 'office'.

IV. Test Results

With 413 test sound signals for the 5 contexts recorded at the same places as the training data, we applied FFT, LDA and SVM to classify the contexts. The total accuracy was around 90%. Figure 6 shows the confusion matrix of the accuracy. As expected from the training data projection in Figure 5, 'bus' and 'office' are almost clearly separable from other contexts, while 'subway' and 'street' are misclassfied to each other relatively many times. In contrast to other contexts, 'subway' and 'street' have many similar sound sources like speech, footsteps and music sounds, so they seem not to have their own unique features as many as others.

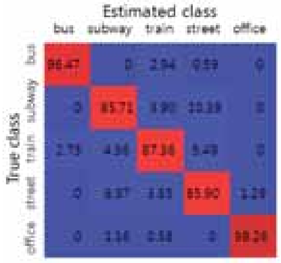


Fig. 6. Confusion matrix to measure accuracy (%)

V. Conclusion

We presented a simple method for smartphone to recognize the context based on sound signals. It could be conducted on a smartphone at a real-time as a background program, contrary to the previous methods, while the performance was similar to the previous methods. In future, we can increase the size of data set for various situations, and combine other modalities such as accelerometer data to improve the accuracy.

References

[1] L. Ma, D. J. Smith, and B. P. Milner, "Context awareness using environmental noise classification," in *Proc. Eurospeech*, Geneva, Switzerland, 2003, pp. 2237–2240.

[2] A. J. Eronen, V. T. Peltonen, and J. T. Tuomi, "Audio-based context recognition," *IEEE Trans. Audio Speech Language Process*, vol. 14, no. 1, pp. 321–329, 2006.

[3] Z. Zeng, X. Li, X. Ma, and Q. Ji, "Adaptive context recognition based on audio signal," in *Proc. Int'l Conf. Pattern Recognition*, Tempa, FL, 2008.

Level-of-Interest Estimation for Personalized TV Program Recommendation

Simon CLIPPINGDALE, Makoto OKUDA, Masaki TAKAHASHI,
Masahide NAEMURA and Mahito FUJII
NHK (Japan Broadcasting Corporation) Science & Technology Research Labs, Tokyo, Japan

***Abstract*—We describe a prototype system that analyzes video from a camera mounted on a TV receiver or set-top-box, showing viewers watching the TV. The system recognizes the faces of registered viewers, and estimates the level of interest that each viewer displays in the program being viewed. This information, along with receiver operation history, can be used to build viewer profiles and to offer personalized program recommendations reflecting each viewer's perceived interests.**

I. INTRODUCTION

Personalization of the user interface, and the provision of targeted advertising, are key tools for boosting page views and revenues on the Internet. The user also benefits through being offered a list of recommendations, e.g. for what video clip to watch next, that are tailored to his/her preferences. To date, however, TV receivers have not been equipped with personalized interfaces because they are not single-user devices (there is usually no login process, and the typical TV viewing model has multiple viewers watching together).

If TV viewers can be identified, however, there is the prospect of building up a profile for each viewer in the same way as a user's web browsing history can be compiled and used to personalize recommended content. In this summary we describe a prototype system that identifies viewers and estimates how interested each viewer is in the program. Viewers who appear to be watching with interest can have the fact recorded in their profiles, while viewers who display little or no interest can have their profiles not updated, or updated with a negative association with the current program.

Section II describes the architecture of the system and the data flow between its various components, and Section III concludes with some remarks about implementation and future directions.

II. ARCHITECTURE

The architecture of the system is shown in Figure 1. Input devices are (i) a camera mounted on or near the TV display, and (ii) a hand-held tablet device that serves as a remote control and as the display component of the user interface. This tablet display offers recommendations for programs related to the program being viewed, or for alternatives thought to be more in line with the user's tastes.

A. Face Module

The Face Module is a face detection, tracking and recognition system based on a prototype developed for automatic video indexing [1]. Skin color regions in the input frame are detected in order to reduce the image area scanned by the boosted cascade face detector [2][3] that follows, which initializes tracking and recognition based on deformable template matching against a number of face templates in the system database. The Face Module outputs, for each face region in the input frame, a recognition result (person ID); an estimate of head pose (left-right and up-down); and the locations of several feature points. It also passes segmented face image regions to other modules that deal with facial expression and action.

B. Expression and Action Modules

The Expression Module is a "bag-of-visual-words" system that extracts SURF features [4] from face images and uses a support vector machine (SVM) [5] to assess how far the expression deviates from neutral. The SVM is trained on images from the CMU Cohn-Kanade database [6]. The Action Module computes an index of upper body movement based on the gross motion of the face region.

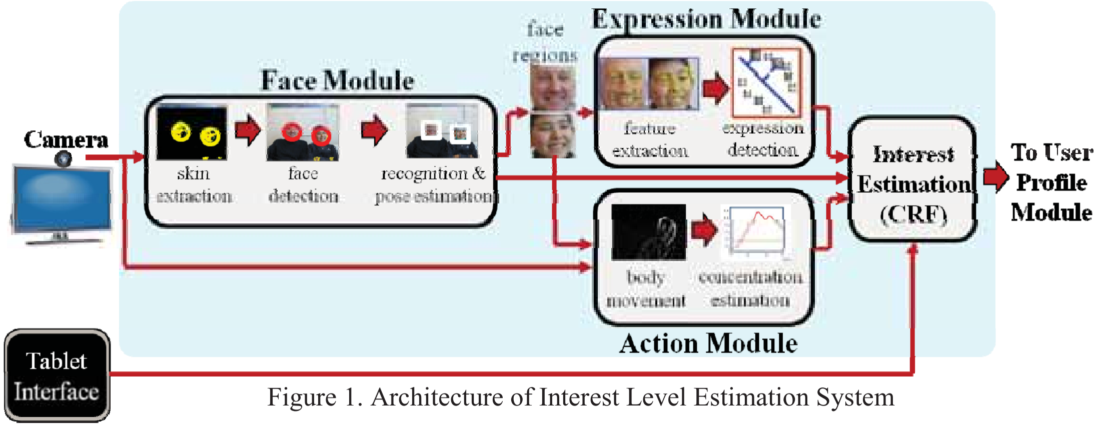


Figure 1. Architecture of Interest Level Estimation System

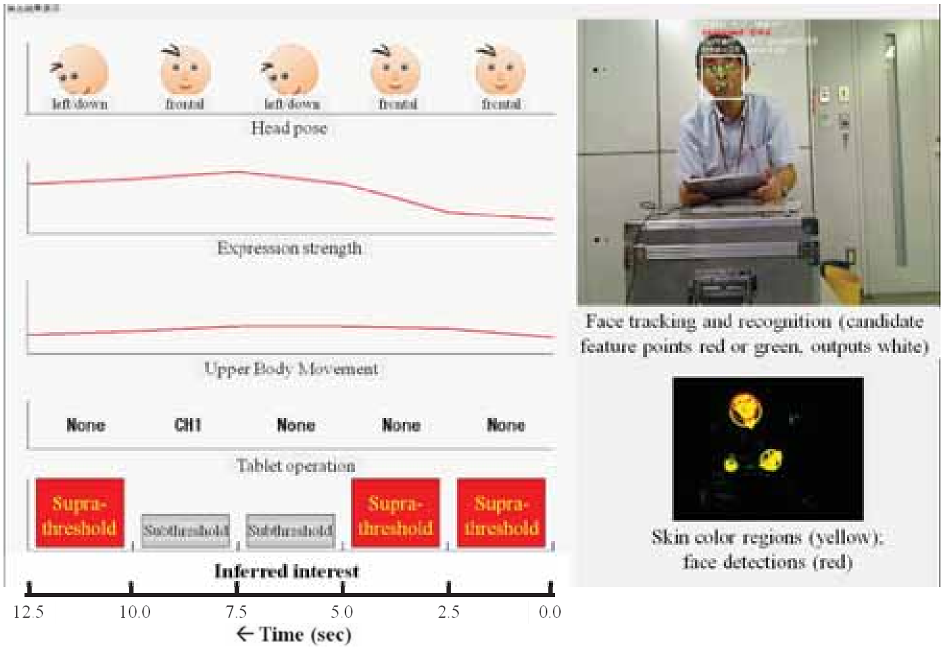


Figure 2. Screenshot of prototype system operation. Right: Face Module. Left: Interest Estimation Module (NB there is a one-slot time lag before inferred interest appears). Level of interest is analog, but displayed as a thresholded binary variable.

C. *Interest Estimation (CRF) Module*

The Interest Estimation Module uses a Conditional Random Field (CRF) algorithm [7] to combine the following outputs from the image processing modules and the tablet interface, with weights learned from labeled training data, to estimate the viewer's level of interest in each time slot (currently about 2.5 seconds):

- *Head Pose*: frontal head pose suggests that the viewer is looking at the TV screen, while non-frontal pose suggests that his/her attention is elsewhere;
- *Facial Expression*: expression changes while looking at the screen are presumed to indicate interest;
- *Upper Body Movement*: little movement is presumed to indicate attention, while large and/or frequent movements suggest inattention;
- *Tablet Operation History*: for example, requested volume increases suggest interest, while channel changes in mid-program suggest a lack of interest.

The operation of the system is illustrated in Figure 2. A preliminary cross-validation experiment using nine subjects (eight for training and the remaining one for testing, in rotation) showed an average accuracy, in inferring intervals in which viewer interest was present, of 71% [8]. A more thorough experiment with a larger cohort is in progress.

III. CONCLUSIONS AND FURTHER WORK

The current prototype is an early attempt at inferring interest from viewer behavior, intended as a testbed for the tuning of various detectors and algorithms, and for the accumulation of a significant volume of training data.

Clearly one can think of exceptions to the "rules" above which will need to be dealt with on a case by case basis. Examples include people who are listening to the TV at a distance (outside the visual field of the camera) while doing something else, and people apparently watching the TV but talking about something entirely unrelated; their facial expressions should presumably not be interpreted as responses to the program, even at frontal head pose.

Despite its restricted scope, though, the system does give a first approximation to a viewer's level of interest. It is expected that both viewer interest, and the target of that interest, will be more accurately inferred by replacing the simple head pose tracking used now with gaze tracking. Other issues that demand careful attention include privacy, if the information extracted is sent to a remote location for any purpose, such as the automatic compilation of viewer statistics or assessing the impact of commercials.

REFERENCES

[1] S. Clippingdale and M. Fujii, "Video face tracking and recognition with skin region extraction and deformable template matching," *International Journal of Multimedia Data Engineering and Management*, 3(1), pp. 36-48, January-March 2012.

[2] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision* **57**(2) pp. 137-154, 2004.

[3] R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection", Proc. ICIP 2002, pp. 900-903 (2002).

[4] Herbert Bay, Andreas Ess, Tinne Tuytelaars and Luc Van Gool, "SURF: Speeded Up Robust Features", *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346—359 (2008).

[5] Vladimir N. Vapnik, "*The Nature of Statistical Learning Theory*", Springer-Verlag, 1995.

[6] T. Kanade, J. F. Cohn and Y. Tian, "Comprehensive database for facial expression analysis," *Proc. Fourth IEEE International Conference on Automatic Face and Gesture Recognition, FG2000.*

[7] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning," in "*Introduction to Statistical Relational Learning*". Edited by Lise Getoor and Ben Taskar. MIT Press. (2006)

[8] M.Naemura et al, "CRFs for Estimation of TV Viewer Interest", IEICE Tech. Report PRMU2011-186 (in Japanese).

About Encouraging Residential Users to Share Upload Bandwidth with CDN/P2P Live Streaming Systems

Leandro de Sales, *Student Member, IEEE*, Kyller Gorgônio, Hyggo Almeida, Angelo Perkusich, *Member, IEEE*

***Abstract* — We encourage residential users to share upload bandwidth with CDN/P2P streaming systems by providing an incentive algorithm for fair peer selection. CDN servers compute a rank of best contributors that obtain better quality of service.**

I. INTRODUCTION

Real time multimedia delivery on the Internet is essential for the most current applications, such as Voice over IP (VoIP), Videoconferencing and WebTV. In applications of this type, the CDN/P2P has shown to be an efficient solution for large-scale video distribution over the Internet [1].

In CDN/P2P systems, such as LiveSky [2], a data-source processes the captured media (video, audio or both) and transmits network packets to CDN servers. These servers are deployed in multiple locations, often over multiple backbones and ISPs. Then, Internet clients connect to one of these servers and start to playback the media. As the clients receive the media content, they become a media relay and share the content with other clients. This constitutes a P2P network, where clients connect directly to each other and also playback the same media, indirectly from one of the CDN servers.

A. THE PROBLEM

An important challenge for CDN/P2P systems is how to encourage end-users to share upload bandwidth with other end-users through its media player. This problem becomes increasingly critical when end-users with idle upload bandwidth (potential relays) are not interested in watching a live streaming interested by others. This is a typical scenario that CDN/P2P systems rely on: the users need to be interested in receiving a stream to allow the infrastructure to ask them to share the content with others. When a potential relay has no interest in receive the stream, it is not possible to use it, causing a waste of idle resource (no interest, no sharing).

B. THE GOAL

This work proposes an incentive algorithm to help CDN/P2P systems improve the end-user quality of experience while watching live streaming on the Internet.

II. THE SOLUTION

The system applies a score-based mechanism for scheduling a relay to a client, while relay contributors are rewarded with tickets. The ticket mechanism gives ToB (Tickets of Bits) to a relay as it shares its upload bandwidth with other users that successfully receive media contents. The algorithm considers the quality of the stream to calculate the amount of ToBs to reward the user, which can convert them in money credit, discounts for online purchases and online movie rentals.

The general aspects of our system can be summarized as following. It provides two modules, one that runs in the user side, usually in a home router (named client module, or client) and the other that runs in the CDN servers (named server module, or server). The client executes an Incentive Algorithm (IA), responsible for allowing a user to register its device as a relay and also receiving live streaming content from other relays or directly from a server. The server executes a Scheduling Algorithm (SA), responsible for selecting the proper relay(s) for a given user according to its available ToB.

As a first step, the router admin user selects the live streaming system and specifies the username and password to sign in a server deployed by a live streaming system. During this process, the client computes its upload bandwidth and delay (in milliseconds), and sends back to its server. This information is sent periodically to the server, which adjusts the input parameters of the algorithm. After finishing the sign in process, the client module is ready to both receive live streaming content and act as a relay.

A. RECEIVING LIVE STREAMING CONTENT

The client application running into the end-user host, connected to the local network through a home router executing the client module, sends a request to a remote live streaming system, the same is selected by the admin user in the router settings. The request message is then captured by the client module, which redirects it to the proper CDN server.

Once a CDN server receives the request for a live streaming transmission, it schedules the proper relay client according to the client ToB. If there is no available relay, the CDN server transmits the live streaming content directly to the client. In this case, if the client has an available ToB in its profile, the server consumes them as it sends data to the client, according to a business model defined by the live streaming system and running into the server module. For example, for each 1Mb transmitted to the client module, the server module consumes 10 ToB from the user profile. Also, a user can buy ToBs from the live streaming system, which is used to start to participate of the system; otherwise the client must act as a relay to obtain ToBs. Even if a client does not have ToBs, the live streaming system can tolerate certain amount of data transfer from the CDN server to the client. In this case, the client owes the system and, as the client contributes to the system, it pays the debits with the system.

B. THE CLIENT AS A RELAY

When a client becomes a relay, the server must bill the relay activities and generates user's ToB. The server works according to the philosophy that the more a client contributes

This work is partially funded by the CNPq/Brazil under the scholarship process number 142595/2010-1.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

to the system, the more it gets ToB and better relays will be assigned to it. As a result, this will improve the quality of user experience that contributes more, as well as worsening the experience of free-rider users, which receive limited relays.

In this manner, the server stores in a relay profile database the Statistics of Contributions (SC) per client account and disseminates this information across the other servers. The statistics of contributions include how many times it acted as a relay and for how long; and what was the quality of video experimented by the other client(s) when it acted as a relay.

The client sends its SCs to the CDN server, which computes a scoring function and performs a relay selection. The contribution level of a relay *i* is converted to a score defined as c_i, which is mapped in a rank R_i. A relay selection depends on the rank ordering of the client interested in receiving a live streaming content. The relay selection scheme schedules relays with equal or lower rank R_i to serve as suppliers. The relay selection process is the realized quality of a streaming session. To evaluate delivery quality and quantify the performance of a relay in a Streaming Session (SS), we define quality Q according to the Equation 1.

$$Q_{c_i} = \frac{\sum_{j=1}^{P} V_j}{P} \times T_{ss} \tag{1}$$

where: P is the number of packets in a streaming session; V_i is a variable that takes value 1 if packet *j* arrives at the client at playout time, or 0 otherwise; and T_{ss} is the time the relay serves the client, in seconds. The ToB of a given relay can be expressed as a function of contributions level and the quantity K of clients the relay serves, counting those that receive packets in multicast mode, when used by the relay. Each client evaluates the Equation 1 and transmits the result to the relay's server. Then, the server uses Equation 2 to compute the ToB in a streaming session ss, defined as $\tau_{i_{ss}}$.

$$\tau_{i_{ss}} = \sum_{j=1}^{K} Q_{c_j} \tag{2}$$

A user may take a long time to use its ToB. This becomes critical in case the user, using a good Internet connection, obtains a good τ_{ss} and later uses a bad Internet connection to become a relay. In order to avoid this situation, the accumulated ToB, defined as τ_f, is calculated with Equation 3.

$$\tau_f = \beta \times \tau_{ss} + (1 - \beta) \times \tau_f \tag{3}$$

where: β is the penalty factor based on the current Q_{c_j} observed in the streaming session. The higher the value of Q_{c_j}, the lower the value of β.

In this way, we avoid outliers when measuring ToB and also avoid the aforementioned situation, forcing the user to always connect from the same Internet connection quality, otherwise it will loose part of its ToB obtained previously. The τ_f value is stored in the user profile for later use as a client. Note that when a user buys ToBs, they are used to select better relays to it, but not to rank the user's client to become a relay.

Moreover, the scheduling mechanism of each server is responsible for exchanging packets with all its relays. Swarm-like content delivery is performed and each relay periodically generates a report that contains a buffer map of its newly received packets and sends it to its clients. Each client periodically requests a subset of required packets from each relay based on the reports. A pull mode is incorporated to the client to fetch absent packets marked in the relay's report as missing. The client asks for them directly to the server its relay is connected to, but the server response is sent to the relay and not to the client that asked the packets. This allows the relay to deliver the missing packets not only to the client that requested them, but also to any other client in its neighborhood, what improves the response time. A packet-scheduling algorithm in execution in the client determines packets to be requested by the pull mode, which is very similar to the data-driven approach in DONet [3].

III. RESULTS

For our case study, we have developed the client and the server modules. The client is an extension of the Linux DD-WRT[1] operating system, while the server module an application in the Omnet++[2] simulator. We have performed simulations and our solution improved in 31% the quality of an MPEG video transmitted using a simulated CDN/P2P live streaming system named Denacast [6], while decreased the quality of the video in 19% observed by free-riders clients.

IV. CONCLUSION

In this paper we have presented a system to encourage end-users to share their upload bandwidth with live streaming systems based on an incentive algorithm. Our proposal allows users to receive the same level of quality of service it provides to the others, while naturally discourages free-riders to use the system. The algorithm also covers changes in the user's upload bandwidth by evaluating the stream quality and penalizing its ToB. The system can be deployed in common Internet-enabled devices, such as home routers, Smart TVs and set-top-boxes, while end-users can exchange their ToB by real goods and services, depending on the business model being used by the live streaming system.

REFERENCES

[1] Fortino, G. et. al., "CDN-Supported Collaborative Media Streaming Control" *IEEE Multimedia*, vol. 14, pp. 60-71, April 2007.

[2] Yin, H. et. al., "Design and deployment of a hybrid CDN-P2P system for live video streaming: experiences with LiveSky." 17th ACM International Conference on Multimedia, pp. 25-34, March 2009.

[3] X. Zhang, et. al., "DONet : A data-driven overlay network for efficient live media streaming" INFOCOM 2005, vol. 3, pp. 2102-2111, March 2005.

[4] S. M. Y. Seyyedi and B. Akbari, "Hybrid CDN-P2P architectures for live video streaming: Comparative study of connected and unconnected meshes," in 2011 International Symposium on Computer Networks and Distributed Systems (CNDS), Tehran, Iran, 2011, pp. 175-180.

[1] http://www.dd-wrt.com/. Last accessed in October 2012.

[2] http://www.omnetpp.org/. Last accessed in October 2012.

Sensor Fusion-Based People Counting System Using the Active Appearance Models

Seung-Wook Kim, June-Young Jung, Seung-Jun Lee, Aldo W. Morales, and Sung-Jea Ko
School of Electrical Engineering, Korea University, Seoul, Korea

Abstract--**The paper presents a novel robust people counting system using the active appearance model (AAM). Conventional people counting methods utilizing the monoscopic or stereoscopic image data often fail due to occasional illumination change and crowded environment. The proposed algorithm uses both the vision and depth image captured by a vision-plus-depth camera mounted on the ceiling. Then, we construct a 3D human model from the depth image using the AAM to segment and recognize human objects. Experimental results show that the proposed algorithm achieves over 97% accuracy in various testing environments.**

I. INTRODUCTION

The robust people counting system is an important tool in surveillance such as monitoring the number of passengers in transportation and counting people at a certain floor of the building to prevent crowded situation [1]. Over the recent years, a number of vision-based algorithms for people counting have been developed. An algorithm which extracts and tracks the human silhouette has been proposed in [2]. Kim *et al.* [3] developed a real-time system in which the person is modeled using box-based feature extraction with background subtraction. Even though these vision-based algorithms have low computational complexity, they cannot extract each individual from the crowd successfully in a crowded condition. In order to segregate the crowd into individuals, Lee *et al.* [4] used the depth information obtained by the stereo vision method. However, the depth data from the stereo images is less reliable than that from the infrared (IR) structured light sensor since the vision-based depth data suffers from illumination changes, shadows, and reflections.

The proposed system is based on an effective fusion of the vision image and the depth image captured by the IR sensor as shown in Figs. 1 (a) and (b). In the proposed algorithm, the input depth image is preprocessed to recover the lost depth data caused by the optical noise and absorption of IR light by the dark surfaces [4]. Then, the resultant image is segmented to discriminate the human objects from the background. To detect the human objects, we introduce a 3D human model based on the active appearance model (AAM) [5]. Finally, the segmented human objects are tracked using the Bhattacharyya similarity [6], and the counter of the object is increased when the object trajectory crosses a counting line.

This paper is organized as follows. Section II explains the proposed system in detail. The experimental results are provided in Section III.

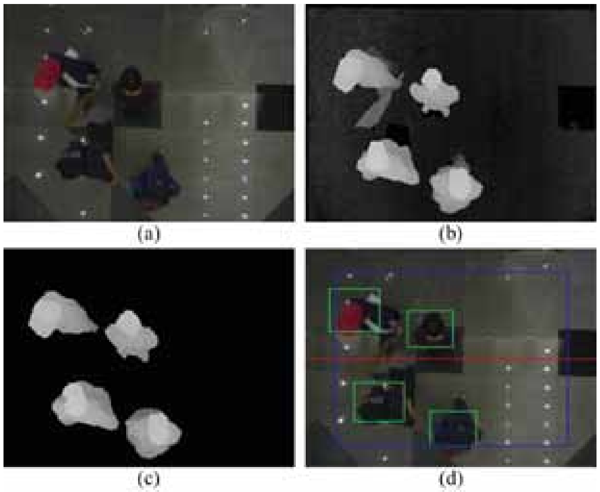


Fig. 1. (a) Vision image. (b) Original depth image. (c) Recovered and clipped depth image. (d) Output image of the proposed system.

II. PROPOSED SYSTEM

The proposed system consists of three stages. In the first stage, the system recovers the lost data in the depth image. In the second stage, the human candidate is first detected using a simple thresholding method and then the pre-trained AAM is utilized to determine whether the candidate is human or not. In the final stage, people tracking and counting are performed.

A. Depth Data Recovery

To recover the lost data in the depth image, object boundaries have to be extracted. We propose a boundary extraction technique that uses binarization of the depth image followed by a series of morphological operations. Let $Z(x,y)$ denote the depth image. Each pixel in the depth image can be simply classified into foreground(1) and background(0) pixels as follows:

$$I(x,y) = \begin{cases} 1, & if\ Z(x,y) - \tau > \sigma, \\ 0, & otherwise, \end{cases} \tag{1}$$

where τ is a depth value corresponding to the height of the camera measured from the floor and σ is a threshold used to eliminate parts of the object which are relatively far from the camera. Then, a refined binary image $\hat{I}$ is obtained applying the morphological closing operation to the binarized depth image I. Finally, in order to extract the object boundary B, the following morphological operation is performed:

$$B(\hat{I}) = (\hat{I} \oplus S) - \hat{I}, \tag{2}$$

where $\oplus$ is a dilation operation and S is a 3×3 square structuring element.

After the object boundary is extracted, the nearest neighbor interpolation method is applied to recover the lost depth data. Let R^+ be the set of pixels inside the object boundary whose

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012R1A2A4A01008384).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

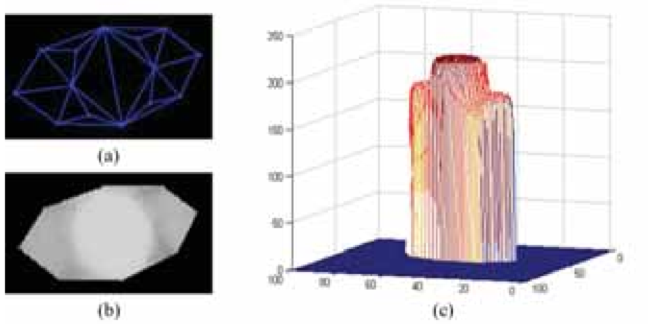


Fig. 2. (a) Shape model of the human object. (b) Texture model of the human object. (c) 3D human model.

depth data $Z^{+}(x,y)$ is successfully acquired by the sensor. Then, the lost depth data $Z^{-}(x,y)$ to be recovered is obtained by

$$Z^{-}(x,y) = Z^{+}(x^{*},y^{*}), \tag{3}$$

where

$$(x^{*},y^{*}) = \underset{(u,v)\in R^{+}}{\arg\min}\left(\sqrt{(u-x)^{2}+(v-y)^{2}}\right). \tag{4}$$

The example of the recovered depth data is shown in Fig. 1 (c).

