

On the estimation of target spectrum for filter-array based spectrometers

Cheng-Chun Chang* and Heung-No Lee

Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh PA,
USA

*Corresponding author: chc55@pitt.edu

Abstract: Miniature spectrometers have been drawn researchers much attention due to its wide variety of possible applications. In this paper, we show the achievability of a fine spectrometer on-a-chip based on a low-performance, low-cost filter-array. A low quality filter-array is augmented with digital signal processing techniques. A series of estimators for recovering target spectrum is introduced. By exploiting non-negative nature of spectral content, a non-negative least-square algorithm is found particularly useful for spectrum recovery. The concept is verified in a hardware implementation.

©2008 Optical Society of America

OCIS codes: (300.6190) Spectrometers; (120.6200) Spectrometers and spectroscopic instrumentation digital signal processing.

References and links

1. C. P. Bacon, Y. Mattley, and R. Defrece, "Miniature spectroscopic instrumentation: applications to biology and chemistry," *Rev. Sci. Instrum.* **75**, 1-16 (2004).
2. D. C. Heinz, and C.-I Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **39**, 529-546 (2001).
3. O. Manzardo, H. P. Herzig, C. R. Marxer, and N. F. de Rooij, "Miniaturized time-scanning Fourier transform spectrometer based on silicon technology," *Opt. Lett.* **24**, 1705-1707 (1999).
4. K. Chaganti, I. Salakhutdinov, I. Avrutsky, G. W. Auner, "A simple miniature optical spectrometer with a planar waveguide grating coupler in combination with a plano-convex lens," *Opt. Express* **14**, 4064-4072 (2006).
5. R. F. Wolffenbuttel, "State-of-the-art in integrated optical microspectrometers," *IEEE Trans. Instrum. Meas.* **53**, 197-202 (2004).
6. R. Shogenji, Y. Kitamura, K. Yamada, S. Miyatake, and J. Tanida, "Multispectral imaging using compact compound optics," *Opt. Express* **12**, 1643-1655 (2004).
7. S.-W. Wang, C. Xia, X. Cheng, W. Lu, L. Wang, Y. Wu, and Z. Wang, "Integrated optical filter arrays fabricated by using the combinatorial etching technique," *Opt. Lett.* **31**, 332-334 (2006).
8. S.-W. Wang, C. Xia, X. Cheng, W. Lu, M. Li, H. Wang, W. Zheng, and T. Zhang, "Concept of a high-resolution miniature spectrometer using an integrated filter array," *Opt. Lett.* **32**, 632-634 (2007).
9. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, 1974.
10. J. G. Proakis, *Digital Communications*, McGraw Hill, 2000.

1. Introduction

Recently, miniature spectrometers have been drawn researchers great attention. Miniature spectrometers provide solutions to a variety of promising applications in biological, chemical, medical, or pharmaceutical industries, in which small, light-weight, and non-fragile properties of spectrometers are demanded [1, 4-6]. Currently, MEMS, CMOS, micro-optic electromechanical systems, or integrated optics technologies are the main means to build a miniature- or micro- spectrometer. Based on the underlying operation principle, spectrometers can be classified such as grating-based, Fourier-transform based, or filter-based [3-5].

Filter-based static spectrometers utilize different filter functions to filter spectral energy emanating from a target. Recent literatures [6-8] have demonstrated that filter-based spectrometers are capable of high-resolution and allowed to fabricate on-a-chip. By implementing multiple filters and detectors, these spectrometers are avoided to have moving elements, and hence are static and rigid. Furthermore, they have the ability to capture the target spectrum in a very short time. This property is demanded for certain applications especially in biological, biochemical or biomedical industries.

For low-cost fabrication, filters may not have delta-function-like shapes with a narrow range response. The spectrum obtained directly from these filtered results is severely distorted, and hence is unacceptable as it is. However, signal processing techniques can be applied to estimate and restore the target spectrum.

In this paper, we discuss and provide digital-signal-processing (DSP) methods for filter-array based spectrometers. Based on a discrete linear system model, a series of estimators for spectrum recovery is introduced. By exploiting the non-negative nature of spectral content, we found the non-negative least-squares (NNLS) algorithm particularly useful to estimate and restore the target spectrum with high fidelity. A hardware implementation for the filter-array based spectrometer is demonstrated via a commercialized CCD camera and a DSP board. This prototype shows the achievability of a fine spectrometer on-a-chip based on a low-performance, low-cost filter-array.

Section 2 shows the system model of the filter-array based spectrometer. Section 3 discusses and provides the estimators for restoring a target spectrum. A hardware implementation and experimental results are shown in section 4. Section 5 draws the summary and conclusion.