B. *Human Detection*

To detect the human object, we construct a human model using the AAM [5]. N depth images marked with user-defined 2D landmark points are utilized to build a training set. The shape vector $\mathbf{x}^{i} = (x_1^i, y_1^i, \cdots, x_n^i, y_n^i)^T$, $i = 1, \cdots, N$ is defined to construct the shape model, where n is the number of 2D landmark points from the ith training image. Using the method in [5], the texture vector $\mathbf{g}^{i}$ is extracted from the shape modeled region in the ith depth image. The dimension of the feature vectors is reduced using the principal component analysis (PCA). As a result, linear models are obtained for both the shape with $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s\mathbf{b}_s$ and the texture with $\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g\mathbf{b}_g$, where $\bar{\mathbf{x}}$ and $\bar{\mathbf{g}}$ represent the mean vectors for the feature vectors **x** and **g**, respectively. $\mathbf{P}_s$ and $\mathbf{P}_g$ are the sets of the eigenvectors obtained using the PCA and the sets of model parameters for the feature vectors are represented as $\mathbf{b}_s$and $\mathbf{b}_g$. Since $\mathbf{P}_s$ and $\mathbf{P}_g$ can still be correlated, the PCA is applied once more to obtain the following combined parameter **c**,

$$\mathbf{c} = \mathbf{P}_c^T \begin{pmatrix} \mathbf{W}_s\mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix}, \tag{5}$$

where $\mathbf{P}_c$ is the sets of the eigenvectors and $\mathbf{W}_s$ is a diagonal weight matrix to compensate the different ranges of the eigenvalues of the shape and texture eigenspaces. Since the gray-level of the depth image represents the height of the object, **x** and **g** denote 3D models of the human object as shown in Fig. 2.

The AAM is initialized by extracting the head cross section $I_{head}(x,y)$ as human candidates using a simple thresholding method as follows:

$$I_{head}(x,y) = \begin{cases} 1, & if\ Z(x,y) > 0.9h \\ 0, & otherwise, \end{cases} \tag{6}$$

where h be the depth value associated with the height of the human candidate. After initializing the model using the head cross section $I_{head}(x,y)$, the AAM search algorithm [5] is applied to match the model with a human candidate. Let $\mathbf{g}_m$ be the texture vector from the trained model and $\mathbf{g}_s$ be the texture vector extracted from the depth data of the human candidate. A simple scalar measure of difference using Euclidean distance between the texture vectors is given by

$$E = |\mathbf{g}_s - \mathbf{g}_m|^2. \tag{7}$$

Finally, the human candidate is classified as a human if the scalar measure E is smaller than the threshold acquired by training results. The human detection results are represented as green boxes in Fig. 1 (d).

C. *Tracking and Counting*

For automatic counting of people passing through a gate or a space, a counting line must be designated. Then, the Bhattacharyya similarity [6] is used in order to match the objects in the successive frames. The system registers the trajectory of each object using the matching information. As the trajectory passes through the counting line, the system records an increase of the number of people.

III. EXPERIMENTAL RESULT

In the proposed system, we used both depth and vision images of QVGA (320 × 240) resolution. The test video sequence was captured in an indoor environment. The camera was mounted at 2.9m height with a top-view angle. A PC with 2.8 GHz CPU was used for the experiments.

TABLE I
PEOPLE COUNTING RESULT

Method	Ground truth	Missed count	Over-count	Total error	Detection accuracy
Vision-based	412	32	53	85	79.3%
Proposed	412	11	1	12	97.1%

As shown in Table I, the vision-based [7] and proposed methods produce the accuracy of 79.3% and 97.1%, respectively. The average processing time per frame of the system is 12ms. Therefore, the proposed method can be successfully applied to real-time people counting system.

REFERENCE

[1] K.-Y. Yam, W.-C. Siu, N.-F. Law, and C.-K. Chan, “Effective bi-directional people flow counting for real time surveillance system,” *IEEE Int. Conf. on Consumer Electron.*, pp. 863-864, Jan. 2011.

[2] J. Segen and S. Pingali, “A camera-based system for tracking people in real time,” in *IEEE Proc. Of Int. Conf. Pattern Recognition*, vol. 3, pp. 63-67, Aug. 1996.

[3] J.-W. Kim, K.-S. Choi, B.-D. Choi, and S.-J. Ko, “Real-time system for counting the number of passing people using a single camera,” *Lecture Notes in Computer Science*, vol. 2781, pp. 466-473, Sep. 2003.

[4] S. H. Lee and S. Sharma, “Real-time disparity estimation algorithm for stereo camera systems,” *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, pp. 1018-1026, Aug. 2011.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681-685, Jun. 2001.

[6] T. Kailath, “The divergence and Bhattacharyya distance measures in signal selection,” *IEEE Trans. Communication Technology*, vol. 15, no. 1, pp. 52-60, Feb. 1967.

[7] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, “Real-time foreground-background segmentation using codebook model,” *Real-Time Imaging*, vol. 11, no. 3, pp. 172-185, Jun. 2005.

An Efficient Path Planning Method for a Cleaning Robot Based on Ceiling Vision

Junho Park, WooYeon Jeong, Hyoung-Ki Lee, and Jonghwa Won*, *Member, IEEE*
Samsung Advanced Institute of Technology, Samsung Electronics, Korea

***Abstract*--We propose an efficient path planning method based on an algorithm that aligns the geometric pattern of a ceiling with the heading angle of a cleaning robot. This algorithm results in a 22% faster cleaning time for a well-aligned cleaning path. We implemented and verified our proposed method for a cleaning robot using a ceiling vision simultaneous localization and mapping (SLAM) algorithm.**

I. INTRODUCTION

Fast cleaning and full coverage are the most important factors to consider when evaluating a cleaning robot [1-4]. An interdigitated zigzag cleaning path instead of a random path or combined path ensures fast cleaning and full coverage [1, 5]. Cleaning robot movement algorithms use information from ultrasonic sensors and laser sensors to detect the heading angle to enable the robot to move straight [2]. However, these sensors have an error rate of more than 2.5% [3]. As an alternative, cameras can be installed in the ceiling to control the heading of a robot [6].

In this paper, we present a new, fast, path planning algorithm that aligns the direction of the heading angle of a cleaning robot with the walls to ensure fast cleaning and full coverage. In many residential homes and offices, the ceiling patterns have the same direction as the walls of rooms and corridors. [3]. Therefore, our algorithm aligns the heading of the robot with the direction of the ceiling patterns by making use of image information provided by the cleaning robot's ceiling-vision camera [1].

CEILING PATTREN ALIGNMENT ALGORITHM

Our proposed ceiling pattern alignment algorithm consists of two steps. First, feature points are extracted from the image of the ceiling and the directionality of the ceiling pattern is determined. Second, the location and heading of the cleaning robot are identified by SLAM. The heading of the cleaning robot is aligned to the direction of the ceiling pattern.

A. Feature extraction and directionality of the ceiling pattern

Our cleaning robot is equipped with a CMOS camera and the viewpoint of the camera is upright toward the ceiling. Several ceiling patterns can be extracted from the camera images, such as lines, boxes, circles, and corners. Furthermore, lines and points extracted from the contours of lamps can be added as ceiling patterns.

To extract useful features from the ceiling pattern, well-characterized feature detection algorithms such as the Harris corner detector algorithm, scale-invariant feature transform (SIFT), and maximally stable extremal region (MSER) can be used. To extract the directionality of the ceiling pattern, we select lines parallel to the ceiling plane and group the patterns according to the direction of the lines. Then, we determine the directionality of the ceiling pattern by the number of lines or/and the length of the lines.

As shown in Fig. 1, the optical center is used to select lines parallel to the ceiling plane. Blue lines pointing toward the optical center are eliminated because they are vertical to the ceiling. Using the Canny edge detector equation (1) and the Hough transform equation (2), the angles of the red lines are calculated as follows:

$$G = \sqrt{G_x^2 + G_y^2}, \theta = \text{arctan}(\frac{G_y}{G_x}) \qquad (1)$$

where Gx and Gy are the horizontal and vertical directions, respectively, and

$$y = (-\frac{\cos\theta}{\sin\theta})x + (\frac{r}{\sin\theta}) \qquad (2)$$

where r is the distance from the origin to the line and θ is the angle of the vector from the origin to this closest point.

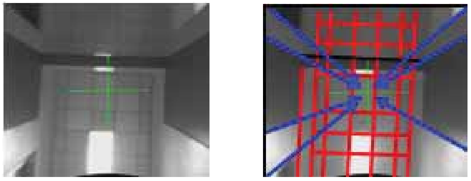


Fig. 1 Original image showing optical center and selected lines.

Another way to obtain the directionality of the ceiling pattern is to use repetitive patterns in the ceiling images such as light bulbs, tiles, etc., as shown in Fig. 2. To do this, small image patches are obtained around feature points and compared. The similarity of local feature images can be compared using normalized cross correlation (NCC), sum of squared differences (SSD), sum of absolute differences (SAD), or the mean of absolute differences (MAD). We selected NCC as the best method for our cleaning robot. We apply NCC to each block as:

$$\text{NCC}(I_N) = \frac{1}{n}\sum_{u,v \in I_N} \frac{\left(I(u,v)_N - \bar{I}_N\right)\left(I(u,v)_{N-1} - \bar{I}_{N-1}\right)}{\sigma(I_N)\sigma(I_{N-1})} \qquad (2)$$

where I_N denotes the N-th feature image patch and $\bar{I}_N$ is the average of gray levels for the N-th feature image patch. $I(u, v)_N$ is the gray level of pixels at the image coordinate of *(u, v)* in each patch, $\sigma(I_N)$ is the standard deviation of gray levels for each patch, and n is the number of pixels in each block.

* Corresponding Author

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

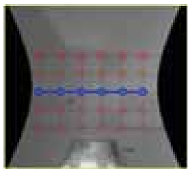


Fig. 2. Repetitive patterns in a ceiling image.

B. Robot localization and well-aligned path planning

A cleaning robot has several sensors such as an odometer, accelerometer, gyroscope, and image sensor to localize and build a map (SLAM). We use a particle filter method to calculate the probability of the robot location and heading angle and Monte-Carlo localization (MCL) to select the most probable particle. The location of the most probable particle is the estimated location of the robot. The simultaneous location of the robot is used to register landmarks, map-buildings, and the robot trajectory. It is also possible to compare the robot trajectory on the map and the directionality of the ceiling pattern.

This allows the robot to clean in the right direction, as shown in Fig. 3, and also allows the robot to deal with unexpected heading changes due to slippage while cleaning.

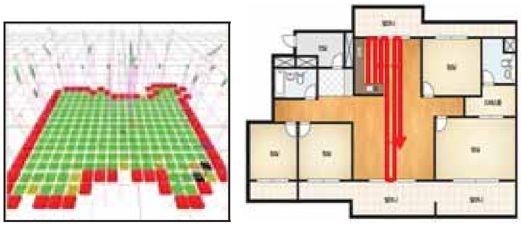


Fig. 3. SLAM and well-aligned cleaning robot path.

II. EXPERIMENTAL RESULTS

We implemented our algorithm in the VC-RE70V cleaning robot that has an embedded ARM9 main processor, one ceiling VGA (640x480) gray camera, and windows CE as the OS. Lines and/or repetitive patterns were detected successfully from the images of residential and office ceilings, as shown in Fig. 4.

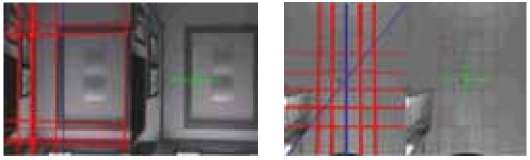


Fig. 4. (a) Residential house (b) office at SAIT.

To determine the efficiency of our proposed algorithm, we tested the robot according to the Korean Standard (KS) for evaluation of a cleaning robot. The test platform is defined in KS B 6934:2011-2.2 as shown in Fig. 5[7]. It was designed for dust removal ability test, but we used it for coverage ability test in this experiment. A cleaning robot should clean an area of 3 m x 3 m squared in 3 minutes. Coverage, which is the area swept by the robot, is a measure of efficiency. In Fig. 6, the random path and semi-random path algorithms showed relatively low coverage within a given time. In contrast, the cleaning robot that used a map showed higher coverage.

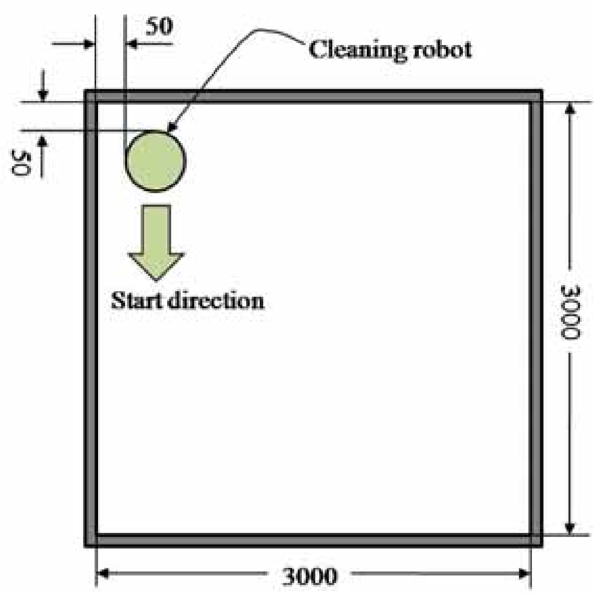


Fig. 5. KS platform for a dust removal ability test (unit:mm)

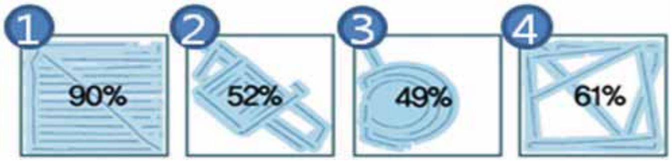


Fig. 6. Coverage comparison according to the Korean standard for cleaning robots.

Finally, we evaluated the performance of our cleaning robot for intentionally ill-aligned paths (angles of 15°, 30° and 45°), as shown in Fig. 7. We measured the time it took for almost 100% coverage; the results are presented in Table I. Full coverage took 220 seconds for the well-aligned path, whereas full coverage took anywhere from 268 ~298 seconds for the ill-aligned paths. Our robot is equipped with infra-red sensors for detecting obstacles and boundaries. Using those sensors, we implemented a water-filling and wall-following methodto cover the whole area. When the cleaning robot detects obstacles, it executes wall-following algorithm while it goes straight forward. For the 30° ill-aligned path shown in Fig. 7, the robot could not maintain a zigzag path using the water-filling method. However, based on the results shown in Figs. 6 and 7, cleaning based on a map is better than the other methods that we tested, regardless of whether the trajectories are aligned or not.

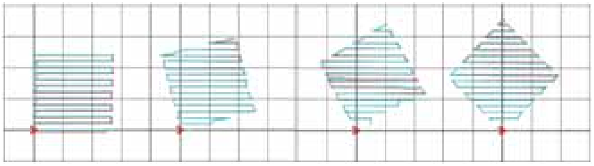


Fig. 7. Well-aligned path and 15°, 30°, and 45° ill-aligned

paths.

TABLE I.
COMPARISONS OF TIME FOR NEAR 100% COVERAGE

Method	Time for Near 100%
Well-aligned path	220 seconds.
15° Ill-aligned path	268 seconds
30° Ill-aligned path	298 seconds
45° Ill-aligned path	284 seconds

III. CONCLUSIONS

We developed and evaluated an efficient path planning method for a cleaning robot. A cleaning robot using our heading angle alignment algorithm cleaned at least 22% faster than cleaning robots that used other types of heading angle alignment algorithms. In future studies, we hope to improve the water-filling algorithm, because our cleaning robot sometimes works inefficiently, as shown for the 30° ill-aligned trajectory (Fig. 7).

REFERENCES

[1] WooYeon Jeong and Kyoung Mu Lee, “CV-SLAM: A new Ceiling Vision-based SLAM technique”, Intelligent Robots and Systems, IROS 2005, pp. 3195-3200, 2005.

[2] Guoping Hu, Zhengwei Hu and Hongbo Wang, “Complete Coverage Path Planning for Road Cleaning Robot”, Int. Conf. on Networking, Sensing and Control, ICNCS 2010, pp. 643-648, 2010.

[3] Zhengwei Hu, Hongbo Wang, Tian Zhang, Xiaoqian Zheng and Xue Yang, “Path Planning and Control System Design for Cleaning Robot”, Int. Conf. on Information and Automation, ICIA 2010, pp 2368-2378, 2010

[4] Jeong H. Lee, Jeong S. Choi, Beom H. Lee and Kong W. Lee, “Complete Coverage Path Planning for Cleaning Task using Multiple Robots”, Int. Conf. on Systems, Man, and Cybernetics 2009, pp. 3618-3622, 2009.

[5] Yu Liu, Xiaoyong Lin and Shiqiang Zhu, “Combined Path Planning for Autonomous Cleaning Robots in Unstructured Environments,” Proc. Of 7th World Congress on Intelligent Control and Automation, 2008, pp 8271-8276, 2008.

[6] Tsutomu Takeshita, Tetsuo Tomizawa and Akihisa Ohya, “A House Cleaning Robot System – Path indication and Position estimation using ceiling camera-” International Joint Conference SICE-ICASE 2006, pp. 2653-2656, 2006

[7] Sungsoo Rhim, Jae-Chang Ryu, Kwang-Ho Park and Soon-Geul Lee, “Performance Evaluation Criteria for Autonomous Cleaning Robots”, Proc. of the 2007 IEEE Int. Symp. On Comp. Intelligence in Robotics and Automation, pp.167~172, 2007

ECA-based Control Interface on Android for Home Automation System

Marcos Santos-Pérez, Eva González-Parada and José Manuel Cano-García
School of Telecommunications Engineering, University of Malaga, 29071 Malaga (Spain)
Email: marcos_sape@uma.es, eva@dte.uma.es, cano@dte.uma.es

***Abstract*—Historically, Embodied Conversational Agents (ECAs) have been used as virtual assistants that make easier the access to information or help in performing complex tasks. Due to their high computational requirements ECAs are usually run on desktop computers, but with the recent development of hand-held devices both in hardware and software, it becames neccessary to move ECAs to that new mobile scenario. Thus, we propose an open-source based platform for developing ECA based interfaces on Android-equipped devices. We also present a prototype for controlling a home automation system.**

I. Introduction

Embodied Conversational Agents (ECAs) are animated virtual characters that emulate human behaviour and communication. ECAs arise as a conversational partner for the user in computer-based environments [1]. Instead of addressing the computer, the user addresses a virtual agent that can be made responsible for certain tasks, as performing a web search, answering a fixed-domain question or controlling the home automation system [2] among others. Other approach is the use of an ECA as a sociable and emotionally intelligent companion for the user [3].

Due to the limited computational power of hand-held devices compared to desktop computers, the most common architectures for ECA-based mobile applications rely on an external server that performs the processor intensive tasks, such as speech recognition, language understanding and text-to-speech [4]. But in the last few years the embedded processors have significantly increased their computational power and it starts to be possible to run an ECA completely within a hand-held device.

This paper describes an platform for developing ECA-based interfaces on Android hand-held devices. The proposed platform is based on free and open source libraries. We developed a prototype installed on a tablet for controlling a home automation system.

II. ECA Platform Overview

The architecture of the platform shown in Figure 1 follows a modular design so that each component can be modified without affecting others. The conversational engine is implemented as a Python module and the rest of the elements are provided as native libraries, accessible through a facade-interface.

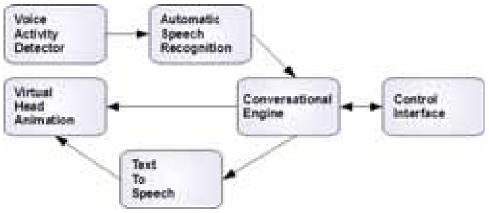


Fig. 1. Architecture of the proposed platform.

A. Voice Activity Detector

The Voice Activity Detector's (VAD) role is to discriminate the user's voice frames from those containing noise. That way, the VAD allows the segmentation of the user's speech into utterances. This module reads the digitized audio samples acquired from a microphone and sends the filtered raw audio to the ASR. The actual implementation of the VAD module is based on the SphinxBase library, which was modified so it can work with the OpenSL ES native audio libraries present on Android.

B. Automatic Speech Recognition

The Automatic Speech Recognition (ASR) module performs speech to text conversion. It takes as input the utterance with the user's speech that come from the VAD and send the resultant text to the CE. In the proposed platform, the ASR module is based on the PocketSphinx speech recognition library. Some changes were made to the original code in order to improve the response time on embedded devices by starting the recognition phase once the first speech frame is detected by the VAD [5] and by choosing the most appropriate language model for the topic of the conversation [6].

C. Conversational Engine

The Conversational Engine (CE) extracts the meaning of the utterance, manages the dialog flow and produces the actions appropriate for the target domain. It generates a response based on the input, the current state of the conversation and the dialog history. The CE module is based on PyAIML, an AIML chatbot. AIML (Artificial Intelligence Markup Language) is an extension to XML that provides symbolic reduction, recursion, context-awareness and history management in order to understand the user's utterance and generate an appropriate response. The original code was improved with an optional lemmatizer submodule that reduces both the response time and the memory usage of the CE module when dealing with inflectional languages. It was also added support for an object-oriented database that can decrease the dynamic memory usage at the expense of an increment of the response time [7].

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

D. Control Interface

The Control Interface translates the commands said by the user to a format that can be understood by the target applications or services running on the same device or accessible remotely. This module is domain-specific and has to be re-implemented or adapted for every new target application.

E. Text-To-Speech

The Text-To-Speech (TTS) subsystem carries out the generation of the synthetic output voice from the text that comes as a response from the CE. For the sake of getting a realistic ECA, it sends to the VHA module a list of the phonemes with their duration so animation and artificial speech match up. The TTS module implementation is based on the eSpeak library.

F. Virtual Head Animation

The virtual head is the embodiment of the conversational agent and the visual counterpart of the TTS module.

This module receives as inputs both the mood information from the CE and the list of the phonemes' durations from the TTS module. By processing the inputs, it generates the visemes (the visual representation of the phonemes) and the facial expression that will be rendered along with the synthetic voice. The purpose of the phonemes' timing list is to modulate the animation speed to achieve perfect lip synchronization. OGRE 3D was used as the rendering engine of the software platform.

III. Prototype Details

The ECA requires a version of Android higher than 4.0 (also named ICS-Ice Cream Sandwich) and a hardware capable of coping with the computational tasks that form the platform. For our prototype we used a tablet with an OMAP 4460, which consists of a 1.2GHz dual-core ARM Cortex-A9 as CPU and a PowerVR SGX540 384MHz GPU. Fig. 2 shows the ECA running on the tablet.

The home automation system was previously implemented in a room and has a tactile manual control interface that is showed in Fig. 3. Likewise the manual control interface, the ECA can control the door lock, the lighting, the blinds and the temperature.

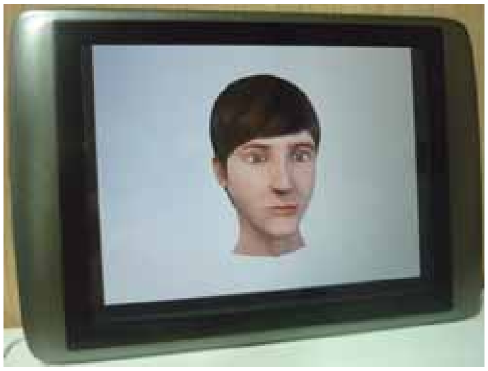


Fig. 2. ECA prototype in happy mood.

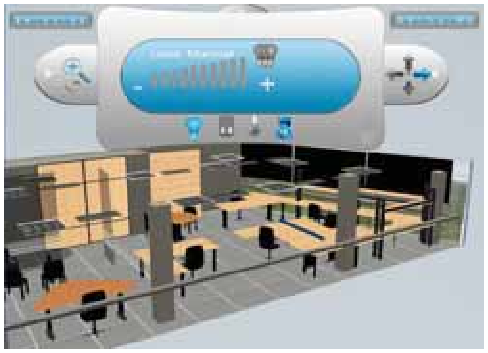


Fig. 3. Manual control interface of the home automation system.

IV. Conclusions and Future Work

The main goal of this work was to describe a platform aimed at developing ECA-based interfaces on hand-held devices equipped with Android. Thus, we proposed a possible architecture and gave implementation details for such platform. The whole platform is based on free and open source libraries and a first prototype was developed for controlling a home automation system.

The future work consists of to convey some experiments with real users to measure the usefulness, usability and performance of the platform.

Acknowledgment

This work was partially supported with public funds by the Spanish National Project TEC2009-13763-C02-01 and by the Andalusian Regional Project P08-TIC4198.

References

[1] M. M. Louwerse, A. C. Graesser, D. S. McNamara, and S. Lu, "Embodied conversational agents as conversational partners," *Applied Cognitive Psychology*, vol. 23, no. 9, pp. 1244–1255, 2009.

[2] B. De Carolis, I. Mazzotta, N. Novielli, and S. Pizzutilo, "Social robots and ECAs for accessing smart environments services," in *Proceedings of the International Conference on Advanced Visual Interfaces*, ser. AVI '10. New York, NY, USA: ACM, 2010, pp. 275–278.

[3] M. Cavazza, R. Santos de la Cámara, M. Turunen, J. Relaño Gil, J. Hakulinen, N. Crook, and D. Field, "How was your day?' An Affective Companion ECA Prototype," in *Proceedings of the SIGDIAL 2010 Conference*, Association for Computational Linguistics. Tokyo, Japan: Association for Computational Linguistics, September 2010, p. 277–280.

[4] H.-J. Oh, C.-H. Lee, M.-G. Jang, and K. Y. Lee, "An Intelligent TV interface based on Statistical Dialogue Management," *Consumer Electronics, IEEE Transactions on*, vol. 53, no. 4, pp. 1602 –1607, nov. 2007.

[5] M. Santos-Pérez, E. González-Parada, and J. M. Cano-García, "Efficient Use of Voice Activity Detector and Automatic Speech Recognition in Embedded Platforms for Natural Language Interaction," in *Highlights in Practical Applications of Agents and Multiagent Systems*, ser. Advances in Intelligent and Soft Computing. Springer Berlin Heidelberg, 2011, vol. 89, pp. 233–242.

[6] M. Santos-Pérez, E. González-Parada, and J. Cano-García, "Topic-Dependent Language Model Switching for Embedded Automatic Speech Recognition," in *Ambient Intelligence - Software and Applications*, ser. Advances in Intelligent and Soft Computing. Springer Berlin / Heidelberg, 2012, vol. 153, pp. 235–242.

[7] M. Santos-Pérez, E. González-Parada, and J. M. Cano-García, "Embedded Conversational Engine for Natural Language Interaction in Spanish," in *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*. Springer New York, 2011, pp. 365–374.

Tri-modal Speech Recognition for Noisy and Variable Lighting Conditions

Steven Anderson, Acm Fong, *Senior Member* and Jie Tang, *Member, IEEE*
[1]Auckland University of Technology, New Zealand
[2]Tsinghua University, China

***Abstract*—Automatic speech recognition (ASR) has found widespread applications in consumer products. Often, ASR performance can be compromised in noisy environments. Previous research has shown that adding visual cues can improve the performance of ASR, particularly in noisy environments. However, audiovisual (AV) ASR is not robust against changing lighting conditions, which are often encountered by end users of consumer products. Since thermal imaging is highly invariant to changing lighting conditions, we propose a tri-modal ASR involving thermal imaging and audiovisual (TAV) data for consumer applications. Experimental results demonstrate the applicability of this approach over a range of signal-to-noise ratios: Tri-modal TAV recognition rates were +39.2% over audio-only and +11.8% over AV recognition rates.**

I. INTRODUCTION

Consumer electronic devices are increasingly equipped with automatic speech recognition (ASR) capabilities to enhance the user experience. After all, giving verbal commands to an electronic device is a much more natural way of communication than other common ways of input mechanisms, such as keyboard entries. In the case of pocket-sized consumer products (smart phones, PDAs, etc.), the small form factor means that navigation using inputs via mechanisms such as hitting a tiny keyboard on the screen can be a challenge.

Often, an ASR application involves machine understanding of a limited vocabulary comprising a predefined set of commands, such as “turn on radio”, “FM”, “call Steven”, “call 123456”, etc. This class of ASR forms a significant type of consumer applications, as opposed to a more generic speech-to-text converter on a personal computer. Clearly, the challenges faced by the designers of these systems are quite different, particularly in view of the potentially very different operating environments.

In this research, we focus on the first class of ASR, which represents scenarios often encountered by consumers of portable consumer devices. In particular, since these devices are carried and used by consumers in different environments, the ASR must be robust under a range of environments with different background noise levels and lighting conditions. In particular, we investigate a fusion of three modes of processing involving audiovisual data and thermal imaging data to enhance the performance of ASR under different operating environments. Experimental results show that our tri-modal approach can outperform audio-only and audiovisual (AV) modes under a range of operating conditions without incurring significant computational costs.

II. RELATED WORK

Research in ASR started in the early 1950’s when Davis *et al.* constructed a system for isolated digit recognition for a single speaker recognizing each of the ten digits, ‘one’ through ‘nine’ and ‘oh’ for determining the identity of an unknown spoken digit [1]. Baum and Petrie [2] proposed a statistical framework based on a Hidden Markov Model (HMM), which was subsequently applied to ASR research. Most contemporary ASR systems are based on this statistical framework. Commercial ASR systems started to appear in the marketplace since the late 1980’s.

Today, ASR is in common usage throughout our daily lives, particularly with the proliferation of Bluetooth devices and mobile phones. However, the reliance on sound processing alone makes speech recognition unsuitable for use in noisy environments and even in clean environments state-of-the-art ASR systems often perform well below that of humans [3]. As such, the use of ASR often becomes a source of frustration for the consumer in practical situations.

It has been known for a long time that combining audio and visual analysis improves speech recognition accuracy in both noisy and clean environments. According to [4], the visual modality benefit to speech intelligibility in noisy environments has been quantified by as far back as in 1954 by [5]. A famous example of this is the *McGurk Effect*, where sound superimposed over the video of a speaker reciting a different sound can result in an observer hearing a third sound. For example when the sound played is /ga/ and the video is of a person saying /ba/, most observers will identify the word as /da/. Audiovisual ASR (AVASR) attempts to use the known benefit of a visual channel as a means to improve ASR.

Among the first AVASR system was developed in 1984 by Petajan [6] who showed that addition of the visual modality can improve the performance of ASR. Since then, a large amount of research has been conducted into AVASR and the majority of research projects have shown improvements to audio-only ASR in a variety of conditions. However, these systems are still highly affected by a number of factors that make them impracticable for real world consumer applications. One of the biggest problems encountered is changes in lighting conditions, which greatly reduces accuracy making AVASR unsuitable for use in uncontrolled environments.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

This research investigates the use of adding thermal imaging as a third modality to AVASR. Thermal images are insensitive to changes in lighting levels and are immune to the differences in skin tone that can make lip-reading difficult.

III. Tri-Modal ASR System

An AVASR system was created for testing the effect of adding a third modality to AVASR. Mel-frequency Cepstral Coefficient (MFCC) based speech recognition was used for audio speech recognition. For visual recognition, the standard video and thermal video data were processed using a modified form of Motion Templates combined with DCT for feature extraction. HMM was used for whole word classification, and high level fusion weighted to the reliability of each stream for each test subject was used for combination. We used high level fusion to allow for separate testing of individual elements. MATLAB R2009b was used for programming and creating the recognition system along with the HMM MATLAB Toolkit [7] and the VOICEBOX [8].

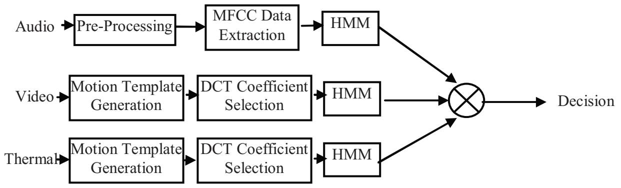

Audio
Pre-Processing
MFCC Data Extraction
HMM
Video
Motion Template Generation
DCT Coefficient Selection
HMM
Decision
Thermal
Motion Template Generation
DCT Coefficient Selection
HMM

Fig. 1. Tri-modal ASR system overview.

Fig. 1 shows an overview of the tri-modal system. Audio preprocessing entails silence removal, noise filtering, signal strength normalization, band-pass filtering and pre-emphasis filter to the signal to boost speech in the higher frequency ranges. The first step taken in MFCC is to split the signal into short time intervals using the Hamming windowing function. Following [9], we use a window size of 30ms in length was selected with a step of 15ms between windows. Spectral analysis is then performed on each window by taking the Discrete Fourier Transform (DFT), the results of which are then passed through a set of band pass filters to obtain the spectral feature of speech. The filters are set along the *Mel* scale. The number of filters was set at 24 because it simulates human ear processing. The MFCC coefficients are calculated by taking the inverse Fourier transform of the logarithm of the filter bank output. The effect of the logarithm scale is that it reduces the component amplitudes at every level.

The same methods were used for standard and thermal video data based on a modified version of Motion Templates [10]. Motion Templates use image subtraction based on intensity to give a single image to represent the series of lip movements within a sequence. This is done in three steps: first the regions of interest (ROIs) are normalized in size, for the second step the difference between consecutive frames is calculated, and in the third step the difference of frames are merged. Once the MTs are generated, 78 DCT coefficients are extracted from both standard and thermal video data for HMM training. The outputs of the HMMs were fused together using multiplication because the magnitudes of the HMM outputs could be greatly different making addition unsuitable.

IV. EXPERIENTS

A thermal-audiovisual (TAV) database was created comprising 11 participants filmed in thermal and standard video reading the words “zero” through “nine” repeated fifteen times each. This resulted in 150 samples from each speaker for a total of 1650 utterances. For each participant 14 samples of each word were used in HMM training with the remainder used for testing. The training for audio used clean data only whereas audio was tested with Gaussian white noise added at levels of 20, 30, 40, 50, and 60 dB SNR. This was repeated five times with different samples selected for testing and the results were averaged to reduce sample bias.

Experimental results showed an improvement of recognition rates for Tri-modal TAV ASR above both combined standard video and audio (AV) ASR and audio only ASR across all noise levels. This was calculated to be a relative average improvement of 11.8% over combined audio and standard ASR and a relative improvement of 39.2% on audio only ASR across all noise levels.

V. CONCLUSION

Automatic speech recognition (ASR) systems for consumer applications often operate in noisy and variable lighting conditions. Our proposed tri-modal thermal-audiovisual (TAV) recognition system uses thermal imaging to supplement audio and standard video modalities. While both standard video and thermal cues improve speech recognition in noisy environments, addition of the latter is particularly useful under different lighting conditions.

References

[1] K.H. Davis, R. Biddulph, and S. Balashek, “Automatic recognition of spoken digits”, *J Acoustical Society of America*, 24, 1950, pp. 627-642.

[2] L.E. Baum and T. Petrie, “Statistical inference for probabilistic functions of finite state markov chains”, *Annuals of Math. Statist.*, 37, 6, 1966, pp. 1554-1563.

[3] G. Potamianos, H. P. Graf, and E. Cosatto, “An image transform approach for HMM based automatic lipreading,” in *Proc. Int. Conf. Image Processing*, vol. I, Chicago, IL, Oct. 4–7, 1998, pp. 173–177.

[4] G. Potamianos, C. Neti, G. Gravier, A. Garg, and A.W. Senior, "Recent Advances in the Automatic Recognition of Audio-Visual Speech", *Proceedings of the IEEE*, 91, 9, Sept 2003, pp. 1-18.

[5] W.H. Sumby and I. Pollack, “Visual contribution to speech intelligibility in noise”, *J Acoustical Society of America*, 26, 1954, pp. 212-215.

[6] E. D. Petajan, “Automatic lipreading to enhance speech recognition,” in *Proc. Global Telecomm. Conf.*, Atlanta, GA, 1984, pp. 265–272.

[7] K. Murphy, *Hidden Markov Model (HMM) toolbox for Matlab* Retrieved May 25 2012, from http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html

[8] M. Brookes, Voicebox: speech processing toolbox for MATLAB. Retrieved May 25, 2012, from http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[9] C. Becchetti and L.P. Ricotti, Speech recognition: theory and C++ implementation Chichester, New York Wiley, 1999.

[10] W.C. Yau, Video analysis of mouth movement using motion templates for computer-based lip-reading. PhD thesis, RMIT, Australia, 2008.

Enabling a Healthy and Connected Home Based on Universal Plug and Play and Personal Health Devices

Danilo Santos, *Member, IEEE*, Angelo Perkusich, *Member, IEEE* and Hyggo Almeida

***Abstract*—This paper presents a UPnP (Universal Plug and Play) architecture definition that enables the communication between Personal Health Devices (PHD) and consumer electronic (CE) devices. The main contribution is the reference design, which allows PHDs with different connectivity interfaces to communicate with television sets and other UPnP enabled devices at home. The solution presented in this paper makes possible to current CE devices to communicate with new PHDs using a UPnP gateway without the need to make changes on the CE device. Also, we introduce a reference implementation based on open source software that makes possible an evaluation of the solution in a real home environment.**

***Index Terms*—e-Health, Personal Health Devices, UPnP, Consumer Electronics**

I. Introduction

Nowadays one of the most promising research and development area is the monitoring of health and wellness of human beings. The focus is on the use of the technology to monitor the daily health of a person, in order to create a direct impact in the health of a whole population [1] [2]. In this context a major current trend is the use of personal and portable devices to measure, process, and export health sensors data, the called e-health.

Some Personal Health Devices (PHDs) are equipped with local connectivity, such as USB, Bluetooth, NFC, or Wi-Fi. This local connectivity is used to export health measures to other devices. The health information is usually exported in order to be stored or processed by other ends. Organizations and groups are defining standards and solutions to enable an interconnected health environment, as well as to certify devices that are compliant with their standards and guidelines.

The main scenario from the CE industry is an end to end path where the health data collected by a PHD is exported to a local device, namely health manager, using standards such as those defined in the context of IEEE 11073 [3] and Bluetooth HDP (Health Device Profile)[1]. Such health data from the health manager is usually sent to a Personal Health Recorder (PHR) on the Internet.

However, the relationship between these personal health devices into a home environment is not well explored. Home and personal networks are already available and deployed in the end-user house. The interconnection between different consumer electronics devices, such as televisions, mobile computers, and smartphones, is a reality due open network protocols and standards such as UPnP [4].

This article presents a UPnP network infrastructure proposal that enables the use of PHDs in a home network, allowing seamless access of health information stored in Personal Health Recorders (PHR) from the home network. The architecture enables the interaction of multiple consumer electronics devices, such as television sets and personal computers, with PHDs and PHRs using the already deployed UPnP network. Also, at the end of the article a reference implementation of the proposed architecture is presented. This architecture promotes the realization of different use scenarios where the interaction between end-users and health/wellness information at home, thus, enhancing their life and wellbeing.

II. Technology Overview

Some technologies are enablers for a connected health scenario, and two standards are used, namely, Bluetooth HDP (Health Device Profile) and IEEE 11073 Family of Specifications. HDP is the Bluetooth profile used to transport health related information over standard Bluetooth links for both basic rate and enhanced data rate modes, known as BR/EDR mode. In a general way, IEEE 110073 is used to define the data format and the control flow between a health manager and a PHD.

On the other end, it is necessary to save user health information. One trend is to save this data using cloud services running Personal Health Recorders (PHR) instances. Some PHRs are already deployed and free to use on the Internet.

Considering the end-user point-of-view, UPnP is the way to go infrastructure to be used in home networks. UPnP uses standard internet protocols, such as UDP (User Datagram Protocol), HTTP (Hypertext Transfer Protocol), and SOAP (Simple Object Access Protocol). As a widely used standard, UPnP applications, more specifically, control point applications, are offered in a vast number of mobile devices and consumer electronic devices, such as television set, videogames, mobile phones, etc.

III. UPnP-Health Architecture Proposal and Scenarios

The proposed UPnP architecture defines one UPnP Device called HealthGateway. This UPnP Device must have both or at least one of the following UPnP Services:

- *UPnP LocalHealth Service*: connects directly to one PHD device and implemnets one UPnP action that lists available PHDs in the area. The service generates UPnP NOTIFY events when one PHD sends new measures. For example, the user make a new measurement using a Bluetooth enabled weighing scale; after receive the measure, the UPnP LocalHealth service sends an event with this new measure to UPnP interested control points. An extension of this service would enable to send control commands to PHDs using UPnP actions.
- *UPnP PersonalHealth Service*: connects to the user personal health database, that can be either local, in the same physical device, or a Personal Health Recorder in the internet. This service provides UPnP actions to retrieve latest measurements of biological signals of the user, to list the number of measures stored into the PHR, as also, to add new measures to the user database.

Besides these two new services, it is also recommended to add an extra UPnP Service called DeviceProtection [5].

The authors are with the Embedded Systems and Pervasive Computing Laboratory, Federal University of Campina Grande, C.P. 10105 - 58109-970 - Campina Grande - PB - Brazil, emails: danilosantos@copin.ufcg.edu.br, perkusic@dee.ufcg.edu.br and hyggo@dsc.ufcg.edu.br

[1] http://www.bluetooth.org

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

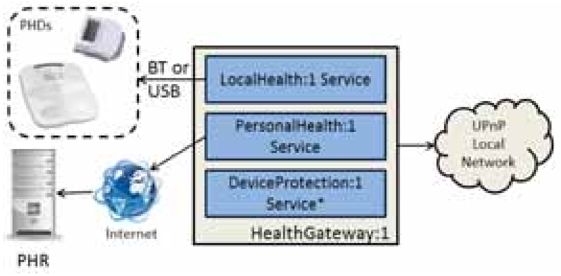


Fig. 1. UPnP HealthGateway Device and Services

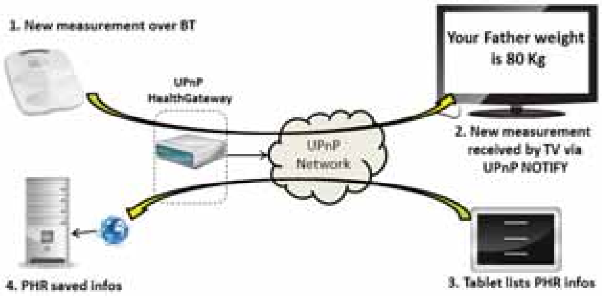


Fig. 2. UPnP Network with a HealthGateway Device

This service can be used to control and manage user access to the new health services. Health information is personal information that should be kept private. Therefore, the use of DeviceProtection service is recommended to add policy control rules to UPnP services based on user profiles. The Figure 1 illustrates the composition of this UPnP services in one UPnP device, and the interaction with other devices in the network.

For the scenario illustrated in Figure 2, the UPnP HealthGateway is a standalone device and it is connected with one PHD using Bluetooth. The HealthGateway is also connected to the Internet and is configured to access one pre-configured user information stored in a PHR. In the home network there is also available one television set with UPnP support, which has a control point application that makes possible to the end-user view his health information on the television set screen. Also, the end-user can view the same information using other devices in the home network, such as a tablet.

IV. Development and Evaluation

This section presents a description of a reference implementation for the proposed solution. The first version does not consider security and privacy features, and therefore, DeviceProtection Service is removed. The main objective is to validate the communication path and relationship between PHDs and the UPnP network.

The base platform is a personal computer. It has three main communication interfaces: Wi-Fi, which is used to communicate with the home network; and Bluetooth and USB that are used to communicate with PHDs.

For the software implementation, two main tools were used: The Brisa UPnP Framework [6], and the Antidote IEEE 11073 library[2]. Brisa UPnP is a well adopted solution with support implementations for multiple platforms using the Qt Framework. Antidote IEEE 11073 is the first and the only open-source implementation of IEEE 11073 specification. It is implemented totally in Ansi-C with a modular architecture, which makes it portable to different platforms. Also, one of the main reasons to use Antidote is its integration with BlueZ HDP and the operating system, which allows the comunicatin with Bluetooth Continua Health Alliance certified devices. Antidote also offers a daemon application (health-d) that communicates with other applications using D-BUS[3]. For the PHR representation is implemented a local database that stores all measures coming from PHDs. In future works this PHR representation can be linked directly with an external PHR.

For the evaluation test the devices used were: (i) a portable computer with a generic UPnP control point (which could represent a UPnP enabled television set); (ii) one Bluetooth HDP/IEEE11073 enabled weighing scale, (iii) and one Bluetooth HDP/IEEE11073 enabled blood pressure device. The overall scenario and communication between devices worked as expected, showing a good set of use cases for the proposed architecture.

One interesting point to analyze is how to define a common way to control PHDs using UPnP. During the implementation was noticed different ways to control PHDs, depending of manufactures and connectivity technologies (different ways to make Bluetooth Pairing, for example). This is a research action point to investigate in the future.

V. Conclusion and Future Works

In this paper we presented an architecture for a UPnP based network integrated with PHDs and PHRs. Also, it was described a reference implementation developed using open-source tools, such as Brisa UPnP Framework [6] and the Antidote IEEE 11073 implementation. The reference implementation validates the viability of the proposed architecture in a real environment.

The architecture introduced in this paper promotes the definition and implementation of a new set of use cases related with the interaction between the end-user, PHDs and consumer electronics devices, such as television sets, tablets, among others. Together with new use cases, new challenges were created, such as how to control PHDs using a standard UPnP interface.

As future work we envision the use and discussion of this proposal in the context of the UPnP Forum. UPnP Forum created the E-Health and Sensors (EHS) Working Committee, which is responsible to address the integration of personal health devices and sensors within UPnP. Although the group did not release any specification until the write of this article, the experience and base implementation presented here can be used as base for initial development.

VI. Acknowledgment

The authors would like to thank COPIN/UFCG and CAPES, Brazil, for the partial support in the development of this work.

References

[1] U. Varshney, “Pervasive healthcare and wireless health monitoring,” *Mob. Netw. Appl.*, vol. 12, pp. 113–127, March 2007. [Online]. Available: http://dx.doi.org/10.1007/s11036-007-0017-1

[2] J. Maitland, M. McGee-Lennon, and M. Mulvenna, “Pervasive healthcare: from orange alerts to mindcare,” *SIGHIT Rec.*, vol. 1, pp. 38–40, March 2011. [Online]. Available: http://doi.acm.org/10.1145/1971706.1971718

[3] *ISO/IEEE 11073-20601: Health informatics - Point-of-care medical device communication - Part 20601:Optimized exchange protocol Standards*, First edition ed., IEEE.

[4] UPnP, “UPnP Forum WebSite,” Last access: June, 2008, *http://www.upnp.org*.

[5] U. Forum. (2011) Device protection v 1.0. [Online]. Available: http://upnp.org/specs/gw/deviceprotection1/

[6] A. Guedes, D. Santos, J. Nascimento, L. Sales, A. Perkusich, and H. Almeida, “Brisa upnp a/v framework,” in *Proceedings of the International Conference on Consumer Electronics*. IEEE Press, 2008, pp. 1–3.

[2] http://oss.signove.com/

[3] http://dbus.freedesktop.org/

A Hierarchical Path Planning of cleaning robot Based on Grid Map

Hyoung-Ki Lee, WooYeon Jeong, Sujin Lee, and Jonghwa Won*
Samsung Advanced Institute of Technology, Samsung Electronics, Korea

***Abstract*— This paper presents a hierarchical path planning method based on the grid map. Hierarchy is composed of a coarse grid map and an original map with fine grid. The coarse grid map is obtained by reducing the resolution of the original grid map. A* path planning algorithm is applied to the coarse grid map instead of the fine grid map. It reduces memory size a lot for A* algorithm without increasing computation time. The proposed method is implemented for cleaning robots to move to a target position finding a collision free path.**

I. Introduction

Path planning is one of the important tasks for a cleaning robot because a cleaning robot should go back to a charging station after cleaning task is done or battery is getting low.

A grid map is usually used to plan an optimal path and it is obtained by the sensory information of the robot. Path planning algorithms based on a grid map is proposed by many researchers. Hierarchy that is composed of a coarse map and a fine map is mentioned in papers [1-6]. Hierarchical path planning is useful for fast computation [1] and avoidance of moving and/or stationary obstacles [2,4]. Voronoi diagram is used for path planning [3] and A* algorithm is generally used in papers [5-7]. The performance indices of a path panning algorithm can be computation time [1,7,9,10] and the length of a path [5,6,9].

Our cleaning robot, VC-RE70V, has an embedded ARM9 as a main processor and an external memory of 32MB. We tried to implement A* algorithm on the processor but it is found that memory size is not enough for the map of 18m × 18m size. It is known that A* algorithm requires 18 bytes per a grid. If a map is as large as 18m × 18m and the grid resolution is 1cm×1cm, then the memory needed for A* algorithm is 58.3 (=18×1800×1800) Mbytes. Therefore, reducing the memory size is the key to implementing a path planning algorithm for our cleaning robot.

In this paper we propose an algorithm to reduce the needed memory size by using the hierarchical path planning. A coarse grip map is generated by spatial sub-sampling and anti-aliasing filter. Path planning on this map leads to the minimization of the memory usage.

II. Hierarchical Path Planning with Map reduction

The proposed hierarchical path planning algorithm is based on the grid map and composed of three steps. Firstly, an original grid map with fine grids is reduced to a coarse grid map. Secondly a coarse shortest path is obtained on the coarse grid map by A* algorithm. Finally a fine shortest path is obtained by adjusting a coarse path to a fine grid map.

A. *Converting a fine map into a coarse map*

A cleaning robot builds the grid map of free space and obstacles using the information from the sensors such as ultrasound sensors, infra-red sensors, and bumper switches. A cleaning robot should a find path to move from a start position to a goal position while avoiding obstacles using the grid map. A* algorithm is widely used and efficient for memory usage for mobile robot applications. But, it is necessary to devise a new path planning algorithm to require less memory than the original A* algorithm for our cleaning robot.

As the first step of the proposed algorithm, an original grid map is transformed into a one in configuration space to avoid the collision of the mobile robot with obstacles in environment. It helps simplify the original path planning problem with arbitrary shaped obstacles considering the robot as a point. For example, when the robot is disc-shaped, a configuration space map can be obtained by applying the image thinning process to the grid map. The left figure of Fig. 1 shows an original gird map and the right figure is one in configuration space.

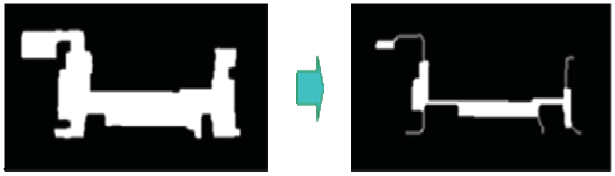


Fig.1 Configuration space for collision free path

The obtained map is converted to a coarse map by merging N×N girds into one grid. The left figure of Fig.2 shows the concept of sub-sampling using a simple spatial sub-sampling. But this results in the loss of path information since the path is disconnected at some regions as shown in the right figure of Fig.2.

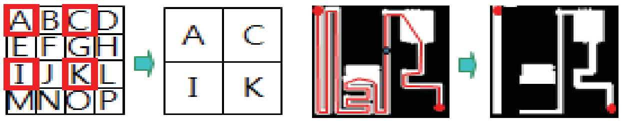


Fig. 2 Sub-sampling and loss of path information

To solve this problem, anti-aliasing filter sampling is used as shown in the left figure of Fig.3. The values of four grids are averaged into one value. This leads to the grey colored map from the black and white grid map. And a thresholding operation is used to create a binary map. It is assumed that obstacles with the color of black have a value of 0 and free spaces with the color of white have a value of 1. If an averaged grid value is larger than 0, then it is set to 1 otherwise it is set to 0. Consequently all the map information is preserved while reducing memory size to 1/(N×N).

* Corresponding Author

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

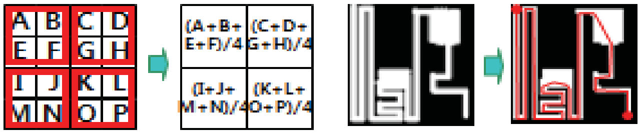


Fig. 3 Anti-aliasing filter sampling and thresholding

B. *Path correction on a fine grid map*

After finding a path on a coarse map using A* algorithm, some minor corrections are needed for obtaining a consistent path on the original grid map. The found path on a coarse map is mapped on the original grid map as shown in Fig. 4(a). In the figure, red, blue and green dots mean a start position, via positions, and a goal position, respectively. Some blue dots are in the obstacle region because of the mismatch of resolutions between the coarse map and the fine map. Wrong paths on the obstacle region are moved to nearest white region. A sky blue line with blue dots shows a corrected path belonging to free space as shown in fig. 4(b).

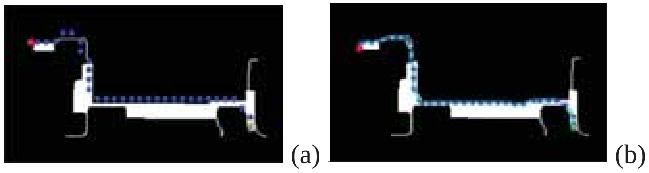
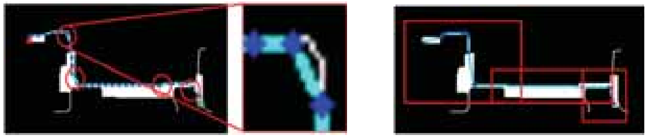


Fig.4 Fitting to nearest white space

Even after the above processing, there are some wrong paths which cross the obstacle region due to limited number of via points referring to red circles in Fig. 5(a). To resolve this problem, a local fine map including the wrong paths is retrieved and A* algorithm is applied to it. When the local map is retrieved, the maximum size of the local fine map is limited by the allowable memory. Red boxes in Fig. 5(b) show some retrieved local maps and blue dots represent locally corrected paths. As a result, A* algorithm is applied several times to a coarse map and several local fine maps.