2. System model of the filter-array based spectrometer

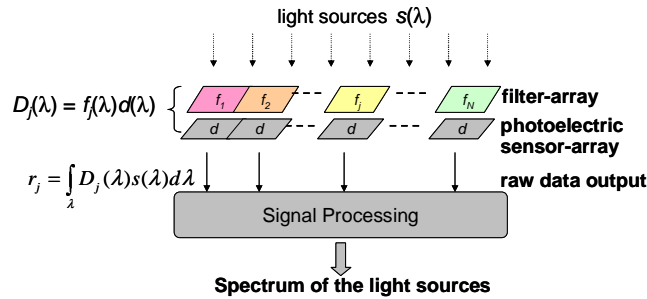


Fig. 1. System structure of the filter-array based spectrometers

Figure 1 shows the basic system model of the static filter-array based spectrometer. An array of filters is directly placed on top of an array of photoelectric sensors such as CCD sensors. A filter may correspond to a CCD sensor or a group of CCD sensors. The outputs from the CCD sensors are then fed into a digital signal processor. We note that the 1-D structure shown in Fig. 1 can be extended to a 2-D structure straightforwardly by placing filters and CCD sensors in a 2-D plane. In this paper, we will restrict our discussion on the drawn 1-D structure.

We assume all the CCD sensors have the same sensitivity function, denoted by $d(\lambda)$, where λ is the continuous wavelength. Denote $f_j(\lambda)$ as the filter transmission function for j^{th} filter with respect to wavelength λ (each filter may have a different transmission function). A filter and its corresponding CCD sensors compose a spectral detector. The overall sensitivity function of the j^{th} spectral detector can be expressed as $D_j(\lambda) = f_j(\lambda)d(\lambda)$. For a target with spectral content $s(\lambda)$, the output from the j^{th} detector is $r_j = \int_{\lambda} D_j(\lambda)s(\lambda)d\lambda$.

We consider the system and the target spectrum based on a *discrete* model. The transformation between the target spectrum and the CCD-sensor outputs is associated by the matrix equation

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (1)$$

where

$$\mathbf{r} = \begin{bmatrix} r_1 \\ \vdots \\ r_j \\ \vdots \\ r_N \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} D_1(\lambda_1) & \cdots & D_1(\lambda_M) \\ \vdots & \vdots & \vdots \\ D_j(\lambda_1) & \cdots & D_j(\lambda_M) \\ \vdots & \vdots & \vdots \\ D_N(\lambda_1) & \cdots & D_N(\lambda_M) \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} s(\lambda_1) \\ s(\lambda_2) \\ \vdots \\ s(\lambda_M) \end{bmatrix}, \quad \text{and } \mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{bmatrix}.$$

The dimensionalities of \mathbf{r} , \mathbf{H} , \mathbf{s} , and \mathbf{n} are $N \times 1$, $N \times M$, $M \times 1$, and $N \times 1$, respectively. \mathbf{r} is an observed signal vector whose elements are the outputs of CCD-sensors. The elements in \mathbf{r} can be observed simultaneously via each individual detector. \mathbf{n} is a noise vector. There are N detectors in this model. \mathbf{H} is a detector sensitivity matrix. The M elements in a row of \mathbf{H} matrix represent the sensitivity function of a detector, obtained by evenly sampling the sensitivity function over a certain wavelength range. \mathbf{s} is a source signal vector, whose elements represent the target spectrum evenly sampled in the wavelength domain. The minimum number of samples required for a given target-spectrum can be obtained through the *sampling theorem*. That is, consider the shape of the target-spectrum as a continuous function of a unit interval. If B is the minimum value such that the Fourier transform of the function over the unit interval $S(\theta)$ satisfying $S(\theta) = 0$ for $|\theta| > B$, $2B$ is the minimum number of samples to specify the continuous function [10, p70]. We call these samples *resolved points* in this paper. The number of resolved points for a given application is critical since it determines the number of required detectors as we will see in section 3. In addition, we note that the sensitivity characteristic of the detectors needs not to be a delta-function shape with a narrow range response.

3. Estimators for restoring the target spectrum

Working on the observation vector \mathbf{r} , an estimator provides an estimation $\hat{\mathbf{s}}$ of the input spectrum by considering all possible source signal vectors \mathbf{s} . One criterion we can use here as the starting point is the *maximum a posteriori* (MAP) rule [10, p242]. The MAP estimator is obtained by maximizing the *posterior* probability, i.e.,

$$\hat{\mathbf{s}}_{MAP} = \arg \max_{\mathbf{s}} P(\mathbf{s} | \mathbf{r}). \quad (2)$$

From the Bayes' rule, the *posterior* probability can be written as $P(\mathbf{s} | \mathbf{r}) = P(\mathbf{r} | \mathbf{s})P(\mathbf{s})/P(\mathbf{r})$.