(a) (b)

Fig. 5 Wrong path and corrected path

III. EXPERIMENTAL RESULTS

We implemented the algorithm in the cleaning robot of Samsung Electronics (product name: VC-RE70V). It's equipped with a camera and inertial sensors and has a function of SLAM (Simultaneous Localization And Mapping) processing ceiling images. It is known to be the world first consumer product with SLAM.

Fig. 6(a) shows a grid map of a residential house with its area 150 square meters. Fig. 6(b) is the configuration map of Fig. 6(a), where a light blue line represents a path generated in a coarse map and red boxes are retrieved local maps. Fig. 6(c) shows a corrected path and it is the same as the result from the fine global map considering all grids on the original map.

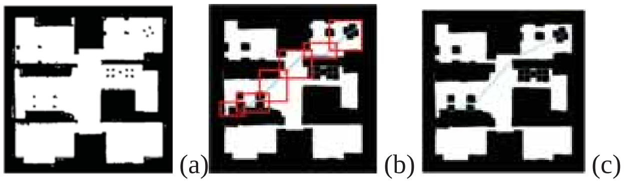


Fig. 6 Experimental results in a residential house

In Table I, we summarize the benefits of proposed hierarchical path planning algorithm in resource hungry embedded system such as cleaning robots. Memory requirement for computing A* algorithm and computation time are compared for the example map of Fig. 2 and 3, where the size of map is 18m × 10m and a grid resolution is 1cm×1cm.

TABLE I.
COMPARISONS OF MEMORY AND TIME

Method	Memory and Time
Original fine map	53.8MB, 258ms
The proposed(1/8 reduced map)	0.9MB, 253ms

IV. CONCLUSION

A path finding algorithm with efficient memory usage is proposed for the embedded system of cleaning robots. A* algorithm runs on a coarse map, which has a smaller size than the original grid map. Thus, the memory size needed for A* algorithm is saved while it doesn't increase the computation time so much. The method is implemented for cleaning robots of Samsung. As a result, cleaning robot having the embedded system of 32MB memory can find a collision free path in an indoor environment with arbitrary shaped obstacles.

REFERENCES

[1] Jae-Yeong Lee and Wonpil Yu, "A Coarse-to-Fine Approach for Fast Path Finding for Mobile Robots", Int. Conf. on Intelligent Robots and Systems, IEEE/RJS, pp. 5414-5419, 2009.

[2] Kikuo Fumimura and Hanan Samet, "A Hierarchical Strategy for Path Planning Among Moving Obstacles", IEEE trans. on Robotics and Automation, vol. 5. No. 1. Feb., 1989.

[3] Xiating Wang, Chenglei Yang, Jiaye Wang and Xiangxu Meng, "Hierarchical Voronoi Diagram-Based Path Planning Among Polygonal Obstacles for 3D Virtual Worlds", IEEE Int. Sym. on Virtual Reality Innovation, pp 175-181, 2010.

[4] F. Khorrami and P. Krishnamurthy, "A Hierarchical Path Planning and Obstacle Avoidance System for an Autonomous Underwater Vehicle", American Control Conference, pp. 3579-3584, 2009.

[5] Minhyeok Kwon, Heonyoung Lim, Yeonsik Kang, Changhwan Kim and Gwitae Park, "Hierarchical Optimal Time Path Planning for an Autonomous Mobile Robot using A* Algorithm," Int. Conf. on Control, Automation and System, pp 1997-2001, 2010.

[6] Adi Botea, Martin Muller and Jonathan Schaeffer, "Near Optimal Hierarchical Path-Finding" Journal of Game Development, vol. 1 Issue 1. 7-28, 2004.

[7] Charles W. Warren, "Fast Path Planning Using Modified A* Method", IEEE Int. Conf. on Robotics and Automation, vol. 2, pp 662-667, 1993.

[8] Iwan Ulrich and Johann Borenstein, "VFH+: Reliable Obstacle Avoidance for Fast Mobile Robots", Proc. of IEEE Int. Conf. on Robotics and Automation, vol. 2, pp. 1572-1577, 1998.

[9] Danny Z. Chen, Robert J. Szczerba and John J. Uhran Jr., "Planning Conditional Shortest Paths through an unknown Environment: A Framed-Quadtree Approach", Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and System, vol. 3, pp. 33-38, 1995.

[10] Subbarao Kambhampati and Larry S. Davis, "Multiresolution Path Planning for Mobile Robots", IEEE Journal of Robotics and Automation, vol. RA-2, No. 5 pp . 135-145, 1986.

Energy Efficient Basestation Operation with Traffic-Specific Energy Consumption

Jinhyeock Choi, Seonmin Jung, Junhyuk Kim, June-Koo Kevin Rhee, Byung Moo Lee, Jongho Bang, and Byung-Chang Kang

Abstract— **Basestation on-off algorithms for mobile-network energy saving beyond 4G can be simplified due to absence of interference in multi-user MIMO systems. We discuss an on-off algorithm based on specific energy and its energy savings performance.**

I. INTRODUCTION

The energy consumption of communication networks has become a major issue not only from its environmental concern but also from its economic impact [1]. Nowadays major service providers consume a few TWh per year [2] which results in significant electricity costs and serious burdens on operational expenditure (OPEX). Moreover, network traffic statistics shows an alarming growth trend [1], especially in the mobile sector. Cisco [3] forecasts that mobile data traffic will increase 26-fold between 2010 and 2015, resulting in 6.3 exabytes per month. On the other hand current technology development trend raises concerns over the ability of energy efficiency improvement to keep the pace with such rapid traffic explosion [1] [4].

Network energy efficiency can be achieved by various approaches [2]. First we may find a way to provide the same service with less traffic. For example, multicast or mobile content distribution access network would reduce the total traffic volume in the mobile network [5]. Second we may improve equipment efficiency to send the same traffic with less energy consumption. A basestation (BS) with more efficient RF power amplifier (PA) would need less energy to serve the same amount of data [2][6]. Third we may operate a network in a more energy efficient way to reduce power waste [2]. In general, current networks are provisioned for peak traffic out of diurnal traffic demand variation [7] and run in full capacity irrespective of traffic load. BSs typically use up to 90% of its maximum power while offering little services [8]. We may shutdown a cell by emptying such small amount of traffic to adjacent cells. Such traffic-adaptive operation is widely applied to data centers and mobile networks.

In this paper, we study management of BS energy consumption that constitutes the majority of mobile network energy consumption. On-off power management of BSs should meet the traffic demand of users even when some BSs are turned off, which leads to an energy optimization problem for minimizing power consumption. In [9], Gupta and Sing first considered network energy savings through rate adaptive operation. Some initial developments such as UMTS cell switching off or cooperative power reduction appeared in [10][11]. Lorincz *et al.* [7] formalized energy optimization problems with power, traffic, and network model, and applied ILP principles to select optimal network configurations in terms of power consumption, user demand satisfaction and guaranteed coverage. Oh *et al.* [8] described a real cellular network layout and traffic profile to estimate the potential of dynamic BS on-off operation. Greedy algorithms were proposed for dynamic BS switching considering energy-delay trade off [12], and for centralized and distributed operations [13]. Further investigation on cell size adjustment for power savings were studied in [14].

The contributions of this paper are summarized as below. First, we discover important energy efficiency metric of traffic-specific energy consumption, or the ‘specific energy’ in abbreviation, for a cellular network. The specific energy is defined as the average energy consumption per bit in a cell. Second, we formulate an energy optimization problem and utilize the notion of BS specific energy to propose a new greedy algorithm. Our new heuristic is based on the observation that BS specific energy varies according to the BS traffic rate. We evaluate its performance with computer simulation. The paper is organized as follows. In Sec. II, we present our reference model and introduce the specific energy metric. We formulate the optimization problem and propose a new greedy algorithm in Sec. III, and assess its performance by simulation in Sec. IV. We conclude the paper in Sec. V.

II. REFERENCE MODEL AND ENERGY EFFICIENCY METRIC

Energy efficiency of a wireless network can be assessed by the measure of energy per traffic (in the unit of Joule per bit) or power per throughput (W/bps). For example, if a BS consumed 1 kW while providing 100 Mbps throughput, the specific energy is 10 W/Mbps.

In our reference model, we set a service area A and time interval $S = [s_1 \leq s \leq s_2]$. Basestations BS_i $(1 \leq i \leq M)$ cover the area A and provide mobile service to users therein. Each BS_i has a corresponding power model $P(p_i^{\mathrm{tx}}(s))$ (measured in watt) which describes how much power BS_i consumes as a function of the transmit power $p_i^{\mathrm{tx}}(s)$. We assume that all BSs share the same power model.

In a multiuser MIMO cell for serving the downstream, transmit power from a BS to each user can be allocated independently from other users. Here, we adopt a multi-user

J. Choi, B.M. Lee, J. Bang, and B.-C. Kang are with Samsung Advanced Institute of Technology, Suwon, S. Korea, Email: jinchoe@samsung.com.

S. Jung, J. Kim, and J.-K.K. Rhee are with the Dept. of Electrical Engineering, KAIST, Daejeon, S. Korea, Email: rhee.jk@kaist.edu.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

MIMO zero forcing scheme as a pre-coding technique, which eliminates intra-cell user-to-user interference if the perfect channel information is known to the BS [15]. Inter-cell interference can be also eliminated if the channel information of the users on the cell boundary is shared with all BSs that reach the boundary. Hence, this model simplifies the user association model among BSs in the neighbor. Then, the transmit power of a BS is the total of transmit powers to individual users in the first-order approximation [15], and denoted as $p_i^{\text{tx}}(s)$.

A BS is sleep enabled and consumes a fixed power P_{sleep} during its sleep period. We define the sleep state vector $\delta_i(s)$ to represent the sleep state of BS_i in time interval $s \in S$ as below:

$$\delta_i(s) = \begin{cases} 1 & \text{if } BS_i \text{ is awake at time } s \\ 0 & \text{if } BS_i \text{ is aleep at time } s \end{cases} . \tag{1}$$

Each BS_i has traffic model $T_i(s)$ which measures its traffic demand as a function of time s. $T_i(s)$ is upper bounded by $T_{i,\max}$ which is the maximum throughput that BS_i can support. Unlike the power model, different BSs may have different traffic models. We can combine the power and traffic models to measure the power consumption $P_i(s)$ of BS_i at time s, and evaluate the network-wide total power consumption as

$$P_{\text{total}}(s) = \sum_{i=1}^{M} P_i(s) = \sum_{i=1}^{M} \delta_i(s) \cdot P(p_i^{\text{tx}}(s)) + (1-\delta_i(s)) \cdot P_{\text{sleep}} . \tag{2}$$

The specific energy of BS_i is defined as $P_i(s)/T_i(s)$. The specific energy of a mobile network depends on reference model which consists of i) deployment model which is about how a given area is covered by BSs, ii) power model which describes how each BS consumes energy while sending bits, and iii) traffic model which represents how much and in which pattern each BS sends traffic. It should be noted that even under the same deployment and power model, different traffic models may result in different traffic-specific energy values.

III. Problem Formulation and Green Greedy Algorithm

A. Problem Formulation

Let us consider users u_k for $1 \leq k \leq N$, randomly uniformly distributed in area A. We define $l_k(s)$ and $d_k(s)$ as the location and the traffic demand of u_k at time s, respectively. If a user is asleep or outside of the area A, we can simply assume its traffic demand is zero. We describe user u_k's association with BS_i by state matrix $x_{i,k}(s)$ as

$$x_{i,k}(s) = \begin{cases} 1 & \text{if } u_k \text{ is associated to } BS_i \text{ at time } s \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

We represent the transmit power from BS_i to u_k as $p_{i,k}^{\text{tx}}(s)$.

Assume that there is a separate control network that provides all the information of BSs (such as the power model and its location) and users (such as its location and traffic demand) at any moment. Then we can put some BSs into sleep mode and assign users to other active BSs in such a way that energy consumption is minimized and user traffic demand is met.

The energy optimization problem is formalized as follows:

Minimize $\quad P_{\text{total}}(s)$

s.t. during time interval $s \in S$,

$$\sum_{i=1}^{M} x_{i,k}(s) = \begin{cases} 1, & \text{if user } k \text{ is in service,} \\ 0, & \text{otherwise,} \end{cases} \quad \text{for all} \quad k \in \{1,..,N\} ,$$

and

$$\sum_{k=1}^{N} x_{i,k}(s) p_{i,k}^{\text{tx}}(s) = p_i^{\text{tx}}(s) \leq p_{i,\max}^{\text{tx}} , \text{ for all } i \in \{1,..,M\} . \tag{4}$$

The first and second constraints ensure that user u_k is attached to only one BS and that BS_i satisfies the maximum transmit power limit, respectively. The problem is NP-hard [16].

B. Green Greedy Algorithm

We can save energy by putting some BSs into sleep mode and controlling associations of users. However, the optimization problem is too difficult to tackle directly due to highly complex coupling of BS operation and user association. Here we present a heuristic greedy algorithm by the name of Green Greedy Algorithm (GGA). The traffic-specific energy of a BS, P/T, measures how energy-efficient the BS is. So it makes sense to turn on BSs starting from the most energy efficient one for given traffic demand distribution, which means the BS with the least specific energy. On the other hand it is our observation that the specific energy $P_i(s)/T_i(s)$ is not constant but varies according to the traffic demand $T_i(s)$. Hence we take a pre-estimated BS throughput into consideration and evaluate its specific energy.

We define $W_i(s)$ as the set of users which can be covered by BS_i at time s. Take notice that $W_i(s)$ is not the set of users that are actually associated to BS_i at time s.

Let $\sigma_i(s)$ be the minimum value between $T_{i,\max}$ and the sum of traffic demand $d_q(s)$ of users u_q in $W_i(s)$. $\sigma_i(s)$ represents the potential traffic demand which BS_i provides if it is turned on and all users in $W_i(s)$ are attached to BS_i at time s. We combine the $\sigma_i(s)$ and the power model P_i to estimate BS_i's instantaneous specific energy as $P_i(s)/\sigma_i(s)$ which represents how much power per bps BS_i will consume if turned on. Since BSs will be turned on from the one which uses the least specific energy, we formulate a greedy algorithm as follows:

Step1: Initialize B as the set of all basestations $\{BS_i \mid 1 \leq i \leq M\}$ and U as the set of all users $\{u_k \mid 1 \leq k \leq N\}$.

Step2: For all $BS_i \in B$, find $W_i(s)$ which is the set of all users that can be covered by BS_i at time s. A user u_q belongs to $W_i(s)$ if its location $l_q(s)$ lies within BS_i's coverage.

Step3: For all $BS_i \in B$, estimate its potential throughput $\sigma_i(s)$ if turned on as

$$\sigma_i(s) = \min[T_{i,\max}, \textstyle\sum_{u_q \in W_i(s)} d_q(s)]\cdot \qquad (7)$$

Step4: For all $BS_i \in B$, estimate its instantaneous specific energy $P_i(p_i^{\mathrm{tx}}(s))/\sigma_i(s)$ and find the BS with the smallest value as below

$$i^* = \arg\min_{BS_i \in B} \frac{P_i(p_i^{\mathrm{tx}}(s))}{\sigma_i(s)}. \qquad (8)$$

Step5: We form a subset $\overline{W}_{i^*}(s)$ of $W_{i^*}(s)$ to determine the users to assign to BS_{i^*}. $\overline{W}_{i^*}(s)$ is the set of users that are to be attached to BS_{i^*} and we try to make $\overline{W}_{i^*}(s)$ as large as possible. If the following condition holds,

$$\textstyle\sum_{u_q \in W_i(s)} p_{i,u_q}^{\mathrm{tx}} \le p_{i,\max}^{\mathrm{tx}}, \qquad (9)$$

we can assign all u_q in $W_{i^*}(s)$ to BS_{i^*}, and $\overline{W}_{i^*}(s) = W_{i^*}(s)$. Otherwise we include into $\overline{W}_{i^*}(s)$ first all the users which are exclusively served by BS_{i^*}, and then the remaining users selected randomly till BS_{i^*} can no longer serve users.

Step6: We turn on BS_{i^*} and assign it all u_q in $\overline{W}_{i^*}(s)$. We turn on the most energy-efficient BS, i.e., one with the least instantaneous traffic-specific energy, and try to assign it as many users as possible while sustaining QoS.

Step7: We reset B as $B \setminus \{BS_{i^*}\}$ and U as $U \setminus \overline{W}_{i^*}(s)$, return to **Step1**, and keep repeating the procedures till B = Ø or U = Ø. ■

Here, the BS inventory B becomes empty, i.e., B = Ø when all BSs are turned on, and U = Ø when all users are associated to some BSs. When algorithm terminates, if B ≠ Ø, the remaining BSs are kept turned off.

In practice, we cannot change BS on-off state and user association continuously. We may divide the time interval into sub-intervals and periodically run energy minimizing algorithm with expected user traffic demand. Or we may make an adjustment upon a certain network event such as a change of user traffic demand from a new user arrival.

IV. Simulation Study

A. System model

As for a system model, we consider multi-user MIMO with zero forcing, which is able to allocate individual transmit power levels from BS to active users. A BS has 3 sectors whose maximum transmit power is 40 W, and all sectors in a BS are simultaneously turned on or off for the sake of management convenience. We also assume that there is no additional cost so that a BS can be turned on or off immediately, and the user location information is provided by the combination of a MU-MIMO control system and a control network system with a separate channel.

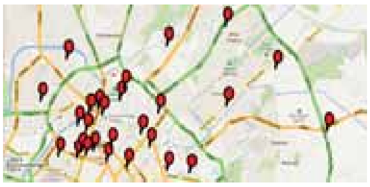


Fig. 1. Cellular network topology model in the Manchester area with 27 LTE-BSs, labeled in red call-out marks.

B. Reference model

We measure the energy efficiency of a mobile network, i.e., the specific energy under the following reference model.

1) Deployment model

As illustrated in Fig. 1, we use the data of the BS locations in the city of Manchester, UK, [17] to form a deployment model for a rectangular area of 5 km x 2 km. The service area is covered by 27 LTE BSs of one operator, of which cell radius is 1 km.

2) Traffic and user demand models

Users are assumed to be randomly located within the area in a uniform distribution for simplicity of the simulation. We assume that a user demand is 2 Mbps, and each sector of BS can provide capacity up to 100 Mbps. Hence, each sector can serve up to 50 users. As a result, 150 users can be served by one BS simultaneously in our model. Considering the population of Manchester, the density is 4,313/km^2. By taking into account the variation of user population in a day, the service area from the cellular network topology model, and the ratio of active users in urban area [18], we assume that the maximum number of active users considered in simulation is 1620 (i.e. 60 users per a BS). We model the number of active users during each period as shown in Table I [7].

TABLE I

Average user demand of users and number of active users.

Time (h)	% of active users	# of active users	Average user demand
(00~06)	15	243	2 Mbps
(06~09)	60	972	2 Mbps
(09~12)	100	1620	2 Mbps
(12~15)	70	1134	2 Mbps
(15~18)	85	1377	2 Mbps
(18~21)	55	891	2 Mbps
(21~24)	30	486	2 Mbps

3) Power model

The power model of BS is a function of traffic rate. We adopt a linear power model of $P_{\mathrm{fixed}} + \Delta_P * p_{\mathrm{tx}}$ with standby power P_{fixed} and traffic dependent part $\Delta_P * p_{\mathrm{tx}}$ as in [18]. P_{sleep} denotes the power in a sleep mode and $P_{\max}$ the maximum power associated with maximum transmission power $p_{\mathrm{tx,max}}$ in Fig. 2. The power model holds for all users within cell coverage.

We employ the Okumura-Hata propagation model to calculate a path loss which is the power decrease from BS to

	LTE-BS
$P_{\max}$	1350 W
P_{fixed}	712 W
P_{sleep}	12 W
$p_{\mathrm{tx,max}}$	40 W(1sector)
Δ_P	5.32 W/Mbps

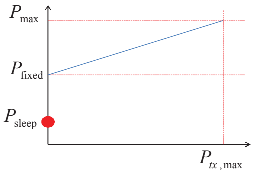


Fig. 2. Linear power model.

each user [19]. The Okumura-Hata parameters of our model for a 900-MHz frequency band are shown in Table II [20]. We assume that the user sensitivity is -104.5dBm for 2-Mbps user service [20]. In our simulation, transmit power of each user is derived by the user sensitivity and path loss.

TABLE II
Okumura-Hata parameters.

Parameters	Urban Indoor
Basestation antenna height (m)	30
Mobile antenna height (m)	1.5
Mobile antenna gain (dBi)	0.0
Slow fading standard deviation (dB)	8.0
Location probability (%)	95
Correction factor (dB)	0
Indoor loss (dB)	20
Slow fading margin (dB)	8.8

C. *Performance Evaluations*

Under the aforementioned reference model, we evaluate the performance of the proposed algorithm as follows. During a day, under the diurnal user demand, we compare the energy consumption in two different network operations. First, we do no management and run all BSs at all times. Second, we use the proposed algorithm to put BSs into sleep according to traffic load as in Table I.

Fig. 3 depicts how BSs operate with our algorithm, GGA. For different time periods, it designates the active BSs with the associated active users (connected with red line). The number of active BSs is different for traffic demand. Fig. 4 describes the instantaneous power consumption (W) and instantaneous traffic-specific energy (W/Mbps) for each case. We can see that the proposed algorithm achieves substantial energy savings. Especially during the low-traffic period, the current scheme uses up almost as much power as in peak time, hence resulting in high traffic-specific energy. Whereas, under our scheme, the power consumption drops down when traffic load does and the traffic specific-energy remains almost the same whole time time.

Compared with the current always-on operation, our algorithm reduced traffic-specific energy from 12.4 W/Mbps to 6.61 W/Mbps and improved the performance by 1.87 times. However, our work assumptions of a simple and ideal model with perfect knowledge of the user profile (such as its location and demand) and instant BS activation and deactivation allowed us for investigation of only the operation principle. Further work is needed for a realistic system operation such as consideration of cost of wake out of sleep state.

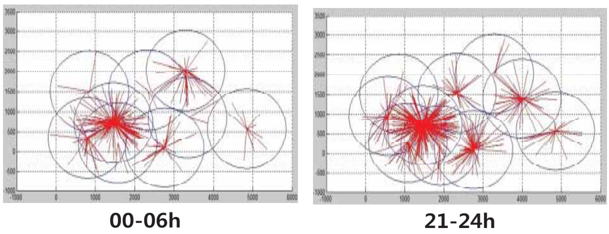

00-06h
21-24h

Fig. 3. Traffic adaptive BS operation with green greedy algorithm (GGA). The red lines indicate user-BS associations.

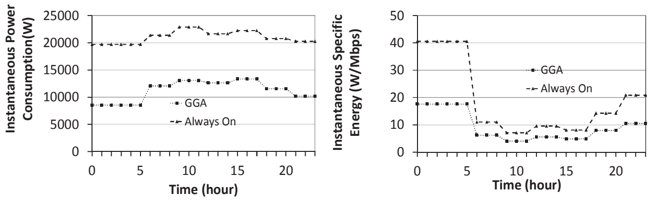


Fig. 4. Instantaneous power consumption and the corresponding instantaneous traffic-specific energy.

V. Conclusions

Energy optimized mobile network management can reduce energy consumption by switching off BSs according to traffic load conditions. We defined traffic-specific energy, i.e., the average energy consumption per bit of a BS, as the energy efficiency metric, and presented how such metric can be incorporated in a greedy algorithm called GGA to operate BS on-off power management. In performance simulations, we describe a way to evaluate specific energy under a reference model of network layout, diurnal traffic, and power profiles of BSs. Simulations showed that the proposed algorithm reduced energy per bit by 1.87 times.

Acknowledgment

The authors thank the members of the BCG2 project of Green Touch Consortium, especially Prof. Antonio Capone, Dr. Oliver Blume, Dr. Andre Santos, and Dr. Yan Chen for helpful discussion on energy efficiency of wireless network.

Reference

[1] D. Kilper, *et al.*, *IEEE J. on Selected Topics in Quantum Electron*., vol. 17, no. 2, pp.275-284, 2011.

[2] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, *IEEE Comm. Surveys & Tutorials,* vol.13, pp.223 - 244, 2011.

[3] Cisco, "Cisco visual networking index: global mobile data traffic forecast update, 2010–2015", http://www.cisco.com/.

[4] GreenTouch Consortium, http://greentouch.org/.

[5] K. Guan, G. Atkinson, D. Kilper, and E. Gulsen, GreenComm4, 2011.

[6] S. Vadgama and M. Hunukumbure, GreenComm4, ICC '11, 2011.

[7] J. Lorincz, A. Capone, and D. Begusic, *Computer Networks*, vol. 55, no. 3, pp. 514-540, Feb. 2011.

[8] E. Oh and B. Krishnamachari, IEEE GLOBECOM 2010, Dec. 2010.

[9] M. Gupta and S. Singh, ACM SIGCOMM'03, August 2003.

[10] L. Chiaraviglio, D. Ciullo, M.Meo, and M.A. Marsan, WPMC'08, 2008.

[11] M.A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, GreenComm09, ICC '09, 2009.

[12] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, *IEEE J. Selected Areas in Comm.*, vol. 29, Iss.8, pp.1525-1536, 2011.

[13] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, ACM MobiCom, Beijing, China, September 2009.

[14] S. Bhaumik, G. Narlikar, S. Chattopadhyay and S Kanugovi, ACM SIGCOMM'10, August 2010.

[15] B. C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, *IEEE Trans. Commun*., vol. 53, pp. 195–202, Jan. 2005.

[16] L. Alfandari, RAIRO Operations Research, vol. 41, pp. 83-93, 2007.

[17] "Sitefinder: Mobile phone base station database," ofcom, http://sitefinder.ofcom.org.uk/ .

[18] Earth Project, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown", https://www.ict-earth.eu/

[19] M. Hata, *IEEE Trans. on Vehicular Technol.*, vol.29, no.3, pp.317-325, 1980.

[20] LTE Encyclopedia, "LTE Radio Link Budgeting and RF Planning", https://sites.google.com/site/lteencyclopedia/lte-radio-link-budgeting-and-rf-planning

Multi-Stage Image Deblurring Using Long/Short Exposure Time Image Pair

Dong-bok Lee and Byung Cheol Song, *Senior Member, IEEE*

Electronic Engineering, Inha University, Incheon, Republic of Korea

Abstract--**This paper proposes a multi-stage de-blurring algorithm which iteratively estimates a blur kernel from a short and long exposure-time image pair, and de-convolutes the long exposure-time image by using the estimated blur kernel. Experimental results show that the proposed algorithm provides sharper details and smaller artifacts than the state-of-the-art algorithms.**

I. INTRODUCTION

Image de-blurring can be generally categorized into two types: single-image de-blurring and multi-image de-blurring. In the single image de-blurring, unknown blur kernel and latent image are estimated and reconstructed from a single blur image [1-3]. Some have used the fact that de-blurring can benefit from consecutive multiple blur images [4]. Another approach was to reconstruct a single de-blurred image from blurred/noisy image pair instead of consecutively blurred images [5-6]. Most of the multiple-image de-blurring algorithms assume that the input images are exactly aligned. So, if local objects actually have large motion between the two images, those algorithms may not work.