When we do not have any information on the source signal such that $P(\mathbf{s})$ is uniformly distributed, the MAP estimator becomes the *maximum likelihood* (ML) estimator. The ML estimator maximizes the likelihood function, i.e.,

$$\hat{\mathbf{s}}_{ML} = \arg \max_{\mathbf{s}} P(\mathbf{r} | \mathbf{s}). \quad (3)$$

For the filter-array spectrometer, the observed signal vector \mathbf{r} and the source signal vector \mathbf{s} can be associated by Eq. (1) as discussed. Now assume the noise vector \mathbf{n} is multivariate Gaussian with zero mean and covariance matrix \mathbf{R}_n , i.e., $E[\mathbf{n}] = \mathbf{0}$, and $E[\mathbf{nn}^T] = \mathbf{R}_n$, where the superscript T denotes the transpose operation. The ML estimator then is obtained by maximizing the likelihood function

$$P(\mathbf{r} | \mathbf{s}) = \frac{1}{(2\pi)^{N/2} |\mathbf{R}_n|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{r} - \mathbf{H}\mathbf{s})^T \mathbf{R}_n^{-1} (\mathbf{r} - \mathbf{H}\mathbf{s}) \right]. \quad (4)$$

To solve for the estimator, it is equivalent to find the vector \mathbf{s} which minimizes the scale exponent $(\mathbf{r} - \mathbf{H}\mathbf{s})^T \mathbf{R}_n^{-1} (\mathbf{r} - \mathbf{H}\mathbf{s})$. The solution can be found by solving the partial differential equation $\partial(\mathbf{r} - \mathbf{H}\mathbf{s})^T \mathbf{R}_n^{-1} (\mathbf{r} - \mathbf{H}\mathbf{s}) / \partial \mathbf{s} = \mathbf{0}$. That is, $\partial(\mathbf{r}^T \mathbf{R}_n^{-1} \mathbf{r} - 2\mathbf{r}^T \mathbf{R}_n^{-1} \mathbf{H}\mathbf{s} + \mathbf{s}^T \mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H}\mathbf{s}) / \partial \mathbf{s} = -2\mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{r} + 2\mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H}\mathbf{s} = \mathbf{0}$. If the matrix $\mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H}$ is nonsingular (i.e., if inverse exists), the solution is

$$\hat{\mathbf{s}}_{ML} = (\mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{r}. \quad (5)$$

Furthermore, if there is no knowledge about the correlation of the Gaussian noise vector (or if the elements are mutually independent), it is reasonable to substitute the covariant matrix \mathbf{R}_n by the identity matrix \mathbf{I} . Thus the ML estimator, Eq. (5), is reduced to the *least-squares* (LS) estimator, i.e.,

$$\hat{\mathbf{s}}_{LS} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{r}. \quad (6)$$

It requires that the inverse of the square matrix $\mathbf{H}^T \mathbf{H}$ exists. Recall that the dimensionality of \mathbf{H} is $N \times M$. For the inverse to exist, M needs to be less than or equal to N and the $M \times M$ $\mathbf{H}^T \mathbf{H}$ matrix should be of full rank M . *That is, the number of filters used in the filter-array spectrometer needs to be greater than or equal to the number of resolved points in the wavelength-domain.* For a practical consideration, we take $M = N$, i.e., \mathbf{H} is a square matrix. Then, the LS estimator can be reduced to

$$\hat{\mathbf{s}}_{inv} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{r} = \mathbf{H}^{-1} \mathbf{r}. \quad (7)$$

It is worth to mention that, for zero-mean noise, the $\hat{\mathbf{s}}_{ML}$, $\hat{\mathbf{s}}_{LS}$, and $\hat{\mathbf{s}}_{inv}$ are unbiased, e.g., $E[\hat{\mathbf{s}}_{ML}] = (\mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H}\mathbf{s} = \mathbf{s}$. Therefore, for a fixed unknown source signal vector \mathbf{s} , we may have the received signal vector \mathbf{r} measured multiple times over either the temporal or spatial domain. This unbiased property ensures the enhancement of estimation accuracy after averaging operation.