In order to solve the above-mentioned registration problem, this paper presents an advanced registration scheme which produces accurate motion fields by applying hierarchical motion estimation (ME) and overlapped block motion compensation (OBMC) to the noisy/blurred image pair. Subsequently, kernel re-estimation and residual de-convolution are performed. This processing is iterated until the acceptable de-blurred output is obtained. In addition, we propose a post-processing to reduce the visually annoying artifacts by adaptively employing original blurred pixels for low-light-level and flat areas.

II. PROPOSED ALGORITHM

Fig. 1 is the overview of the proposed algorithm which consists of three stages: Pre-processing, main-processing, and post-processing. Prior to the detailed description of each module, we define the long exposure-time (LE) image **B** and short exposure-time (SE) image **N** in matrix form as follows:

$$\mathbf{B} = \mathbf{Ik} \tag{1}$$

$$\mathbf{N} = \mathbf{I} + \mathbf{n} \tag{2}$$

where **I**, **k**, and **n** stand for unknown latent image, blur kernel, and noise, respectively.

A. Pre-processing

In a dim-light environment, **N** is very noisy and dark due to insufficient intensity of radiation. So, we equalize the intensity level of **N** with that of **B** by exploiting a typical histogram matching. Subsequently, the histogram-matched **N** is de-noised by a state-of-the-art algorithm to remove noise factors without loss of sharpness. Here, we employ a famous BM3D (block matching and 3-D filtering) [7]. Let $\mathbf{N}_D$ be the histogram matched and de-noised **N** image.

Generally, it is difficult to find an accurate motion vector (MV) field from registration between **B** and the relatively sharp image $\mathbf{N}_D$. So, we artificially blur $\mathbf{N}_D$ and use it for initial image registration. In order to derive the initial blur kernel, we adopted a simple blur kernel estimation, i.e., Cho's algorithm [2]. Let $\mathbf{N}_D^b$ be the blurred $\mathbf{N}_D$.

B. Image registration

At the first iteration, the MV field is obtained from **B** and $\mathbf{N}_D^b$. In order to minimize the MC error and artifacts, we employed a hierarchical ME. First, the MV for an overlapping $M\times M$ matching block is searched by a full search, and the selected MV for the 16×16 block is allocated to its central 4×4 block. For artifact-free MC, we adopt the so-called OBMC, exploiting the four-connected neighboring MVs. In this paper, the OBMC is performed on a 4×4 block basis. Note that even though the ME is performed between **B** and $\mathbf{N}_D^b$, the MC blocks are derived from $\mathbf{N}_D$. Finally, we can obtain the sharp reference image $\mathbf{I}_R$ from $\mathbf{N}_D$. At further iterations, the same image registration is performed between sharper LE and SE images, that is, **I'** and $\mathbf{N}_D$. Here, **I'** is the de-blurred output image of the main-processing stage as shown in Fig. 1.

C. Kernel re-estimation

Next, we re-estimate the blur kernel using both $\mathbf{I}_R$ and **B**. Note that as $\mathbf{I}_R$ becomes closer to an original latent image, the re-estimated kernel may be more similar to its original kernel. Let **k'** be the estimated kernel. We can derive the best **k'** via minimization process of Eq. (3).

$$\mathbf{k}' = \min_{\mathbf{k}} \left\| \mathbf{Ik} - \mathbf{B} \right\|^2 + \lambda \left\| \mathbf{k} \right\|^2 \tag{3}$$

This research was financially supported by the Ministry of Knowledge Economy (MKE) and the Korea Institute for Advancement of Technology (KIAT) through the Human Resource Training Project for Strategic Technology, and was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012000446).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

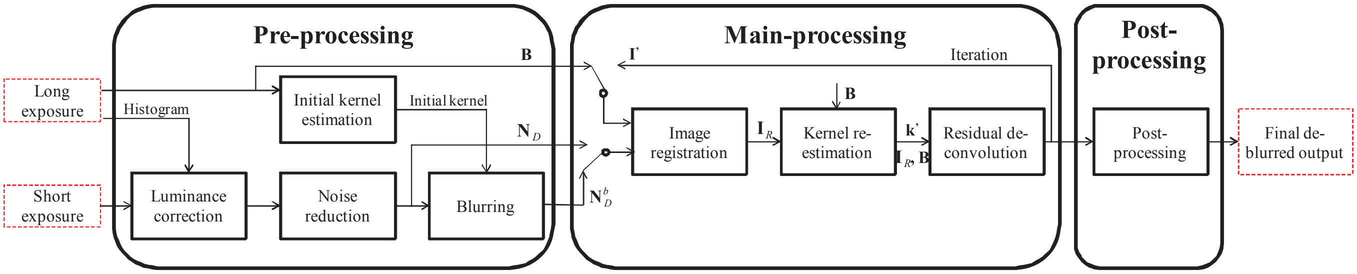


Fig. 1. The overview of the proposed algorithm.

In Eq. (3), Tikhonov regularization is employed to find a stable solution, and λ is set to 5. In order to solve Eq. (3), we used a so-called conjugate gradient (CG) method.

D. Residual de-convolution

We represent the MC error, i.e., latent residue as $\triangle\mathbf{I}$. Then, the latent image can be re-defined by Eq. (4).

$$\mathbf{I} = \mathbf{I}_R + \Delta\mathbf{I} \tag{4}$$

Applying the blur kernel **k**' to Eq. (4), $\triangle\mathbf{B}$ is derived from the following equation:

$$\Delta\mathbf{B} \equiv \Delta\mathbf{I}\mathbf{k}' = \mathbf{B} - (\mathbf{I}_R\mathbf{k}') \tag{5}$$

From Eq. (5), $\triangle\mathbf{I}'$ is derived via de-convolution. Here, we employed a simple de-convolution algorithm using Gaussian prior. Finally, the reconstructed image is obtained via $\mathbf{I}' = \mathbf{I}_R + \Delta\mathbf{I}'$. For more accurate de-blurring, this main-processing is iterated as shown in Fig. 1.

E. Post-processing

Low-light-level and flat areas in $\mathbf{I}_R$ often show some artifacts because of limited de-noising performance or inaccurate registration. Such artifacts can be propagated to **I**'. In order to avoid this phenomenon, we replace the pixels in the low-light-level and flat areas of **I**' with original blurred ones.

III. Experimental Results

We acquired several test images (2448x1376) by using Sony α55V camera under a dim-lighting condition, e.g., outdoors after sunset. The exposure time and ISO for **B** were set to 1 sec and 100, while those for **N** image were set to 1/100 sec and 1600. Also, motion search range was set to ±64. We compared the proposed algorithm with two state-of-the-art algorithms: Cho's [2], and Xu's algorithms [3] in terms of subjective visual quality. Note that all the processing was performed on YUV domain. For example, Fig. 2 shows that the proposed algorithm provides clearer texture for the tree and sharper edges for the bench than Cho's and Xu's algorithms. Actually, the computation time of the proposed algorithm is comparable with Xu's algorithm because we employ fast block matching algorithm for image registration.

IV. Conclusions

This paper proposed a de-blur algorithm which reconstructs a single high quality de-blurred image from a short-exposure and long-exposure image pair. The proposed algorithm could improve the accuracy of estimated kernel via precise registration using hierarchical motion estimation and overlapped block motion compensation. Also, the proposed residual de-convolution effectively suppressed ringing artifacts. The experimental results show that the proposed algorithm provides clearer details and smaller artifacts than the state-of-the-art algorithms.

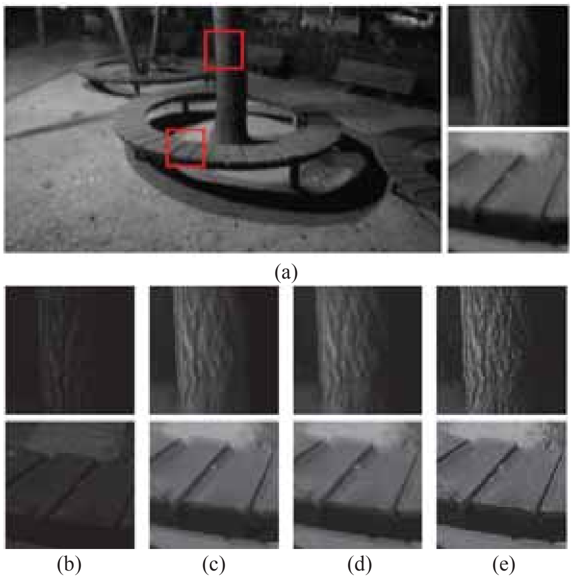


Fig. 2. De-blurring results for a real blurred image. (a) The blurred LE image (b) the noisy SE image (c) Cho's (d) Xu's (e) the proposed algorithm.

References

[1] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion de-blurring from a single image," *ACM Trans. Graphics*, vol. 27, no. 3, August 2008.

[2] S. Cho and S. Lee, "Fast motion de-blurring," *ACM SIGGRAPH ASIA*, 2009.

[3] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," *Proc. ECCV*, 2010.

[4] W. Li, J. Zhang, and Q. Dai, "Exploring aligned complementary image pair for blind motion de-blurring," *Proc. CVPR*, 2011.

[5] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image de-blurring with blurred/noisy image pairs," *ACM Trans. Graphics*, vol. 26, no. 3, 2007.

[6] S. Sato, Y. Okada, and T. Azuma, "Blur-free high-sensitivity imaging system utilizing combined long/short exposure green pixels," *Proc. ICCE*, 2012.

[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image de-noising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080-2095, August 2007.

A Cross-Channel Bilateral Filter for CFA Image Denoising

Yong Min Tai, Young-Su Moon, Junguk Cho, Shihwa Lee
SAIT, Samsung Electronics, Korea

***Abstract*— High ISO image which is captured at low light environment has serious sensor noise. To avoid this problem, we propose a novel denoising framework for CFA image. We create cross-channel correlation map and explore method to denoise with the map to distinguish the noise from signal. And we propose the modified bilateral filter algorithm with the map. Experimental result show that proposed method has a good balance between denoising noise and edge preservation in high ISO image.**

I. INTRODUCTION

Color images are usually acquired by digital cameras using a single sensor on which a color filter array (CFA), such as the Bayer pattern [1]. Because of using CFA, the image processing pipeline has a demosaicking stage to interpolate the sub-sampled CFA data to full-color image. However, it is challenge to restore the noise-free image from noisy sensor data. Especially the high ISO image is captured at low light environment and signal has very small intensity. After amplifying the image, it is hard to distinguish signal and noise.

Many denoising method have been developed over the years, the bilateral filter [2, 3] is one of the approaches. It is popular denoising algorithm with sigma filter [4] and SUSAN filter [5]. The bilateral filter has proven to be useful, because it is easy to implement. It preserves edges while noise is averaged out when noise level was estimated well. However, the high ISO image captured at low light environment is hard to estimate noise level. The intensity of noise is similar to intensity of edge in the high ISO image. Although the bilateral filter is being used widely, there is not a clear theory on selecting noise level in the high ISO image.

In this paper, we propose a denoising algorithm based on bilateral filter for the high ISO image. The proposed algorithm has framework using multi-resolution motivated by work of Ming Zhang et al. [6]. In our work, we explore novel method to suppress the visually unpleasant artifacts in high ISO image using modified bilateral filter. The modified filter algorithm suppresses the noise with cross-channel correlation map to distinguish the noise from signal. The rest of the paper is organized as follows. In Section II, our denoising method is proposed. Experimental results are put in Section III. Finally, conclusion is given in Section IV.

II. PROPOSED ALGORITHM

A multi-resolution denoising framework is effective to get rid of the coarse-grain chrominance noise [6]. Thus, proposed Denoising framework is illustrated in Fig.1 is using multi-resolution denoising with Laplacian pyramid [7] at every denoising modules.

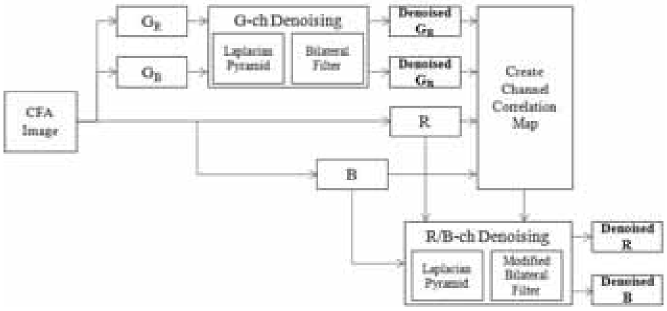


Fig. 1. Illustration of the proposed denoising framework

First, CFA image from sensor is decomposed into each color channels. G_R and G_B are denoised before R and B denoising. The cross-channel correlation map is created by denoised G_R, G_B with R and B. Lastly, R and B are denoised with the cross-channel correlation map. We explain the specific denoising process below.

A. G channel Denoising

The multi-resolution filtering based on [6] is applied to the G_R and G_B channel, but it has a little difference at multi-resolution scheme. The input image is decomposed each levels with Laplacian pyramid [7]. And bilateral filtering is applied to the image of each level before the image is reconstructed. At a pixel location p, the output of the bilateral filter is calculated as follows:

$$\hat{I}(p) = \frac{1}{C}\sum_{p' \in N(p)} e^{\frac{-|I(p')-I(p)|^2}{\alpha\sigma_r^2}} e^{\frac{-\|p'-p\|^2}{2\sigma_d^2}} I(p') \qquad (1)$$

Where σ_d and σ_r characterizes the intensity and spatial domain variance, respectively. These are parameters controlling the fall-off the weights in each domains. N(p) is a spatial neighborhood of p and C is the normalization constant. We use α to control denoising strength for denoising process on G channel image.

B. Cross-Channel Correlation Map for R and B channel Denoising

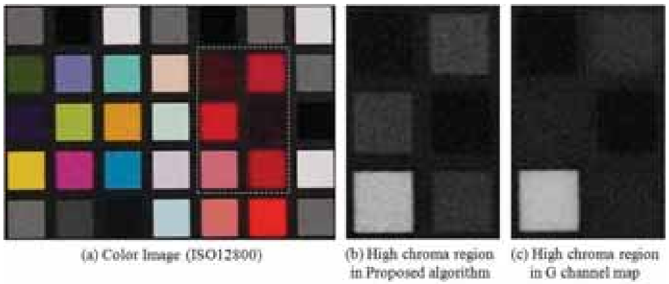
R and B channel image from CFA are too noisy to distinguish edge and noise in high ISO level. We propose the cross-channel correlation map for R and B channel denoising efficiently. The proposed map is mainly made from denoised G-ch image for reference of R (or B) channel denoising. However, Very low intensity of G-ch image in high chroma region can’t express an edge well (Fig. 2 (a)). Thus, proposed map uses R and B channel only if the pixel in high chroma region.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

The example of cross-channel correlation map ($I_{\text{Ref\_R}}$) for R channel denoising as in (2).

$$I_{\text{Ref\_R}} = \begin{cases} I_{G^{NR}_{R\_pos}}, & (\left|I_R - I_{G^{NR}_{R\_pos}}\right|) < \delta \textit{ and } (\left|I_{B_{R\_pos}} - I_{G^{NR}_{R\_pos}}\right|) < \delta \\ \frac{1}{N}\left(W_G I_{G^{NR}_{R\_pos}} + W_R I_R + W_B I_{B_{R\_pos}}\right), & \text{otherwise} \end{cases} \quad (2)$$

Where $I_{G^{NR}_{R\_pos}}$ and $I_{B_{R\_pos}}$ are G-ch intensity and B-ch intensity by interpolation at R pixel position, respectively. W_G, W_R and W_B are weight of each channels and the normalization constant N is $W_G+W_R+W_B$. This equation shows difference between G channel and other channel are used as criterion for high chroma region. The cross-channel correlation map is defined as weighted sum of all channels when the difference values are larger than threshold. Here, we should note that the weight of G channel (W_G) should larger than others. G channel value has less noise and fine detail. In Fig. 2, It is hard to find the edge in high chroma region at the map only consists of G channel while the map of proposed method could express edge in high chroma region.



(a) Color Image (ISO12800)

(b) High chroma region in Proposed algorithm

(c) High chroma region in G channel map

Fig. 2. Comparisons of cross-channel correlation map. Input image(12bit, CMOS) is taken at 50 lux and ISO12800 with $W_G : W_R : W_B = 6 : 1 : 1$; $\delta = 120$. (a) is an input image. (b) and (c) are high chroma region in the map by proposed algorithm and the map only consists of G channel, respectively (dotted box in input image).

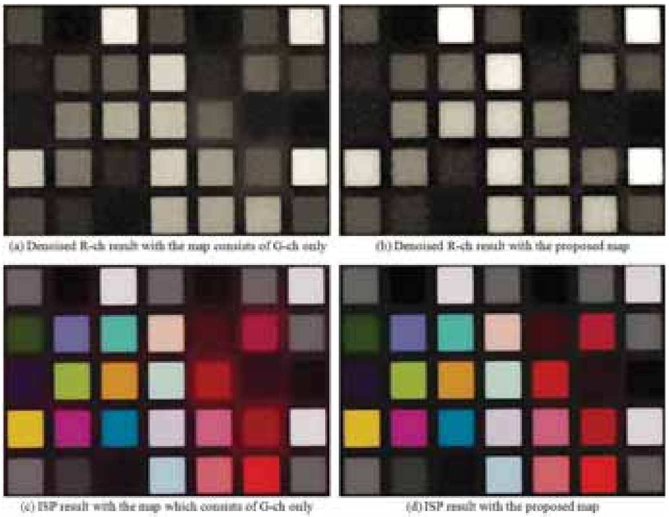
It is similar to make cross-channel correlation map ($I_{\text{Ref\_B}}$) for R channel denoising as follow:

$$I_{\text{Ref\_B}} = \begin{cases} I_{G^{NR}_{B\_pos}}, & (\left|I_B - I_{G^{NR}_{B\_pos}}\right|) < \delta \textit{ and } (\left|I_{R_{B\_pos}} - I_{G^{NR}_{B\_pos}}\right|) < \delta \\ \frac{1}{N}\left(W_G I_{G^{NR}_{B\_pos}} + W_R I_{R_{B\_pos}} + W_B I_B\right), & \text{otherwise} \end{cases} \quad (3)$$

We can see that the proposed map outperforms in R channel denoising is high chroma region in Fig.3. Denoised R channel image with the map which consists of G channel has Blurring in high chroma region (Fig. 3(a)). It will amplify in ISP pipeline and the result image of final ISP pipeline has serious artifact (Fig. 3(c)). We will denoise R and B channel image by using this map for reference of edge.

C. R and B channel Denoising with Adaptive Bilateral Filter

Directly denoising R and B channel is hard because of R and B channel image from CFA are too noisy to distinguish edge and noise. In this case, it is often to use the data which has the edge to be preserved in computational photography application [8][9].



(a) Denoised R-ch result with the map consists of G-ch only

(b) Denoised R-ch result with the proposed map

(c) ISP result with the map which consists of G-ch only

(d) ISP result with the proposed map

Fig. 3. Result of denoising R-ch image. The proposed method provides good noise suppression (b) than using only G-ch (a) in high chroma region. (c) and (d) are result images of full ISP chain of (a) and (b), respectively.

We define the new bilateral filter for R and B channel denoising by using the proposed map. This map characterizes that the edge is preserved and noise is suppressed. We use the map in the bilateral filter to modify distance of pixel intensity, $(I(p')-I(p))$. The proposed bilateral filter is defined as (4).

$$\hat{I}(p) = \frac{1}{C}\sum_{p' \in N(p)} e^{\frac{-\left|D_{\text{intensity}}(p')\right|^2}{\alpha\sigma_r^2}} e^{\frac{-\|p'-p\|^2}{2\sigma_d^2}} I(p') \quad (4)$$

Where $D_{\text{intensity}}(p')$ is the modified distance of pixel intensity, as in (5).

$$D_{\text{intensity}}(p') = \left(1 + \gamma_1\gamma_2\left(\frac{\text{th}_{\text{map}}}{\text{th}_R} - 1\right)\right)(I(p') - I(p)) \quad (5)$$

Where γ_1 and γ_2 are parameters controlling distance of pixel intensity. Both parameters have range between zero to one as defined in (6) where α_1 and α_2 are the fall-off slope in γ_1 and γ_2, respectively. γ_1 has non-zero value when distance of pixel intensity in R channel is larger than threshold(th_R). γ_2 has non-zero value when distance of pixel intensity in the map is larger than threshold(th_{map}). There is a peak noise or an edge on I(p') when γ_1 is non-zero. And I(p') is proved a noise pixel when γ_2 has non-zero. Using these two parameters, we can smooth only noise pixels with the edge to be preserved while the classical bilateral filter had problem with high ISO noise.

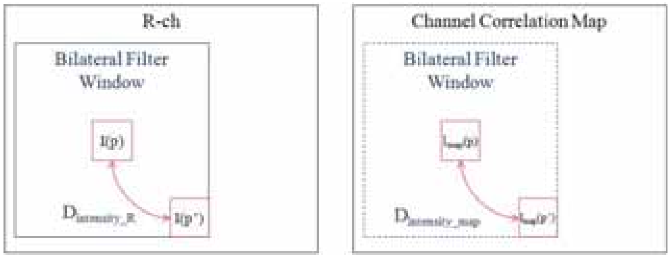


Fig. 4. Proposed bilateral filter on R channel denoising with cross-channel correlation map.

$$\gamma_1 = \frac{1}{1+e^{-\alpha_1(|D_{\text{intensity}\_R}|-\text{th}_\text{R})}}$$

$$\gamma_2 = \frac{e^{-\alpha_2(|D_{\text{intensity}\_map}|-\text{th}_\text{map})}}{1+e^{-\alpha_2(|D_{\text{intensity}\_\text{map}}|-\text{th}_\text{map})}} \tag{6}$$

Fig.4 illustrates an example of proposed bilateral filter on R channel denoising. If the difference between I(p) and I(p') - distance of pixel intensity - is larger than th_R in bilateral filter window, it will be considered as edge in classical bilateral filter. It makes that impulse noise remained in high ISO image with classical bilateral filter. The proposed bilateral filter uses not only the distance of intensity at R channel ($D_{\text{intensity\_R}}$) but also the distance of intensity at cross-channel correlation map ($D_{\text{intensity\_map}}$) which made in *Section B*. If $D_{\text{intensity\_map}}$ is smaller than th_map when $D_{\text{intensity\_R}}$ is larger than th_R, I(p') is classified the noise pixel. And the modified distance ($D_{\text{intensity}}$(p')) will decrease. The distance of intensity at cross-channel correlation map can be the effective criteria for noise and edge. Here, we should note that the threshold of R channel (th_R) should be larger than the threshold of channel correlation map (th_map).

The simple version of modified distance is defined as in (7). It is less complex than (5) to implement.

$$\text{when } \left(\left(|D_{\text{intensity\_R}}| > \text{th}_R\right) \text{ and } \left(|D_{\text{intensity\_map}}| < \text{th}_\text{map}\right)\right)$$
$$\hat{D}_{\text{intensity\_R}} = D_{\text{intensity\_R}} \times \frac{\text{th}_\text{map}}{\text{th}_R} \tag{7}$$

III. EXPERIMENTAL RESULTS

In this section, we have compared our method to some of the available denoising algorithm, and these are shown in Fig. 5. To ensure the fair comparison, we take the same 7x7 filter size of every filters and same noise variance in all the experiments. Especially, noise variance calculated from sensor noise modeling using the Skellam distribution [10]. Test CFA image was captured with ISO 12800 at 50lux environment.

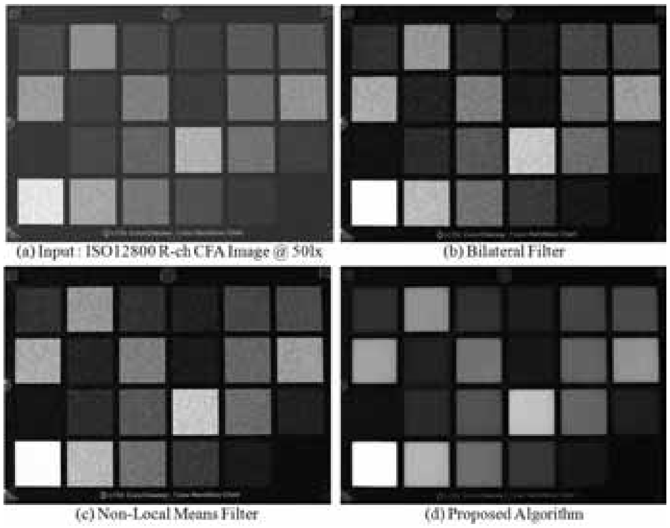
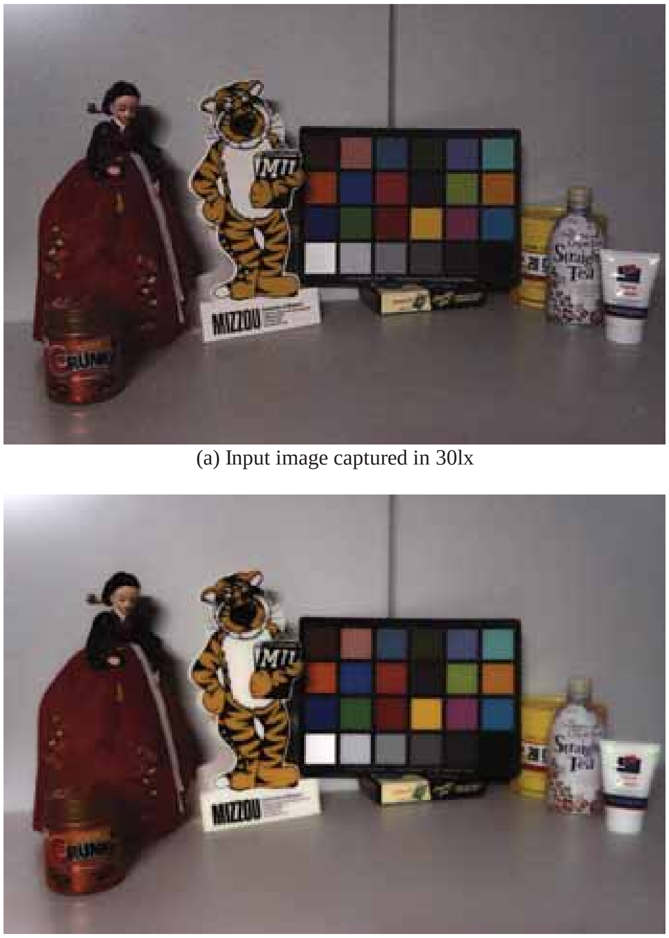

(a) Input : ISO12800 R-ch CFA Image @ 50lx
(b) Bilateral Filter
(c) Non-Local Means Filter
(d) Proposed Algorithm

Fig. 5. R channel image comparison of Bilateral Filter, Non-Local Means Filter and Proposed algorithm.

In Fig. 5, we compare the proposed algorithm (Fig. 5(d)) to Bilateral Filter [3] (Fig. 5(c)) and Non-Local Means Filter [11] (Fig. 5 (c)) in R channel of CFA image. It shows that the proposed algorithm has higher denoising quality in high ISO level image which was captured in low light environment (50lx). While proposed method has a good balance between image denoising and edge preservation, the results of bilateral filter and Non-Local Means filter have annoying artifact in flat


(a) Input image captured in 30lx

(b) Result of proposed method

Fig. 6. The example result of proposed method (30lx).

region (color patch region) as shown in Fig. 5. For computational purpose, the proposed algorithm is clearly less complex than Non-Local Means algorithm and similar with Bilateral filter.

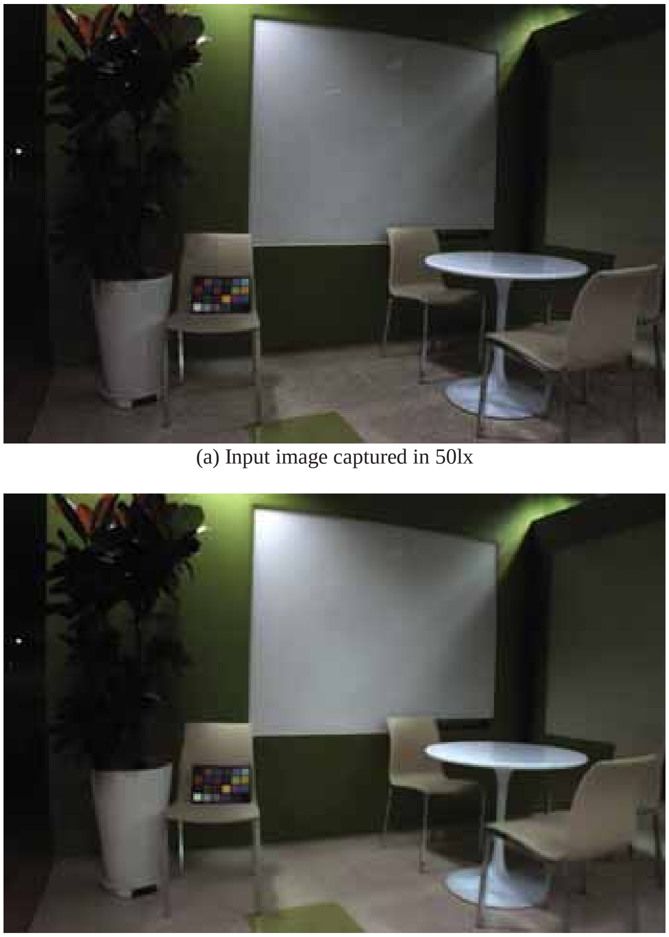
In Fig. 6 and Fig. 7, there are some results of proposed method in low light condition. Fig. 6 is captured in 30lx indoor condition and Fig. 7 is captured in 50lx indoor condition. Both images are captured with ISO 12800 setting. We can see that serious noise in original image is suppressed with edge preservation in result image of proposed method.

IV. CONCLUSION

In this paper, we proposed a robust and efficient bilateral filter algorithm for image denoising. The sensor noises can cause artifacts that are especially hard to remove in high ISO image which is captured at low light environment. Especially, R channel and B channel image from CFA are too noisy to distinguish edge and noise in high ISO level. The proposed algorithm uses cross-channel correlation map which uses mainly denoised G channel and adds R and B channel only if the pixel in high chroma region. This map characterizes that the edge is preserved and noise is suppressed enough and it is used in the bilateral filter to modify distance of pixel intensity of R and B channel. When R(or B) channel is denoised, the map can be the effective criteria for noise and edge. It is useful to remove maximum or minimum intensity noise in the high ISO CFA image.

The experimental result shows that the noise which is created in high ISO level image is well suppressed while edges are well preserved. Compared with other denoising algorithms, proposed algorithm has higher denoising quality in high ISO level image which was captured in low light environment.

The proposed denoising algorithm can be used not only CFA domain but also other domains. Further works include the development of image denoising based on the proposed algorithm on other domains.


(a) Input image captured in 50lx

(b) Result of proposed method

Fig. 7. The example result of proposed method (50lx).

REFERENCES

[1] B. E. Bayer, “Color imaging array,” U.S. patent 3 971 065, 1975.

[2] V. Aurich and J.Weule, “Non-linear gaussian filters performing edge preserving diffusion,” In Proceedings of the DAGM Symposium, 1995.

[3] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in Proc. Int. Conf. Computer Vision, 1998, pp. 839–846

[4] J. S. Lee, “Digital image smoothing and the sigma filter,” CVGIP: Graph. Models and Image Process, vol. 24, no. 2, pp. 255–269, Nov. 1983.

[5] S. M. Smith and J. M. Brady, “Susan—A new approach to low level image processing,” Int. Journal of Computer Vision, vol. 23, pp. 45–78, 1997.

[6] M. Zhang and B. Gunturk, “Multi-resolution bilateral filtering for image denoising,” IEEE Trans. Image Processing, vol.17, no. 12, pp. 2324 – 2333, Dec. 2008.

[7] Peter J. Burt and Edward H. Adelson, "The Laplacian Pyramid as a Compact Image Code," IEEE Trans. Communications, COM-3l, no. 4, April 1983

[8] E. Eisemann and F. Durand, “Flash photography enhancement via intrinsic relighting,” ACM Transactions on Graphics, 23(3), Proceedings of the ACM SIGGRAPH conference, July 2004.

[9] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, “Digital photography with flash and no-flash image pairs,” ACM Transactions on Graphics, 23(3), Proceedings of the ACM SIGGRAPH conference, July 2004.

[10] Youngbae Hwang, Jun-Sik Kim and In-So Kweon, “Sensor noise modeling using the Skellam distribution: Application to the color edge detection,” Computer Vision and Pattern Recognition (CVPR), 2007

[11] Antoni Buades, Bartomeu Coll and Jean-Michel Morel, "A non-local algorithm for image denoising," Computer Vision and Pattern Recognition(CVPR), 2005

A Fast Motion Deblurring Based on the Motion Blur Region Search for a Mobile Phone

Nam-Joon Kim, Sungjoo Suh, Changkyu Choi, Dusik Park, and Changyeong Kim
Samsung Advanced Institute of Technology, Korea

***Abstract*— In this paper, we propose a fast motion deblurring technique based on a motion blur region search for a mobile camera phone. The technique simultaneously captures two differently exposed images in one shot. By comparing edge information from short and normal exposed images, blur regions are identified and deblurring is only performed in the identified regions based on the proposed weighted color interpolation. Further, we modify an Android camera capture flow to remain in the required camera shot-to-shot time even including the proposed technique. Experimental results using an Android camera phone demonstrate the effectiveness of our proposed method.**

I. INTRODUCTION

After the advent of mobile camera phones, people can take a picture anytime and anywhere. However, it is not easy to take satisfactory photos with a mobile camera phone due to various artifacts. Among them, motion blur is one of the most common artifacts. In low light environments, a slow shutter speed is used to acquire enough light that reaches an image sensor, and it leads to blurred images usually introduced by a moving object or camera shake.

To restore the blurred image, multi-exposed images for the same scene have been used [1-3]. In [1], a gain-controlled residual deconvolution method was proposed to reduce ringing artifacts by estimating a blur kernel with two differently exposed images. In [2], the restoration problem was formulated as a MAP estimation problem by utilizing the degradation models of differently exposed images. However, these methods require iterative computations to restore the motion blurred image and thus, they are not suitable for mobile devices with limited computational resources. A non-iterative deblurring method using a multi-exposure camera system has been proposed in [3]. It gets the restored image by a motion-based image merging technique. Since this method is based on a global image registration, it might not work well for blurs caused by different directional movements of each object in the same scene.

In this paper, we propose a fast motion deblurring algorithm based on a motion blur region search. The proposed algorithm uses two differently exposed images of a normal exposed image (NEI) and a short exposed image (SEI) obtained during a single shot as shown in Fig. 1. A NEI is an image obtained by exposing during the exposure time automatically determined by a camera. A SEI is an image captured during an exposure time shorter than the exposure time automatically determined by a camera. Since human eyes are sensitive for the blur around the object boundaries with high frequency details, we search for blur regions around object boundaries. Then, the restored image is generated by performing the proposed weighted color interpolation for the identified motion blur regions. Since our deblurring method is restrictedly performed for blur regions around object boundaries, it is fast enough to be used for most mobile camera phones. Further, because the proposed method locally restores the blur based on the identified blur regions regardless of blur types, it can restore not only the blur resulting from camera shake but also the blur introduced by movements of each object in the scene.

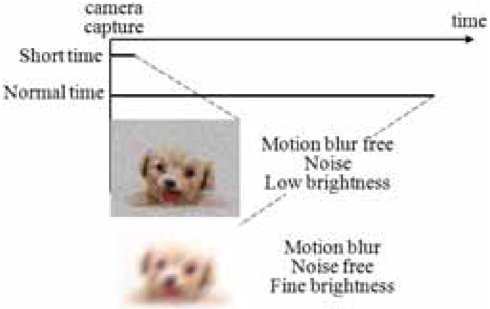
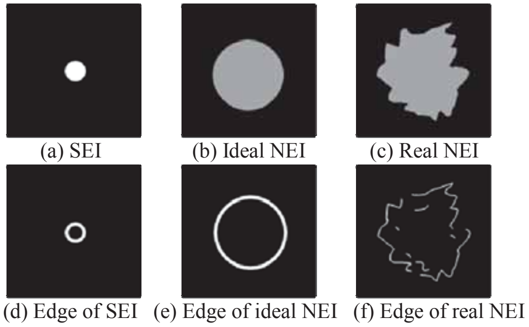


Fig. 1. The short and normal exposed images

II. THE PROPOSED SYSTEM

A. Motion Blur Region Search

To restore motion blur caused by camera shake or the movements of objects, the motion blur region should be identified. Since the SEI is robust to motion blur and the edges in the NEI is slightly shifted due to motion blur, the motion blur regions can be identified by searching for the regions between the SEI edge information and the NEI edge information where their edges are positioned at different locations. If the motion blur exists, a NEI has edges which are blurred and spread out from the SEI edges as shown in Fig. 2. When NEI edges are equally blurred in all directions as shown in Fig. 2(b), the blur region forms a closed path in the edge image as depicted in Fig. 2(e). If the blur happens with different amounts in each direction as illustrated in Fig. 2(c), some of edge information is lost and thus, the blur region does not form a closed path in the edge image but forms a disconnected path as shown in Fig. 2(f). This disconnected path can be identified by checking the connectedness in


(a) SEI (b) Ideal NEI (c) Real NEI

(d) Edge of SEI (e) Edge of ideal NEI (f) Edge of real NEI

Fig. 2. Motion blur and edge information for SEI and NEI

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

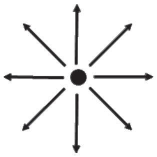


Fig. 3. Angular search direction for the connectedness

angular space as

$$\sum_{i=0}^{7} I_{\frac{\pi}{4}i}(x,y,r), \qquad (1)$$

where $I_{\theta}(x,y,r)$ is 1 if there is any edge within r pixels toward the θ direction for the (x,y) position in the NEI edge image as depicted in Fig. 3. If Eq. (1) is equal to 8, (x,y) belongs to the blur region. Thus, we check for the connectedness in angular space for all pixels in the NEI and identify the motion blur regions through Eq. (1).

B. Image Restoration by a weighted color interpolation

To restore the identified motion blur regions, we propose a weighted color interpolation approach using adjacent pixel values for a blurred pixel. The pixel value in (x,y) position of the identified motion blur region is restored by referring the adjacent pixels $(x+i,y+j)$ in the NEI which do not belong to the motion blur regions as

$$R(x,y) = \Big(\big(\sum_{i}\sum_{j}\big(NEI(x+i,y+j)\times w(i,j)\big)/\sum_{i}\sum_{j} w(i,j)\big) \times C_2 + HSEI(x,y)\times C_3\Big)/\big(C_2+C_3\big), \qquad (2)$$

$$w(i,j) = C_1 - (i^2 + j^2 + 1). \qquad (3)$$

where $HSEI(x,y)$ is the histogram equalized value of the SEI at (x,y) and $NEI(x+i,y+j)$ is the NEI pixel value at $(x+i,y+j)$ that is not in the motion blur region. C_1 is a constant value depending on the mask size. The weighted color C_2 and C_3 are some constant weight values depending on the region where (x,y) is located. Fig. 4 shows the motion blur region and SEI edge information. If there are pixel values in (a) or (e) which do not belong to the motion blur region adjacent to the blurred pixel at (x,y) in (b) or (d), neighboring pixel values from (a) or (e) can be used to restore the blurred pixel at (x,y). In this case, pixel values in the NEI are more weighted than pixel values in the histogram equalized SEI since the regions in (a) and (e) are not blurred and the NEI has small noise in (a) and (e) regions. Thus, the C_2 should be larger than C_3. If there is no pixel value which do not belong to the motion blur region adjacent to the blurred pixel at (x,y) such as regions (c) with high frequency details, pixel values in the SEI are more weighted than the NEI. So, for (c), C_3 is much larger than C_2. Finally, the blocking effect made around the boundary of the motion blur region during the motion deblurring process is suppressed by a 3×3 deblocking filter.

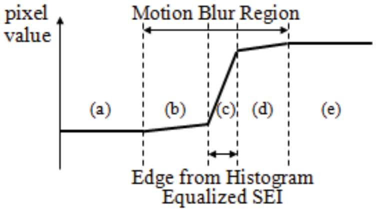


Fig. 4. Edge Region division on motion blurred image

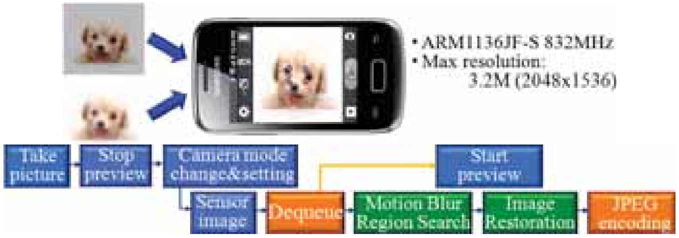
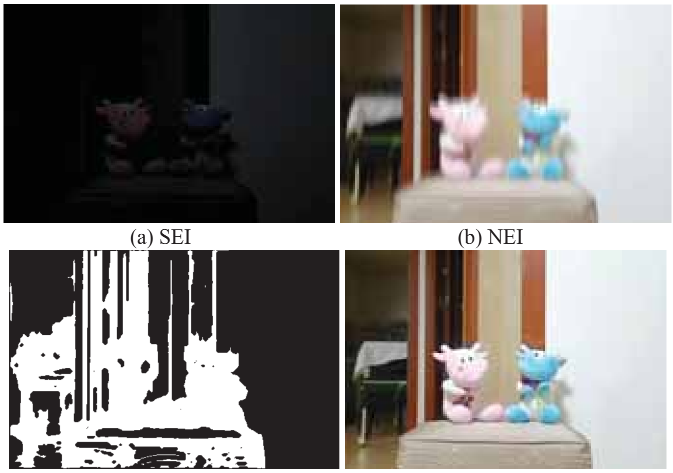


Fig. 5. Modified flow of Android camera's image capture flow

III. EXPERIMENTAL RESULTS

The proposed approach is implemented with the Samsung Android phone by modifying the Android camera capture flow as shown in Fig. 5. This phone should meet the shot-to-shot time in the specified requirement (2.5sec). The shot-to-shot time is the time from "Take picture" to "Start preview". In Fig. 5, after "Take picture", the camera sensor gives the SEI and NEI. To obtain the SEI and NEI during a single shot, the SEI and NEI were stored in the camera buffer. After "Dequeue", "Motion Blur Region Search" and "Image Restoration" are performed on AP followed by "JPEG encoding". The visual result for a real image is shown in Fig. 6.



(a) SEI (b) NEI

(c) Motion Blur Region (d) Restored Image

Fig. 6. Motion deblurring for a real image

IV. CONCLUSION

We have presented a fast motion deblurring approach to implement in the shot-to-shot requirement on an Android phone. The proposed method is applied to the Android camera structure as two separate parts. Although the image restoration time depends on the image complexity, real images are restored to meet the shot-to-shot time requirement and give a good restoration result.

REFERENCES

[1] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image Deblurring with Blurred/Noisy Image Pairs," in ACM Transactions on Graphics, vol. 26, no. 3, Article 1, July 2007.

[2] M. Tico, M. Vehvilainen, "Image Stabilization based on Fusing the Visual Information in Differently Exposed Images", Proceedings of ICIP, vol. 1, pp. 117-120, October 2007.

[3] B. D. Choi, S. W. Jung, and S. J. Ko, "Motion-blur-free camera system splitting exposure time," IEEE Transactions on Consumer Electronics, vol. 54, no. 3, pp. 981-986, August 2008.

Spatially Adaptive Antialiasing for Enhancement of Mobile Imaging Systems Using Combined Wavelet-Fourier Transforms

Eunjung Chae[1], Wonseok Kang[1], Eunsung Lee[1], Sangjin Kim[1], and Joonki Paik[1]

[1] Image Processing and Intelligent Systems Laboratory, Graduate School of Advanced Imaging Science, Multimedia, and Film, Chung-Ang University, Seoul, Korea

Abstract--**In this digest, we present a novel adaptive antialiasing method using combined wavelet-Fourier transforms for enhancement of mobile imaging systems. The proposed method can remove aliasing artifacts while preserving high-frequency details using frequency-domain analysis and adaptive shrinkage of wavelet subbands. The proposed algorithm consists of three steps; i) wavelet transform of the input degraded image, ii) analysis of Fourier transform coefficients and aliasing reduction in the LL subband of the wavelet transform, and iii) selective reduction of aliasing artifacts according to the classification between details and aliased components in the LH, HL, and HH subbands. Based on experimental results, the proposed algorithm can successfully remove aliasing artifacts while preserving high visual quality.**

I. Introduction

As the digital technology advances, the demand for high-resolution images keeps increasing. Mobile imaging devices such as smart phone cameras cannot provide high-resolution images because of the limited sensor resolution, computational power, and power consumption. For this reason, mobile images are provided in the form of down sampled version, which results in critical image degradation caused by aliasing artifacts and noise.

Over the past few decades, many aliasing reduction algorithms have been proposed. The conventional aliasing reduction methods generally used low-pass filters in either spatial [1] or frequency domain [2]. Although these methods are simple and fast, they cannot completely remove the aliasing artifact, and loses the high-frequency components which represent image details. Gun and Taubman have recently proposed a packet lifting method using wavelet transform [3][4]. Although it can effectively remove noise and aliasing artifacts, the resulting image is still blurred because of significantly suppressed high-frequency components.

To remove aliasing artifacts while preserving high-frequency components, we present a novel approach to analyzing aliased frequency components and adaptively shrinking wavelet coefficients using the combined wavelet-Fourier transform and edge-map generation. More specifically, the proposed antialiasing algorithm first performs the discrete wavelet transform (DWT) to decompose the input image into a set of band-limited components, called HH, HL, LH, and LL subbands. We use the Fourier transform of the LL subband for analyzing aliased components, and remove them using the notch filter. From the resulting aliasing-free LL band we compute the edge-map for the later use. Aliasing components in HH, HL, and LH subbands are removed using the edge-map of the LL band and adaptive wavelet shrinkage. The resulting enhanced image is obtained by the inverse DWT (IDWT).

This research was supported by Basic Science Research Program through National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (2009-0081059), by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2011 (00045420-1), and by the Technology Innovation Program, Industrial Strategic Technology Development Program(Development of SR(Super Resolution) Image Scaler for 4K UHD), under Grant K10041900.

II. Antialiasing Method Using Adaptive Wavelet Shrinkage Based On Combined Wavelet-Fourier Transform

Aliasing is a common problem in the image down sampling process. For removing the aliasing artifact while preserving high-frequency components, we present the adaptive wavelet shrinkage method based on combined wavelet-Fourier transforms as shown in Fig 1. The proposed method consists of three steps; i) wavelet transform of the input degraded image, ii) analysis of Fourier transform coefficients and aliasing reduction in the LL subband, and iii) selective reduction of aliasing artifacts according to the edge map of the LL band and appropriately classified LH, HL, and HH subbands.

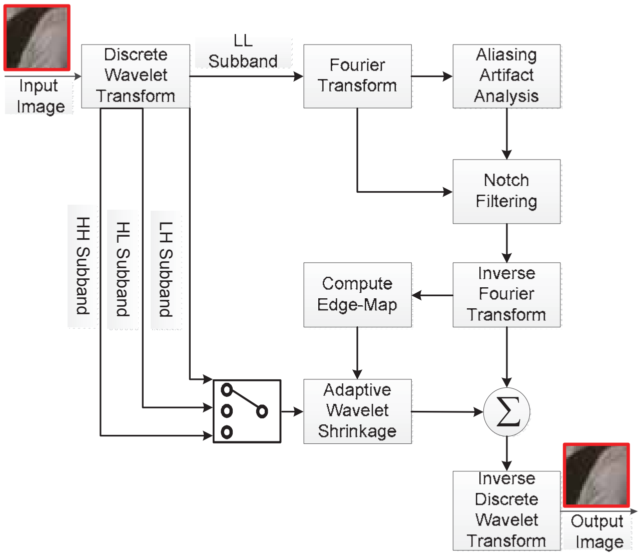

Input Image
Discrete Wavelet Transform
LL Subband
Fourier Transform
Aliasing Artifact Analysis
Notch Filtering
HH Subband
HL Subband
LH Subband
Compute Edge-Map
Inverse Fourier Transform
Adaptive Wavelet Shrinkage
Σ
Inverse Discrete Wavelet Transform
Output Image

Fig. 1. Block diagram of the proposed antialiasing method.

We first perform the one-level DWT to decompose the input image into four subbands including one low-frequency (LL) subband and three directional high frequency (HL, LH and HH) subbands.

We then analyze the power spectrum of the LL subband using the Fourier Transform, and selectively remove aliased components using a notch filter in the LL band. The resulting aliasing-free DWT coefficients in the LL subband provides correct edge information in the form of the local variance edge-map as

$$E(x, y) = \frac{1}{1+\sigma v(x, y)}, \quad (1)$$

where σ represents a control parameter that evenly distributes the edge values in $[0,1]$, and $v(x, y)$ the local variance of a patch centered at (x, y).

We classify wavelet coefficients in the LH, HL, and HH subbands into either “detail” or “aliased” components using the edge map estimated in the LL subband. For removing aliased components in the i -th subband, we shrink the coefficients as

$$\hat{W}_{\psi}^{i}(x, y) = W_{\psi}^{i}(x, y)\{1-E(x, y)\}, \text{ for } i \in \{H, V, D\} \quad (2)$$

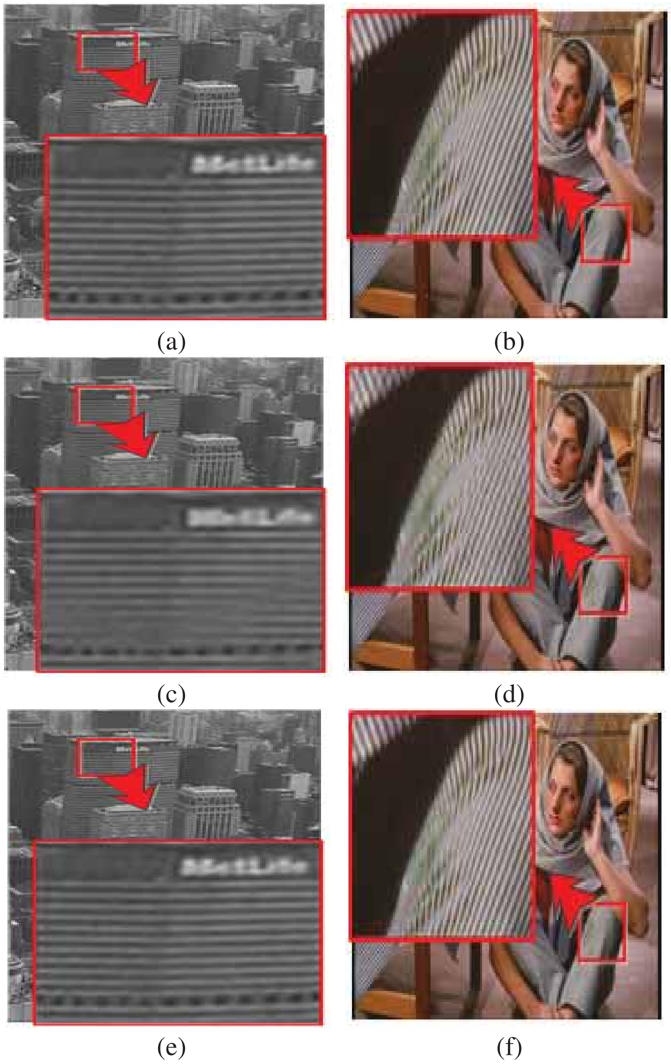
Finally, adaptively shrinked wavelet coefficients of each subband are inversely transformed to obtain the high-quality restored image.

III. Experimental Results

Fig. 2 compares results of aliasing reduction using the packet lifting and the proposed methods. The proposed algorithm gives better performance in the sense of both suppressing the aliasing and preserving high-frequency components compared with the existing state-of-the-art method.

IV. Conclusion

In this digest, we proposed a novel, spatially adaptive antialiasing method based on combined wavelet-Fourier transforms for enhancing the image quality of mobile imaging systems. The proposed method can remove aliasing artifacts while preserving high frequency information using the analysis of frequency components and adaptive shrinkage of wavelet coefficients. Experimental results demonstrated that the proposed algorithm can effectively remove aliasing artifacts while preserving higher visual quality compared with the existing state-of-the-art method. The proposed antialiasing method is suitable for not only mobile imaging systems but also commercial low-cost, high-quality imaging devices such as smart TVs, DSLR cameras, and scanners.



(a) (b)

(c) (d)

(e) (f)

Fig. 2. The experimental results of different antialiasing methods; (a,b) the original image, (c,d) packet lifting[4], and (e,f) the proposed result

Reference

[1] S. Martucci, “Image resizing in the discrete cosine transform domain”, IEEE International Conference on Image Processing, vol. 2, pp. 244-247, Oct. 1995.

[2] P. Vandewalle, S. Susstrunk, and M. Vetterli, “A frequency domain approach to registration of aliased images with application to super-resolution”, EURASIP Journal on Applied Signal Processiong, vol. 2006, pp. 233-247, Jan. 2006.