The estimation-error covariance-matrix of the ML estimator, Eq. (5), can be calculated and expressed as $E[(\hat{\mathbf{s}}_{ML} - \mathbf{s})(\hat{\mathbf{s}}_{ML} - \mathbf{s})^T] = (\mathbf{H}^T \mathbf{R}_n^{-1} \mathbf{H})^{-1}$. We note that it is a function of the filter matrix \mathbf{H} . Thus, it can tell us how good an estimator can be for a particular filter-array. We note that, although the covariance matrix of system noise \mathbf{R}_n is fixed, the variance of the estimation error can be amplified by the detector sensitivity matrix \mathbf{H} . In this paper, we are interested in the case that \mathbf{H} is a square matrix. Conventionally, the singular value decomposition (SVD) is considered as a powerful technique to deal with the noise amplification issue. This method computes the inverse of the \mathbf{H} matrix based on the singular value decomposition where an eigenvalue less than a certain threshold can be discarded. The threshold needs to be carefully chosen. A larger threshold results in a worse approximation of the \mathbf{H} matrix, but less noise amplification.

By exploiting the non-negative nature of the spectral content, we found the non-negative constrained least-squares (NNLS) algorithm work particularly well to estimate the target spectra. NNLS can be seen as a member of the family of the least squares estimator. NNLS returns the vector $\hat{\mathbf{s}}$ that minimizes the norm $\|\mathbf{H}\hat{\mathbf{s}} - \mathbf{r}\|_2$ subject to $\hat{\mathbf{s}} > \mathbf{0}$ [2]. The original design of the algorithm was by C. L. Lawson, and R. J. Hanson [9, p158]. Although the NNLS algorithm solves the solution iteratively, the iteration always converges.

It is worth to point out the weakness of using pseudo-inverse intending for "high" resolution estimation. Consider the same model of Eq. (1), and assume $N < M$, i.e., the number of filters is less than the number of resolved points in the wavelength-domain. Suppose we could find an $M \times N$ matrix \mathbf{H}^* (by any methods) such that $\mathbf{H}^* \mathbf{H} = \mathbf{I}$, where \mathbf{I}

is the identity matrix. Then we could have $E[\mathbf{H}^* \mathbf{r}] = E[\mathbf{s} + \mathbf{H}^* \mathbf{n}] = \mathbf{s}$. In this case, we could recover the source signal vector with high resolution by only using a few filters. However, this supposition can not be fulfilled, and the estimation done this way is always *biased*. Since \mathbf{H} is an $N \times M$ matrix with $N < M$, the $M \times M$ $\mathbf{H}^* \mathbf{H}$ matrix is not possible to be full rank and thus can not be the identity matrix, i.e., $\mathbf{H}^* \mathbf{H} = \mathbf{A}$, where \mathbf{A} is an arbitrary matrix other than the identity matrix. Also the inverse of \mathbf{A} (in the strict sense) does not exist. Therefore, $E[\mathbf{H}^* \mathbf{r}] = E[\mathbf{A} \mathbf{s} + \mathbf{H}^* \mathbf{n}] = \mathbf{A} \mathbf{s}$, the estimated source signal vector is always a biased version of the original input spectrum even in a no noise environment.

4. DSP implementation and experimental results

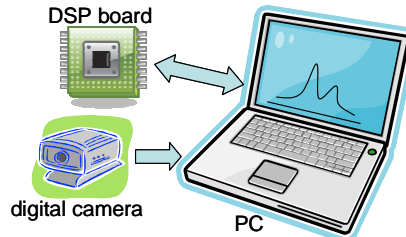


Fig. 2. System set-up for DSP implementation

Figure 2 illustrates the system set-up for our experimental filter-array based spectrometer. A digital camera (IMx-1040FT), a DSP board (TMDSDSK6713), and a personal computer (PC) are used for the preliminary demonstration. The PC serves as a *bridge* and *monitor*. The camera is connected to the PC via the FireWire interface carrying digital signals from each CCD sensor in 8-bit depth. The PC passes the digitalized signals received from the camera to the DSP board via the USB interface. Digital signal processing is performed on the DSP board. The processed data are then acquired by the PC and shown on the PC screen. Forty filters arranged in a straight line configuration are directly placed on top of CCD-sensors. Each of the filters occupies 5 CCD sensors along the horizontal direction. In this system, we only adopt the signal from one CCD sensor near the center of the filter for reducing the possible stray-light effect, which may come from the gap between filters and CCD sensors. We note that this preliminary system serves as a complete prototype for spectrometers on-a-chip consisting of a detector unit and a DSP unit. We implemented both the SVD-inverse algorithm and the NNLS algorithm on the DSP board.

Ideally, if the sensitivity functions are delta-function-like with narrow response ranges, the output directly from the detectors would compose the spectral content. However, the spectral detectors used in this system are far from delta-function-like for the consideration to low-cost fabrication. As illustrated in Fig. 3, the detectors show broad ranges of response while each has its own peak response spot. The shapes of the responses are different as well. Therefore, the spectral outputs obtained directly from the detectors are very different from the input spectral shape.

Figure 4 shows the experimental results of the filter-array based spectrometer measured by shining a red LED whose center peak is at 650nm (LED model: HLMP-4100). The original data were in 10nm intervals. The curves shown in Fig. 4 are obtained by a cubic interpolation processing as suggested by the *Commission Internationale de l'Eclairage* (CIE). Fig. 4(a) depicts the spectrum of the LED provided by the manufacturer. Fig. 4(b) shows the spectral content directly obtained from the detector outputs, whereas Fig. 4(c) shows the estimated spectral content by the NNLS algorithm. The NNLS algorithm shows an excellent estimation of the target spectrum. We have also tried different LEDs whose center peaks vary from 450nm to 700nm. All of the estimations show excellent agreements with the peak-wise target spectra. On the other hand, Fig. 4(d) shows the estimated spectral content by the SVD-inverse

algorithm with the threshold set to zero. We have tried every possible threshold, and note that the results from the SVD-inverse algorithm do not reflect the LED spectra correctly.

We note that the required memory size of both algorithms are dominated by the pre-stored digitalized coefficients of the detector sensitivity matrix \mathbf{H} , and are less than 200 kilobyte. However, the required processing power is significantly different. In our implementation, the total number of *execution cycles* for the NNLS algorithm is 64,898, whereas that for the SVD-inverse algorithm is 1,200,737, which are equivalent to 0.1102 second and 0.0013 second, respectively, execution time on the TI 6713 DSP chip operating at 225 MHz. The NNLS is two orders of magnitude slower than the SVD-based method. Thus, a further research on high-performance but low-complexity algorithms is desired.

5. Summary and conclusions

In this paper, we consider a filter-array based spectrometer in which a low-quality but low-cost filter-array is used. Target spectra can be estimated and recovered accurately through DSP techniques. By exploiting the non-negative property of the spectral content, NNLS algorithm is found particularly useful for this application. Through hardware demonstration, we verified the achievability of a fine spectrometer based on a low-quality, low-cost filter-array.

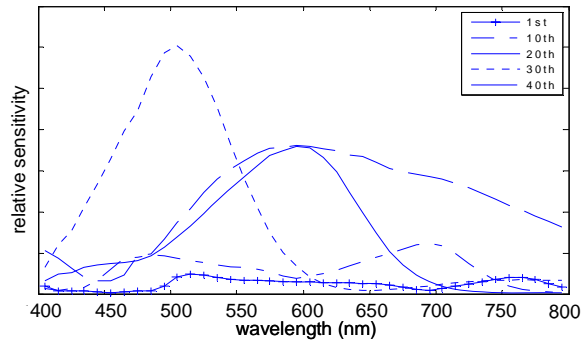


Fig. 3. Sensitivity response of the 1st, 10th, 20th, 30th, and 40th spectral detectors.

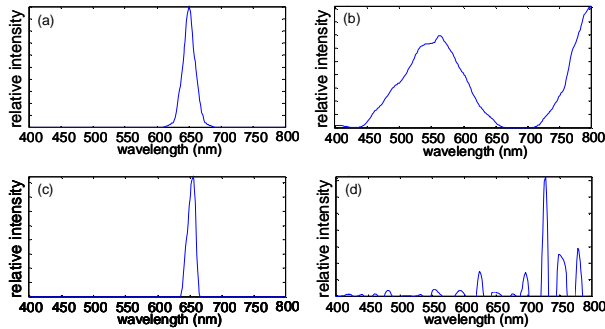


Fig. 4. Experimental results for HLMP-4100 red LED, peak at 650nm: (a) spectrum of the LED provided by the manufacturer, (b) spectrum obtained directly from the spectral detector outputs, (c) spectrum obtained after digital-signal-processing (DSP) based on the NNLS algorithm, and (d) spectrum obtained after DSP based on the SVD algorithm.

Acknowledgments

This project was supported by Nanolambda Inc, Pittsburgh PA, and was carried out in part at the Nanolambda Inc. We would like to take this opportunity to thank Mr. Bill Choi for continuous supports, and Dr. Byounghee Lee and Dr. Min Kyu Song for valuable discussions.