[3] J. Gan and D. Taubman, “A content-adaptive wavelet-like transform for aliasing suppression in image and video compression”, IEEE International Conference on Image Processing, pp. 3821-3824, Nov. 2009.

[4] J. Gan and D. Taubman, “Non-separable wavelet-like lifting structure for image and video compression with aliasing suppression”, IEEE International Conference on Image Processing, vol. 6, pp. 65-68, Oct. 2007.

[5] G. Abhayaratne, “Reducing aliasing in wavelets based downsampling for improved resolution scalability”, IEEE International Conference on Image Processing, vol. 2, pp. 898-901, Sep. 2005.

Application and Evaluation of Texture-Adaptive Skin Detection in TV Image Enhancement

Bahman Zafarifar, Erwin B. Bellers, and Peter H. N. de With, *Fellow* IEEE

Abstract*— This paper evaluates a case study where a previously reported texture-adaptive skin detection algorithm is applied for TV image enhancement. A color-only skin detector of an existing high-end TV chip is extended with a texture feature, enabling exclusion of skin-colored textured areas. We report the performance in terms of detection result, and in terms of image quality in a cascade of three image enhancement functions. In terms of detection score, at 80% true positive rate, the false positive rate of the texture-adaptive skin detector is 29% lower than that of the color-only skin detector, forming a clear improvement. With respect to its application in enhancement, we assess the enhancement quality by measuring the RMS error of the enhancement output compared to an optimally enhanced image based on ground-truth skin areas. When using the texture-adaptive skin detector, the enhancement RMS error is 44% lower than the RMS error when using the color-only skin detector, thereby confirming the applicability of the proposal. Subjective evaluation indicates that the proposed algorithm is better suitable for mid/high-frequency boosting applications like sharpness enhancement, and less suitable for enhancements that operate on low frequencies like color correction functions [1].

I. INTRODUCTION

In TV image enhancement, obtaining a good image quality and natural look in human skin areas is of primary importance for the overall viewing experience and acceptation. Without skin-specific treatment, image enhancement can result in unnatural colors, and excessive sharpness and contrast in skin areas. Such image impairments are easily noticed, since the viewer has prior knowledge about the appearance of skin.

In currently available TV video processing chips[2], skin detection is performed by examining the pixel color in 2D or 3D color spaces. Such color-only skin detectors are prone to erroneous detection of skin-colored non-skin areas (false positives), leading to unintended or suboptimal enhancement of misclassified areas.

In [2], we have presented the details of a skin detection algorithm that extends a skin color feature with a color-adaptive texture feature, and hereby removes textured areas from the skin map (less false positives). In this paper, we compare the performance of this texture-adaptive skin detector versus a color-only detector in terms of skin detection result and its application in image enhancement.

We use three image enhancement functions of a high-end TV chip[2]; skin-tone correction, sharpness enhancement, and local contrast enhancement. These enhancement functions require a skin map for controlling their enhancement level in skin areas. Enhancement quality is assessed by measuring the RMS error of (1) the enhancement output obtained by executing the enhancement functions when using the skin map of our choice (e.g. the color-only skin detector), versus (2) an optimal enhancement output obtained by executing the enhancement functions when using a manually annotated skin map (ideal skin detector).

Our objective measurements indicate an overall improvement in skin detection and skin-dependent enhancements, when using the texture-adaptive skin detector. In subjective assessment, we have found that the proposed skin detector is most suitable for enhancement functions that boost mid/high-frequency image contents, e.g. sharpness enhancement. However, near skin-colored object boundaries, having a slow gradient in the skin map, render the texture-adaptive detector less suitable for enhancements that modify the DC level of the image, e.g. global color corrections.

In the following paragraphs, Section II describes the skin detection algorithm, Section III introduces the enhancement functions, Section IV presents the objective measurement method and the results, and Section V concludes the paper.

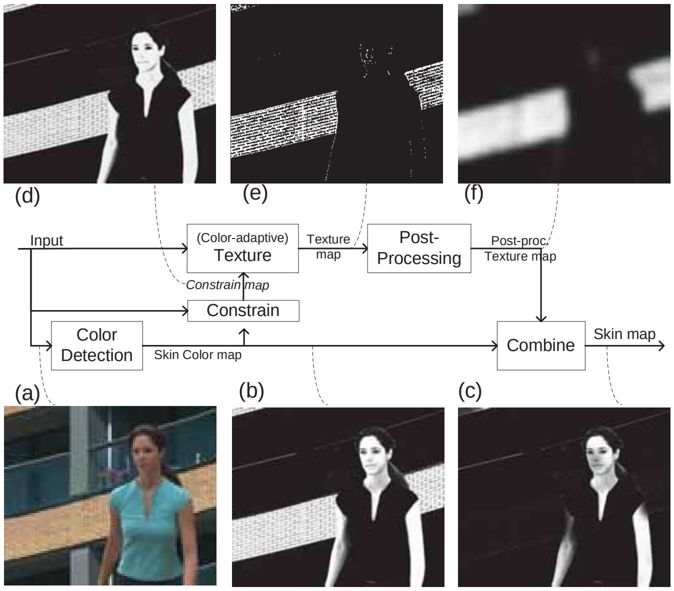


Fig. 1 Schematic overview of the texture-adaptive skin detection algorithm. (a) Input. (b) Color map. (c) Final texture-adaptive skin map. (d) Constrain map. (e) Texture map. (f) Post-processed texture map. When comparing (b) and (c): the proposed method (output shown in (c)) correctly rejects the skin-colored background wall, whereas the color-only method (output shown in (b)) wrongly accepts the wall.

[1] This work has been sponsored by the Casimir program of the Netherlands Organization for Scientific Research (NWO) under grant number 018.002.013, and by NXP Semiconductors, the Netherlands.

Bahman Zafarifar is with Sigma Designs, Eindhoven, the Netherlands and Eindhoven University of Technology. Erwin B. Bellers is with Sigma Designs. Prof. Peter H. N. de With is with Eindhoven University of Technology.

[2] "PNX5100 Video Back-End Processor"

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

II. TEXTURE-ADAPTIVE SKIN DETECTION ALGORITHM

Fig. 1 shows the schematic overview of the texture-adaptive skin detection algorithm [2]. In the figure, the *Color Detection* block first computes a skin-color map. The *Constrain* block generates a constrain map (Fig. 1(d)) that delineates input measuring areas for deriving the texture feature. The constrain map includes skin-colored areas, but excludes skin boundaries and dark facial features, which limits an excessive texture feature output at these areas. The pixel-level texture map (the output of the *Texture* block, Fig. 1(e)) is strongly low-pass filtered in the *Post-processing* block (result in Fig. 1(f)). The skin segmentation map is finally computed by combining the post-processed texture map (Fig. 1(f)) with the skin-color map (Fig. 1(b)), hereby removing skin-colored textured areas.

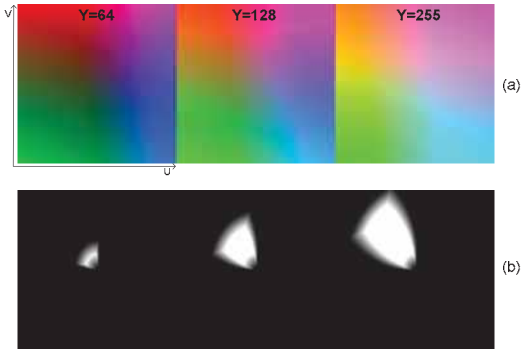


Fig. 2 Skin detection based on a parametric model defined in HSV space. (a) Input image containing 3 UV planes with constant Y values. (b) skin color detection output.

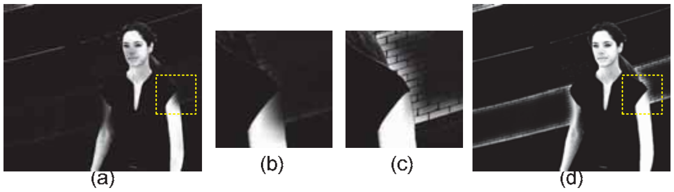


Fig. 3 Applying erosion on the texture map. (a),(b) Texture-adaptive skin detection without erosion. (c),(d) TASD when erosion is applied. Upper-right part of the arm is not removed in (c), at the cost of some false acceptance of the skin-colored brick wall boundary.

In [2], we have employed a generic and flexible color detector (classifier) that uses a 3D color histogram, representing the class-conditional probability of skin. Such classifiers have been shown to outperform classifiers based on parametric probability density estimations [3] [4]. We have concluded in [2] that the 3D color histogram method is expensive to implement in hardware or software. In order to limit the implementation cost, in this paper we use a simpler color detector defined by a parametric model in HSV color-space (Fig. 2), tuned to detect almost all skin areas in commonly occurring lighting conditions.

It is noted in [2] that using a texture feature may partially remove small skin areas or skin boundaries. To reduce this problem, in this paper we optionally perform an additional 3×3 erosion operation on the texture map, in the *Post-processing* step of Fig. 1. More specifically, the post-processing steps (see [2] for details) consists of a cascade of the following operations: downscaling by factor 16, spatial filtering by a 5×5 Gaussian kernel, the above-mentioned 3×3 gray-scale erosion, an 1st order IIR temporal filtering, and bi-linear upscaling by factor 16. Fig. 3 shows that applying erosion on the texture map helps to reduce the removal of skin boundaries, at the cost of allowing some false positives in skin-colored non-skin objects.

Fig. 4 shows that at 80% True Positive Rate (TPR), the False Positive Rate (FPR) is reduced from 0.153 for the color-only detector, to 0.133 for the texture adaptive detector (13% relative reduction), and to 0.108 when the additional erosion of the texture feature is applied (29% relative reduction).

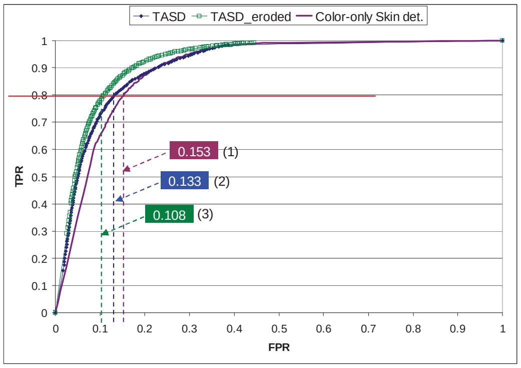


Fig. 4. Comparing performance (ROC) of (1) the color-only skin detector, (2) the Texture-Adaptive Skin Detector (TASD), and (3) TASD with additional erosion operation on the texture map.

III. IMAGE ENHANCEMENT FUNCTIONS

In this section we briefly describe the three skin-dependent enhancement functions used for our evaluations.

The *skin-tone correction* operation ([1]-Sec.5.4.3) detects skin-colored areas and modifies their color towards a desired target color, hereby restoring possible skin color deviations resulting from signal coding/transmission.

The *sharpness enhancement* block increases the image sharpness perception by combining a (a) 2-dimentional Linear Peaking operation ([1]-Sec.5.1.1) that enhances high-frequency details, (b) non-linear Luminance Transient Improvement ([1]-Sec.5.1.2) that increases the transition steepness of luminance edges, and (c) texture booster that improves the contrast of high-frequency textured areas. The contribution of operations (a) and (c) are attenuated in skin areas, in order to avoid an unnatural skin appearance.

The *local contrast enhancement* function improves the contrast of lower-mid-to-high frequency content (object size smaller than 10% of the image size) by combining the contributions of a differential mid/high frequency booster, and a luminance booster that amplifies the luminance in darker areas. The differential booster first computes a differential map by subtracting the image from its strongly low-pass filtered version. Next, a desired contrast gain is applied to the differential map, where the gain is reduced for skin areas to avoid producing an unnatural skin appearance. In the enhanced image, the chrominance is adjusted for areas that undergo luminance boosting, so as to maintain correct color perception.

IV. MEASURING ENHANCEMENT PERFORMANCE

The image enhancement functions of Section III require a skin map for their operation. Our purpose is to objectively measure the effect on image quality, when the skin map is obtained from the existing color-only skin detector, or by the proposed texture-adaptive detector. To this end, we first generate an *optimally enhanced* output I_{Opt} by executing the enhancement functions while using a manually annotated *ground-truth skin map* $P_{Opt,Skin}$ (delineating skin, non-skin and unknown areas). Now, the output I_{Enh} of each configuration is compared against the optimal enhancement I_{Opt}, to compute the Root Mean Square Error (RMSE), representing an objective measure of distance between I_{Enh} and I_{Opt} .

Assigning subscripts N and S to non-skin and skin objects, respectively, we denote N_N and N_S as the number of pixels of non-skin and skin objects, respectively. We further denote d_N and d_S as the sum of squared pixel value differences between an enhanced image I_{Enh} and the optimal enhancement I_{Opt}, for non-skin and skin objects respectively. The RMS errors E_N, E_S and the total error E are now computed as

$$E_N=\sqrt{d_N/N},\quad E_S=\sqrt{d_S/N},\quad E=\sqrt{(d_N+d_S)/N},\quad N=(N_N+N_S)\cdot(1)$$

Using the enhancement functions described in Section III, we set up the following enhancement configurations:
(1) Only skin-tone Correction (SKCR),
(2) Only sharpness enhancement (SHR),
(3) Only local contrast enhancement (LOCO),
(4) Enhancement cascade SKCR, SHR and LOCO.

We perform image enhancement on the test set in each of the above configurations, while using one of the following skin detection methods:

- Color-only skin detector,
- Texture-adaptive skin detector (TASD),
- TASD including erosion of texture map (TASD_eroded),
- Ground truth skin map.

The above procedure generates 12 enhanced versions of the test set (to be evaluated), and 4 optimal enhancement sets (to compare against). The RMS errors between the above 12 enhanced versions and the corresponding optimal enhancement sets are computed as described earlier in this section.

Using a test set of 174 manually annotated HD skin images, we obtain the results presented in Fig. 5. We observe the following for all 4 enhancement configurations.

- The error of the color-only method is the highest,
- The texture-adaptive method results in the lowest error,
- Including texture erosion in the texture-adaptive method slightly increases the error.

For example, in case of the enhancement cascade, the RMS error decreases from 6.47 to 3.6 for the texture-adaptive detector (44% relative decrease), and to 4.51 when also the additional texture erosion is applied (30% relative decrease).

Fig. 6 shows the relative RMS errors E_N, E_S of non-skin and skin areas. We can see that including the texture erosion operation leads to a higher error in the non-skin areas, but a lower error in the skin areas. Taking the enhancement cascade as an example (the left-most two bars), we can see that the error of the eroded TASD in skin areas (0.56) is smaller than that of TASD (0.87), but that the error in non-skin areas of the eroded TASD (4.47) is higher than that of TASD (3.49). This corresponds to the trade-off depicted in Fig. 3, which shows that the eroded TASD has a better detection of skin areas (lower error), versus more false positives in non-skin areas (higher error).

The results in Fig. 5 and Fig. 6 show that including the erosion operation in TASD slightly increases the total error. However, since this objective measure is not perceptually weighted, visual validation of the enhancement results is required when choosing an appropriate skin detection method for each enhancement application.

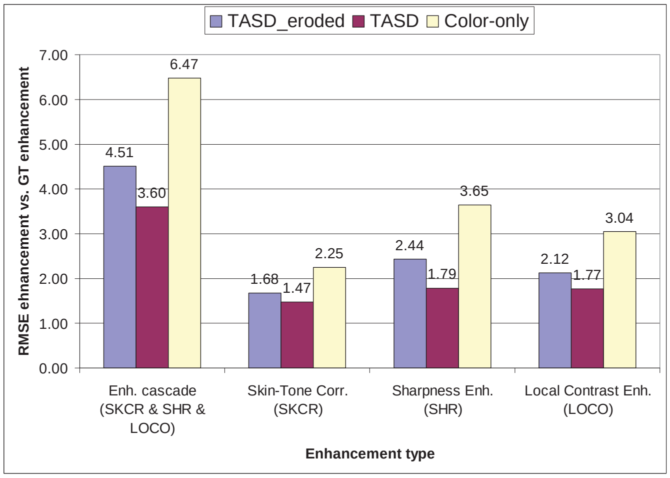


Fig. 5 Performance evaluation of image enhancement functions for different skin detection methods. Bars show error *E*, the RMSE of each enhancement vs. the optimal enhancement (lower values are better).

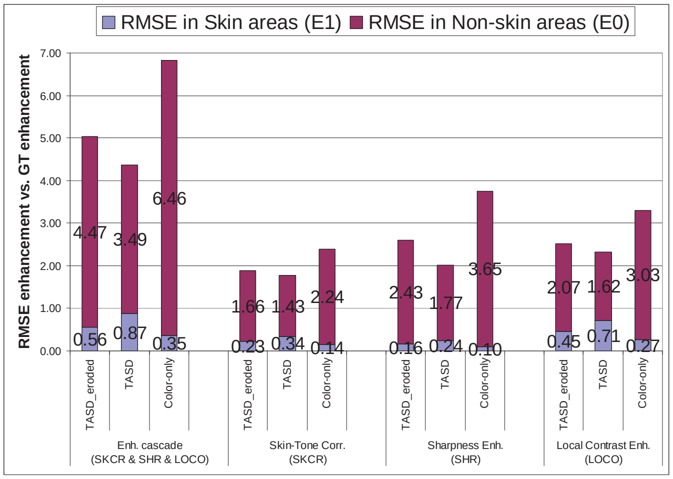


Fig. 6 Contribution of enhancement error E_N in non-skin areas and E_S in skin areas. Bars show RMSE of each enhancement vs. the optimal enhancement (lower values are better).

V. CONCLUSION

In this paper, we have evaluated a few TV image enhancement applications for a previously reported texture-adaptive skin detection algorithm. We have measured the detection performance of the texture-adaptive skin detector, and have evaluated the effect on image quality enhancement by using three TV image enhancement functions.

Our simulations show an overall improvement both in skin detection and in skin-dependent enhancement, when using the texture-adaptive skin detector. In terms of *detection score*, compared to the color-only detector, a relative reduction of 13% and 29% in false positive rate is achieved for two versions of the texture-adaptive skin detector, while detecting 80% of the skin pixels. In terms of *enhancement score* when using a cascade of 3 image enhancement functions, the RMS error of the color-only detector is reduced by 44% and 30%, for two versions of the texture-adaptive skin detector.

Subjective assessment of the enhancement results on large TV screens have indicated that the texture-adaptive skin detector is most suitable for enhancements that boost the mid/high-frequency contents, such as the evaluated sharpness enhancement and local contrast enhancement functions (see Fig. 7). In its current state, the proposed texture-adaptive skin detector is less suitable for enhancements that modify the DC level of the image, such as the evaluated skin-tone correction function. In this case, the shallow slopes of the low-pass filtered texture feature leads to a reduction of the skin-map value near skin-colored object boundaries, resulting to visible luminance or color deviations in the enhanced image. When implementation costs are not of primary concern, more elaborate filtering methods like image-guided bilateral filtering [5] may be used to better align the texture map with object boundaries.

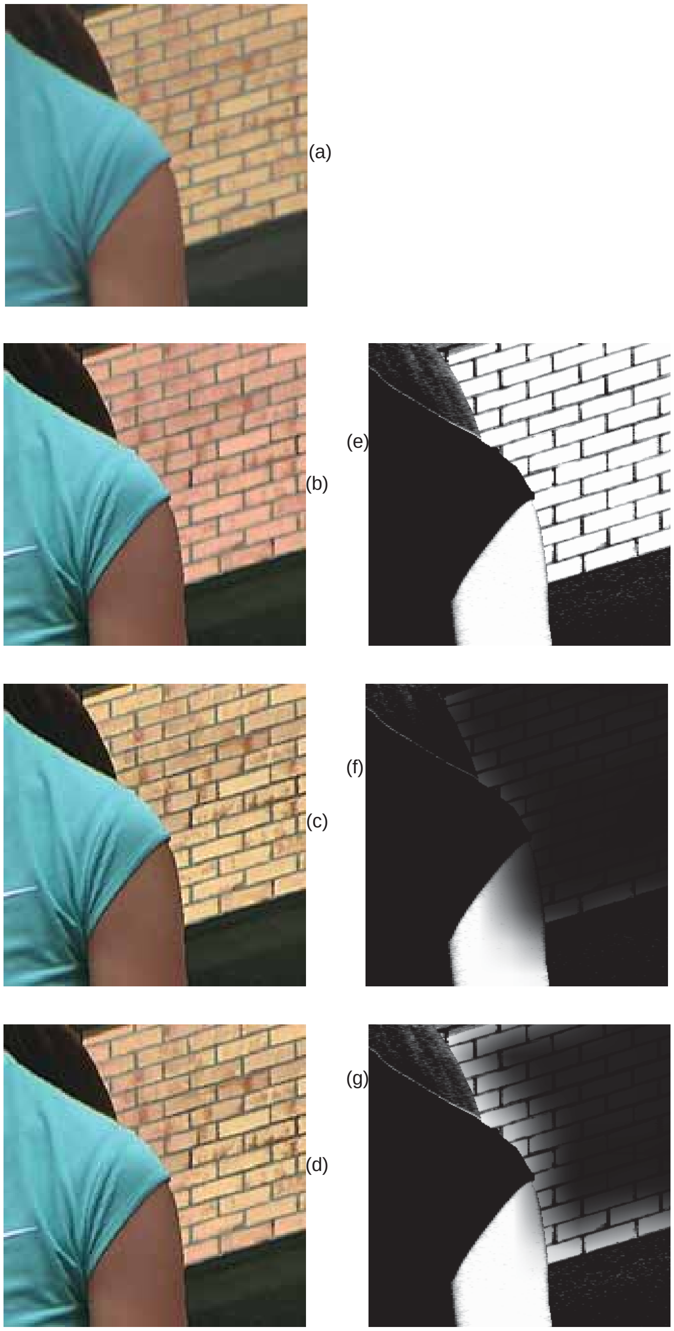


Fig. 7 Skin detection result of the 3 evaluated methods, and the corresponding image enhancement outputs (cascade of 3 enhancement functions). (a) Input. (e),(b) Detection/enhancement using the color-only detector, (f),(c) Detection/enhancement using the texture-adaptive skin detector (TASD). (g),(d) Detection/enhancement using the TASD when including erosion of the texture map. Image (b) has a reddish color deviation and low sharpness and contrast on the brick wall, caused by the false detection of background wall by the color-only detector. In image (c), the background wall has no color deviation, is sharp and has a high contrast (the wall is correctly rejected by the texture-adaptive skin detector). In (d), a slight color deviation occurs at the boundary of the background wall where approaching the person (caused by an incomplete rejection of the wall boundary due to the texture erosion operation).

REFERENCES

[1] "Digital Video Post Processing," Gerard de Haan, publication of Eindhoven University of Technology, July 2008, chapter 5, Section 5.4.3: Skin-tone correction, Section 5.1: Sharpness improvement (5.1.1: Linear peaking, 5.1.2 Non-linear edge enhancement).

[2] Bahman Zafarifar, Tim van den Kerkhof and Peter H. N. de With, "Texture-adaptive skin detection for TV and its real-time implementation on DSP and FPGA," *IEEE Transactions on in Consumer Electronics*, Volume: 58, Issue: 1, 2012, pp. 161-169.

[3] Phung S.L, Bouzerdoum A. and Chai D., "Skin Segmentation Using Color and Edge Information," *Proc. Int. Symposium on Signal Processing and its Applications*, July 2003, pp. 525- 528.

[4] Michael J. Jones and James M. Rehg, "Statistical Color Models with Application to Skin Detection," *International Journal of Computer Vision*, Vol. 46, Number 1 / Jan. 2002.

[5] O. P. Gangwal, E. Coezijn, and R.-P. Berretty, "Real-time implementation of depth map post-processing for 3D-TV on a programmable DSP (TriMedia)," Digest of IEEE International Conference on Consumer Electronics, Jan. 2009.

Enhanced Forwarding Engine for Content-Centric Networking (CCN)

Jaehoon Kim, Myeong-Wuk Jang, Joonghong Park, SungChan Choi, and Byoung-Joon (BJ) Lee
SAIT, Samsung Electronics, Korea

Abstract— **While majority of Internet traffic is fast becoming of high quality multimedia variety, efficient sharing of such multimedia traffic with intuitive usability and strong privacy protection remains an elusive goal. CCN is considered a promising networking technology which supports end-to-content communication paradigm instead of existing end-to-end communication. Although an open source implementation of CCN protocol is available, it is still in its early concept prototype stage. In this paper, we propose an enhanced CCN forwarding engine to support practical deployment of various CCN-enabled applications and services, thus demonstrating the full potential of CCN benefit. This paper also presents experimental results obtained by prototype implementation to demonstrate the effectiveness of proposed approach.**

I. INTRODUCTION

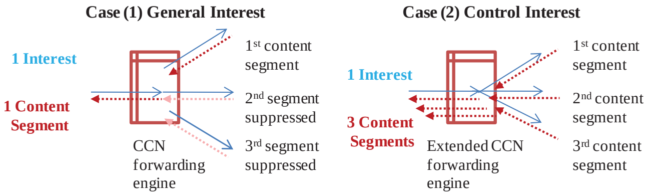
Today, average consumer just wants to access content, e.g., “what”, but do not want to bother to find out exact location of the desired content, e.g., “where”. However, the current Internet, to be more precisely the IP (Internet Protocol), operates in a host-centric way which requires users to know “where” the content is exactly located [1,2,4]. Such gap between content-centric requirements and underlying host-centric network plumbing results in various architectural and performance inefficiencies in scalability, security, flexibility, and mobility. For this reason, research on Content-Centric Networking (CCN) has gained considerable attention in recent years, and is regarded as a promising paradigm for the future Internet [1,3,4]. This network architecture is based on named and signed data networking and provides significant traffic reduction on content dissemination. CCN implementation is also available as an open source code, CCNx[5].

In CCN, a content requestor retrieves content by initiating *Interest* messages. Therefore, CCN operates on the pull-based mechanism. In response to the Interest, at most one matching content segment is delivered to the requestor to maintain the flow balance, suppressing any redundant copies. However, this one-to-one (1:1) Interest-Data segment communication mode may not be efficient when a requester needs to retrieve multiple content segments, also from multiple sources as in the case of content browsing, searching and content name enumeration. Therefore, one Interest-to-multiple Data response (1:*n*) mode of operation is necessary in CCN.

Also, in order to send a message to others without any previous requests, the push mechanism is required. In this paper, we present an enhanced CCN forwarding engine to support such extension mechanism in CCN. This paper also describes a usable and intuitive content sharing method in CCN based on the proposed forwarding engine extension, i.e., browsing-based content navigation, keyword-based content search, and push-based content distribution.

II. ENHANCED FORWARDING ENGINE

We propose a new type of control Interest to support an efficient content sharing services in CCN which are not addressed by the current CCNx implementation. The Control Interest is composed of existing name components and new name components with a specific marker starting with *‘*.C.I.*’*. The Control Interests is not compatible with the current CCNx forwarding engine, because a requester sending a Control Interest often expects to receive multiple content segments with the same name from multiple sources.



1:1 communication	1:n communication
Same content segments suppressed after the Interest satisfied by the first matching content segment	Reverse forwarding path maintained for multiple content segment retrieval
Multiple Interests in sequence to retrieve multiple content segments from multiple sources, e.g., Browsing, Name enumeration, Search, etc	Single Control Interest for multiple content segments from multiple sources
#of Interest : O(n) *n: # of responders	#of Interest : O(1)

Fig. 1 Interest/Data exchange mode of operation in CCN (1:1 vs 1:n)

To support simultaneous multi-source handling of the Control Interest on a CCN node, the CCN forwarding engine is extended to adaptively deliver one or multiple identical content segments according to the type of the Interest. Fig 1 illustrates the operation of such extension mechanism.

Case (1) of Fig. 1 shows the basic CCN forwarding engine operation with 1:1 Interest/Data communication mode, where the multiple copies of same content segments are being suppressed. In order to retrieve multiple content segments with the same name simultaneously from multiple sources, it is necessary to send identical Interests multiple times in sequence, causing additional processing and congestion overhead, and potentially up to *n* times more delay.

As shown in Case (2) of Fig. 1, the proposed mechanism only requires single Control Interest for simultaneous multiple content segments. In this case, the reverse path for the content segments is maintained for a certain duration. By using 1:*n* communication mode, the number of Interests generated from a content requestor does not increase, even when the network size grows in terms of the number of content sources. In the remainder of this section, we describe the use case services running on our prototype implementation, where the extended CCN forwarding engine provides performance benefit.

2.1 Content and Member Browsing

In order to locate desired content or range of relevant content names within a group of CCN-enabled devices such as VPC (Virtual Private Community) in [3], a tree-based browsing model is implemented with an 1:n Control Interest mechanism.

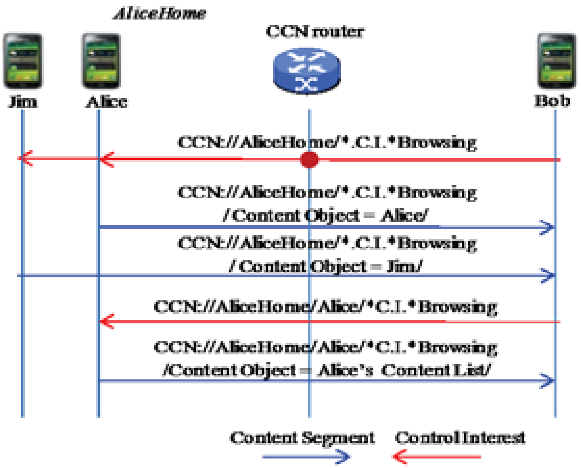

AliceHome
CCN router
Jim
Alice
Bob
CCN://AliceHome/*.C.I.*Browsing
CCN://AliceHome/*.C.I.*Browsing /Content Object = Alice/
CCN://AliceHome/*.C.I.*Browsing / Content Object = Jim/
CCN://AliceHome/Alice/*C.I.*Browsing
CCN://AliceHome/Alice/*C.I.*Browsing /Content Object = Alice's Content List/
Content Segment
Control Interest

Fig. 2 Content and member browsing use case using 1:n Control Interest

Fig. 2 illustrates a browsing use case, where Bob sends out a Control Interest with a predefined marker, i.e., **.C.I.*Browsing* in CCN:*//AliceHome/*.C.I.*Browsing*, both Alice and Jim in a group *AliceHome* reply with their names. In this case, the name CCN:*//AliceHome/*.C.I.*Browsing/ Alice* designates the content prefix held by Alice. After acquiring all the content prefix list available within the *AliceHome* group, Bob can iteratively ask for the content list available under such content prefix, e.g., CCN:*//AliceHome/Alice/*.C.I.*Browsing*. In this way, Bob acquires a list of content that Alice is willing to share, and then can select desired content for downloads by generating a general Interest.

2.2 Content Search

A user can also retrieve a list of full content names by keyword, or a substring of content name. If a content requestor expresses a Control Interest including a *keyword*, e.g., */prefix*/**.C.I.*Search/keyword*, all responses from the content owners with original or copies of content coinciding with *keyword* can simultaneously be delivered to the requestor, using 1:*n* communication mode. In this search mode, a community or group name prefix, content type, and keyword are used as basic name components of the Control Interest. When a group prefix such as *AliceHome* is inserted before the Control Interest marker '**.C.I.*Search*', the Interest, e.g., *AliceHome/***.C.I.*Search**/keyword*, is only disseminated within such closed user group.

2.3 Push-based Content Sharing

The original CCN concept is based on the pull mechanism, which is very useful when a user accesses content. The pull mechanism, however, is not inherently compatible with the use cases where a user wants to announce information about its own content for distribution. Therefore, a push mechanism with a 4-way handshaking is implemented using our proposed 1:*n* Control Interest mechanism.

When a content owner wants to push content to a user or group of users who might be interested in it, a Push Interest, e.g., CCN:*//AliceHome/*.C.I.*Push/Bobhome/Bob/sea.photo*, including a target prefix and a full name of the content can be sent as a form of notification. The target prefix *AliceHome* for routing is located before the Control Interest marker, '**.C.I.*Push*'. In response to the Control Interest for push, the target users willing to receive the announced content sends the content name in the general Interest.

III. Implementation and Experiment

To demonstrate the feasibility and effectiveness of our proposed 1:*n* Control Interest mechanism, we implemented prototype on Android mobile phones and Linux PCs. Table 1 shows experimental results which compare the content browsing delay with various device types. In all cases, a Linux PC is used as a CCN routing hub. The delay measurement takes the average value out of 10 test runs each. As shown in Table 1, the results confirmed the effectiveness of our proposed Control Interest mechanism in fast content browsing.

Table. 1 Browsing Delay: Original CCN vs. Enhanced CCN

Prototype Testbed Configuration			**Forwarding Engine**	
Content Requester Device type	CCN Routing Hub	# Content Sources (Android Galaxy S)	Original CCN	Enhanced CCN
Android Galaxy S	Linux PC	4	1196ms	228ms
Linux PC		3	122ms	85ms
		4	169ms	100ms

IV. CONCLUSION

The proposed 1:n Control Interest mechanism for enhanced CCN forwarding engine is designed to reduce the amount of overall network traffic and delay. The performance benefit will particularly be well highlighted in the practical CCN-enabled applications and services such as content browsing, keyword-based content search and push-based content distribution among others.

Reference

[1] Van Jacobson, Diana. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking Named Content," In Proceedings of CoNEXT'09, Rome, Italy, Dec. 2009.

[2] A. Detti and N. Blefari-Melazzi, "Network-layer solutions for a content-centric Internet," 21st International Tyrrhenian Workshop on Digital Communications (ITWDC), Ponza, Italy, Sep. 2010.

[3] Jaehoon Kim, Myeong-Wuk Jang, Byoung-Joon (BJ) Lee, and Kiho Kim, "Content Centric Network-based Virtual Private Community," In Proceedings of ICCE12, Las Vegas, US, Jan.2011

[4] Lixia Zhang, et al, "Named Data Networking (NDN) Project," PARC Technical Report NDN-0001, Oct. 2010.

[5] Project CCNx. http://www.ccnx.org, Sep. 2009.

Prioritized Dual Caching Algorithm for Peer-to-Peer Content Network

Jong-Geun Park[†], and Hoon Choi[‡], *Senior Member, IEEE*
[†] Electronics & Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea
[‡] Chungnam National University, Daejeon, Republic of Korea

***Abstract*—Content caching is a fundamental strategy for improving the performance and quality of service perceived by users by storing popular objects that are likely to be used in the near future. P2P traffic has its own characteristics, such as flattened head and long heavy tailed popularity distribution. Therefore, a new cache management algorithm which incorporates well P2P traffic characteristics needs to be considered.**

In this paper, we propose a prioritized dual caching algorithm which addresses P2P traffic and present results from simulation experiments. Our results show that our algorithm achieves relatively high performance improvement. In particular, the proposed algorithm is very effective when the cache size is relatively small.

I. Introduction

Peer-to-peer (P2P) file sharing has been one of the most dominant Internet services in the past few years. It generates a major portion of the Internet traffic, and this portion is expected to increase continuously in the future. However, most of cache management algorithms have been focused on Web caching.

It is a well known feature that the popularity distribution of references for Web proxies follows a Zipf distribution, where the relative probability of a request for the i-th most popular object is proportional to $1/i^{\alpha}$, where α is the Zipf exponent [1], [2]. But, this property is no more valid to P2P traffic. Recently, many measurement studies for P2P file sharing [2], [3], [4] have shown that the curve of the popularity distribution of P2P objects has a flattened head at the lowest ranks, whereas Zipf distribution should show obviously linear when plotted on a log-log scale. Therefore, modeling P2P popularity as a Zipf-like distribution yields a significant error. The user's fetch-at-most-once behavior, distinct from the fetch-repeatedly behavior of Web users, is the cause of the flattened head property [3]. The flattened head nature of P2P objects can be modeled as a generalized form of the Zipf distribution, called Mandelbrot-Zipf distribution [4].

Another popularity characteristic of P2P traffic is that there may be numerous singly or rarely accessed objects in the reference streams. Generally, a Zipf distribution and a Mandelbrot-Zipf distribution are a heavy tailed distribution. This characteristic means that rarely accessed objects account for a large fraction of the total objects in Zipf or Mandelbrot-Zipf reference streams. To account for this, a mechanism that de-emphasizes the accesses made to unpopular objects such as recency-based cache placement policy (RBP) [5] needs to be adopted.

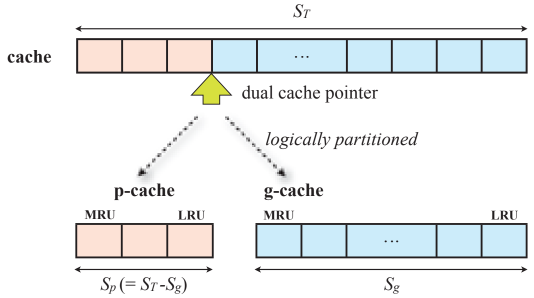


Fig. 1. General structure of the prioritized dual caching policy. The original cache with size S_T is partitioned into the p-cache of size S_p and the g-cache of size S_g.

The RBP policy addresses heavy tailed characteristics for unpopular objects, but the flattened head feature was not considered to design a cache management policy. In this paper, we propose a novel P2P caching scheme based on popularity in order to reflect the flattened head characteristic of P2P traffic.

II. Prioritized Dual Caching Algorithm

Let us consider a P2P file sharing service which has N subscribed users. For cache efficiency, a content server has a content cache with size of S_T, and the content caching strategy has two threshold values, the minimum number of references in order to ignore caching, T_H, and the minimum number of references for hot popular content, $T_L (0 < T_L < T_H < N)$. We assume that the size of each content is uniform and the number of content which the cache can contains is also at most S_T.

The cache can be divided into a premium cache, the p-cache, which contains highly popular contents that are at least T_L times requested; and a general cache, the g-cache, in which a general content that is not highly popular can be cached. The size of each cache is S_p and S_g, respectively. These are not constant, but adaptively decided depending on dynamic reference stream to prioritize content. This is the prioritized dual caching (PDC) algorithm.

Simplifying slightly, on the reference to a content i, if the total number of references to a content i, n_i, does not exceed a threshold T_L, then the PDC places it in the most recently used (MRU) position of the g-cache. On the other hand, if n_i exceeds T_L, the content i should be cached or moved from the

This work was partially supported by the IT R&D program of Ministry of Knowledge Economy (MKE) and Korea Evaluation Institute of Industrial Technology, Republic of Korea, 10039156.

This work was also supported by the MKE, Republic of Korea, under the ITRC support program supervised by the NIPA (NIPA-2012-H0301-12-3005).

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

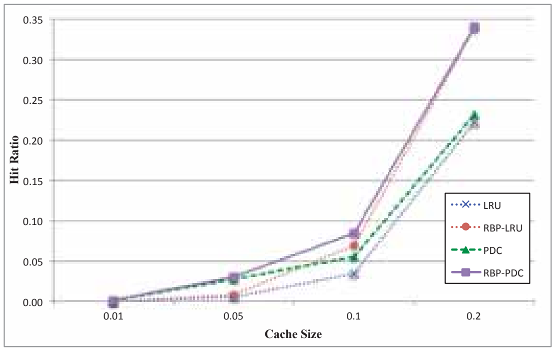


Fig. 2. Performance comparison among LRU, RBP-LRU, PDC, and RBP-PDC

g-cache into the p-cache. During the content is in the p-cache, it cannot be evicted from the p-cache even though it is the least recently used (LRU) one. But, when there is no more low priority content which means $S_p = S_T$ and $S_g = 0$, the least recently used hot content in the p-cache is evicted from the cache. Furthermore, when the n_i exceeds the maximum threshold, T_H, the content i cannot continue to be cached anymore. And it should be evicted from the cache even if it is the most recently used content, since its popularity can be considered as almost finished and the effect of caching may be not critical. The reason why we introduce the threshold T_H is to reflect the flattened head characteristics of P2P popularity distribution. General structure of the prioritized dual caching policy is depicted in Fig. 1.

III. Simulation Experiments

In this section, we discuss the results from simulation experiments to evaluate our proposed PDC algorithm for P2P content caching.

First of all, in order to generate P2P request stream which follows the Mandelbrot-Zipf distribution, we simulated the request distribution generated by 250 user population. For each user, we simulated the same initial Zipf distribution with parameter $\alpha = 1.0$ over 10,000 distinct objects, and inter arrival time between consecutive references follows the Poisson process with the arrival rate 0.002 for 1,000,000. For the most important property, the fetch-at-most-once, we prevented each user from making subsequent requests for the same object. This yields the flattened head popularity distribution.

The cache size used in the simulation experiments were chosen by taking a fixed percentage of the total unique number of objects. The percentages are 1%, 5%, 10%, and 20% which are logically relative to the infinite cache size.

Fig. 2 shows the results of performance comparison among LRU, RBP-LRU (that is, LRU replacement with recency based cache placement), PDC, and RBP-PDC (that is, PDC cache with recency based cache placement) as a function of the relative cache size. When the cache size is relatively small, the hit ratio of PDC and RBP-PDC is relatively very high over LRU and RBP-LRU. The relative hit ratio improvement of PDC and RBP-PDC over LRU is more than 400% and in some case which it exceeds 950%. This performance improvement is due to the fact that PDC policy does tend to contain more popular objects in the cache. The effect of this tendency is maximized, when the cache size is small. For larger cache sizes, the hit ratio of RBP algorithm is relatively high over no placement algorithms. This improvement is achieved by selective admission for caching, which results in preventing heavy tailed unpopular object from evicting popular objects.

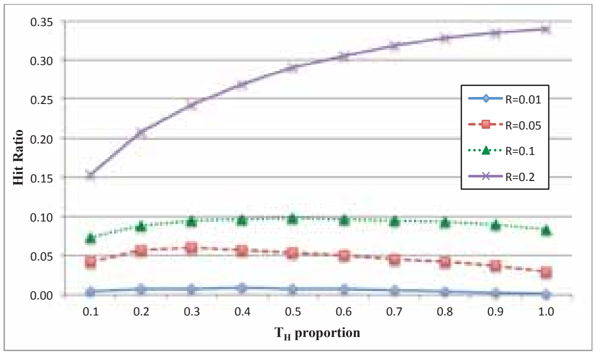


Fig. 3. Performance evaluation of RBP-PDC while changing the T_H value as a proportion to the total number of users and R value as a relative cache size (T_L was set as 80% of T_H value for all experiments)

Fig. 3 shows the hit ratios when the relative cache size, R, is 1%, 5%, 10%, and 20%, respectively, while the maximum number of references for discarding caching, T_H, is varying from 0.1 to 1.0 of the total number of users. The performance of PDC algorithm depends on the value of T_H. When the cache size is relatively large, large T_H value guarantees high hit ratio because the cache is enough to continue to contain highly referenced objects. But, when the cache is relatively small, T_H value needs to be tuned for the optimal performance.

IV. Conclusion

In this paper, we proposed a prioritized dual caching algorithm that incorporates the flattened head characteristic of P2P popularity distribution. According to our simulation experiments, the proposed caching algorithm achieves relatively high hit ratio when the cache is small. Furthermore, if it adopts a selective cache placement policy to address heavy tailed unpopular objects, it is superior to other replacement algorithms.

References

[1] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: evidence and implications," in *Proc. IEEE INFOCOM '99*, New York, NY, Mar. 1999, pp. 126–134.

[2] Z. Liu and C. Chen, "Modeling fetch-at-most-once behavior in peer-to-peer file-sharing systems," in *Proc. APWeb Workshop 2006*, vol. LNCS 3842, Harbin, China, Jan. 2006, pp. 717–724.

[3] K. P. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan, "Measurement, modeling, and analysis of a peer-to-peer file-sharing workload," in *Proc. SOSP '03*, Bolton Landing, NY, Oct. 2003, pp. 314–329.

[4] M. Hefeeda and O. Saleh, "Traffic modeling and proportional partial caching for peer-to-peer systems," *IEEE/ACM Trans. Netw.*, vol. 16, pp. 1447–1460, Dec. 2008.

[5] J.-G. Park and H. Choi, "On selective placement for uniform cache objects," in *Proc. IEEE ICAIT'12*, Paris, France, Jul. 2012.

WiCUBIC: Enhanced CUBIC TCP for Mobile Devices

Yongsu Gwak and Young Yong Kim
Yonsei University
Seodaemun-gu, Seoul, 120-749, Korea

Ronny Yongho Kim
Korea National University of Transportation
Uiwang, Gyeongki, 437-763 Korea

***Abstract*— CUBIC is widely used as a default TCP variant in android based smart phones and recent Linux kernels. However, since CUBIC is designed to improve the scalability of TCP over fast and long distance networks in wired environments, its throughput performance needs to be improved under erroneous wireless environments. This paper proposes a novel TCP variant, Wireless CUBIC (WiCUBIC), which is an enhanced CUBIC TCP for mobile devices. By considering the different characteristics of packet losses caused by network congestion and wireless channel impairment, WiCUBIC is designed to distinguish losses caused by wireless channel from losses caused by network congestion. With such capability, WiCUBIC shows substantially large throughput performance improvement while providing fairness to other TCP protocols. Simulation results corroborate the efficiency and fairness of WiCUBIC.**

I. Introduction

CUBIC [1] is a default TCP variant in Android based smart phones and Linux kernels since version 2.6.19. Therefore, most of Andriod or Linux based mobile devices transfers their data complying with CUBIC's congestion control algorithm. However, since CUBIC is originally designed to improve the scalability of TCP over fast and long distance in wired environments, it shows degraded throughput performance in wireless environments due to its inability to differentiate wireless packet losses from congestion packet losses which is the principal problem of TCP variants inherited from wired TCP [2]. In contrast to the standard TCP's and many other TCP variants' additive increase of congestion window (*cwnd*) after *cwnd* reduction due to congestion, in order to provide the scalability, CUBIC multiplicatively (cubic function) increases *cwnd* after *cwnd* reduction due to congestion.

Algorithm 1
WiCUBIC Procedure

```
Packet loss:
begin
    $t_{last} \leftarrow t_{current}$
    $t_{current} \leftarrow tcp\_time\_stamp$
    $W_{\max}(current) \leftarrow cwnd$
    $eW_{\max}(m) \leftarrow \frac{\sum_{n=current-m+1}^{current} W_{\max}(n)}{m}$
    if $W_{\max}(current) < eW_{\max}(m)$ and $t_{current} - t_{last} > T$ then
        wireless channel impairment():
    else network congestion():
end
network congestion():
begin
    $cwnd \leftarrow (1-\beta)W_{\max}(current)$
    $K \leftarrow \sqrt[3]{\frac{W_{\max}(current)-cwnd}{C}}$
    $origin\_po\operatorname{int} \leftarrow cwnd$
    $t\arg et \leftarrow origin\_po\operatorname{int} + C(t-K)^3$
end
wireless channel impairment():
begin
    $cwnd \leftarrow W_{\max}(current)$
    $origin\_po\operatorname{int} \leftarrow cwnd$
    $t\arg et \leftarrow origin\_po\operatorname{int} + Ct^3$
end
```

Several cross layer approaches ([3]-[5]) have been proposed to improve the inefficiency of TCP variants in wireless environments. Despite of their performance improvement in wireless environments, they are not widely used in practice since they requires additional control signal to provide ability to distinguish losses caused by wireless link from network congestion. There have been TCP layer approaches including [6] providing ability to distinguish cause of packet losses. Even though they don't require additional control signal, since such approaches typically employ a certain cycle e.g., probe cycle in [6], to make decision on the cause of packet losses, they causes significant inefficiency.

In this paper, we propose a novel TCP variant, Wireless CUBIC (WiCUBIC) in order to improve the throughput performance of CUBIC in wireless environments. WiCUBIC has two important requirements:

1) ***Easy implementation***: TCP layer solution without additional control
2) ***Agile CWND Control***: fast differentiation of cause of packet losses without an additional cycle

In order to meet the requirements, WiCUBIC's *cwnd* control algorithm is designed based on careful observation on network capacity and patterns of packet losses. With salient features of WiCUBIC, WiCUBIC is able to provide significant performance improvement in wireless environments while providing fairness to the conventional TCP protocols.

II. Design of WiCUBIC

The proposed TCP variant, WiCUBIC is based on CUBIC. Whenever packet loss is detected, CUBIC regards the packet loss as network capacity is full (i.e., network congestion) and stores *cwnd* as $W_{\max}$. Therefore, change of $W_{\max}$ is closely related to network capacity. To differentiate wireless packet losses from congestion packet losses, WiCUBIC calculates a moving average value ($eW_{\max}$) of $W_{\max}$:

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

$$eW_{\max}(m) = \frac{\sum_{n=current-m+1}^{current} W_{\max}(n)}{m} \quad (2)$$

where m is a moving average factor and, with appropriate m, $eW_{\max}$ is able to effectively reflect change of network capacity.

WiCUBIC also stores last window reduction time (t_{last}) to utilize packet loss patterns for differentiation of packet loss cause. Since a pattern of congestion packet losses is more bursty than that of wireless packet losses, time difference between current and last window reduction ($t_{current} - t_{last}$) is another measurement to distinguish a cause of packet losses. The detailed procedure of WiCUBIC is described in Algorithm 1.

III. Simulation Results

An application server transfers TCP packets of 1460 bytes to mobile devices. The CUBIC parameter (C) is set to 0.4 and the multiplicative factor (β) is set to 0.2. The moving average factor (m) and the burstiness factor (T) of WiCUBIC are set to 3 and 0.5 respectively. In wireless environments, throughput performance of WiCUBIC is demonstrated and the fairness issue in wired environments is also investigated.

A. *Throughput in wireless environments*

We assume 3G network which has the bottleneck capacity of 1.8Mbps per each TCP session. With a two-state Markov chain packet error model [7], each packet is lost in wireless link. Fig. 1 shows the throughput performance comparison between CUBIC and WiCUBIC in wireless environments for various packet error rates. The throughput performance of WiCUBIC outperforms CUBIC for all error rates. Especially at 10% PER, the throughput gap between two protocols is largest and it is influenced by the value of T. If T is smaller than 0.5, the throughput of WiCUBIC at 20% PER increases, but it may break fairness in wired environments. Therefore, finding the optimal T is very important to improve the performance of WiCUBIC which is left as our future work.

B. *Fairness in wired environments*

In Fig. 2, CUBIC and WiCUBIC flows share a wired link of 1.8Mbps. WiCUBIC flow starts about 5 seconds later than CUBIC. After approximately 30 seconds, *cwnd*s of two flows converge. With appropriate m and T, WiCUBIC is able to provide fairness to CUBIC in wired environments while showing better throughput performance in wireless environments.

IV. Discussion and Future Works

In this paper, we propose WiCUBIC which is an enhanced CUBIC TCP for mobile devices. WiCUBIC is able to distinguish losses induced by wireless channel from losses induced by network congestion. Simulation results corroborate WiCUBIC's outstanding throughput performance and fairness to other TCPs. Finding the optimal T for further WiCUBIC performance improvement is left as our future work..

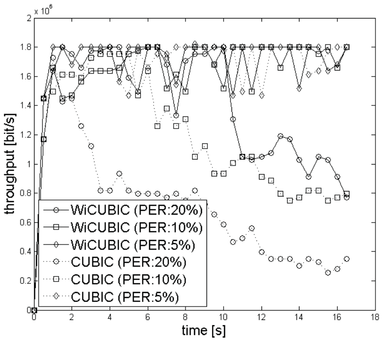


Fig. 1. The comparison of throughput performance between CUBIC and WiCUBIC in wireless environments

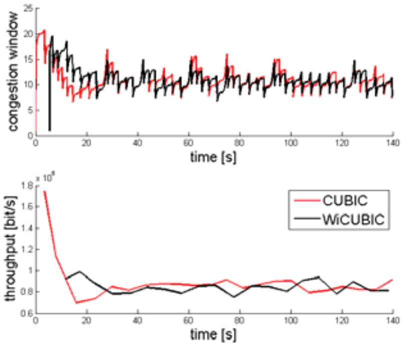


Fig. 2. CUBIC and WiCUBIC flows in wired environments

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2012R1A1A1014610). R. Y. Kim is the corresponding author.

References

[1] I. Rhee and L. Xu, "CUBIC: A New TCP-Friendly High-Speed TCP Variant," *in proc. of PFLDnet 2005*, Lyon, France, February 2005.

[2] V. Tsaoussidis and I. Matta, "Open Issues on TCP for Mobile Computing," *Journal of Wireless Communications and Mobile Computing*, vol. 2, no. 1, pp. 3-20, Feb. 2002.

[3] H. Balakrishnan, V. Padmanabhan, S. Seshan, and R. Katz, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links," *ACM/IEEE Transactions on Networking*, December 1997.

[4] Z. Haas and P. Agrawal, "Mobile-TCP: An Asymmetric Transport Protocol Design for Mobile Systems," *In Proceedings of the IEEE International Conference on Communications (ICC'97)*, 1997.

[5] K. Ramakrishnan and S. Floyd, "A Proposal to Add Explicit Congestion Notification (ECN) to IP," *RFC 2481*, January 1999.

[6] V. Tsaoussidis and H. Badr, "TCP-Probing: Towards an Error Control Schema with Energy and Throughput Performance Gains," *In Proceedings of the 8th IEEE International Conference on Network Protocols*, 2000.

[7] A. Chockalingam, Michele Zorzij and Velio Trallij, "Wireless TCP Performance with Link Layer FEC/AKQ*," *in proc. of IEEE International Conference on Communications. ICC '99*, Vancouver, Canada, 6-10 June 1999.