Digital holography for quantitative phase-contrast imaging

Authors: Etienne Cuche, Frederic Bevilacqua, and Christian Depeursinge

Publication: Optics Letters March 1,1999

Speaker: Nitin Rawat

Short summary:

In the paper, they have presented a new application of digital holography for phase-contrast imaging. The technique uses a CCD camera for recording of a digital Fresnel off-axis hologram and a numerical method for hologram reconstruction. The method simultaneously provides an amplitude-contrast image and a quantitative phase-contrast image.

I. INTRODUCTION

The holographic process is described mathematically as follows:

$$O(x,y) = o(x,y)\,\mathrm{e}^{i\varphi_o(x,y)} \qquad \ldots (1.1)$$

Is the complex amplitude of the object wave with real amplitude o and phase φ_o and

$$R(x,y) = r(x,y)\,\mathrm{e}^{i\varphi_R(x,y)} \qquad \ldots (1.2)$$

Is the complex amplitude of the reference wave with real amplitude r and phase φ_R

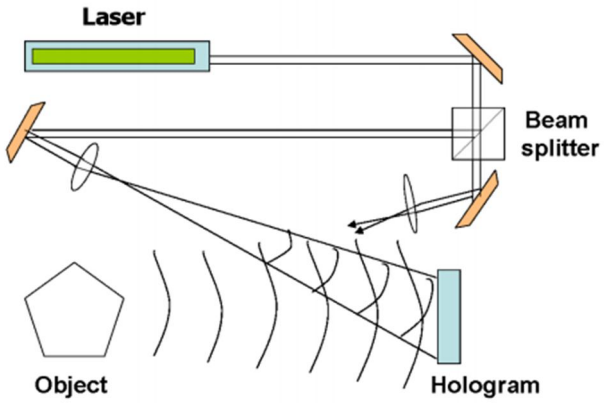


Fig. 1 Construction of a hologram

Both waves interfere at the surface of the recording medium. The intensity is calculated by

$$I_H(x,y) = \left|O(x,y) + R(x,y)\right|^2$$

$$= (O(x,y) + R(x,y))(O(x,y) + R(x,y))^*$$

$$= R(x,y)R^*(x,y) + O(x,y)O^*(x,y)$$
$$+ O(x,y)R^*(x,y) + R(x,y)O^*(x,y) \quad \ldots (1.3)$$

$$I_H(x,y) = |R|^2 + |O|^2 + R^*O + RO^* \quad \ldots (1.4)$$

Digital Recording

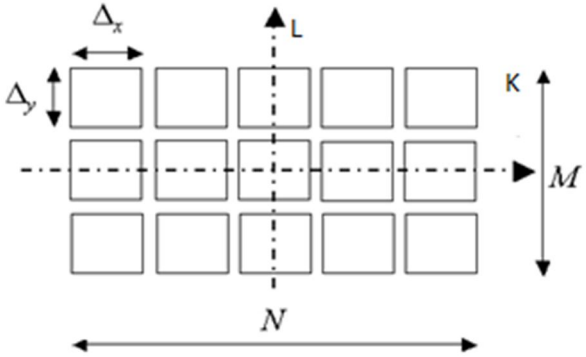


Fig. 2 Concept diagram of a surface-pixel sensor

The digital recording will (depending on the directions x and y on the recording plane) consist of M x N pixels. Each of these pixels is of a dimension $\Delta_x \times \Delta_y$. In CCD sensors, Matrices made up of photosensitive elements called pixels are generally square-shaped. In our case it is $12 \times 12 \mu m$

In their case, the hologram intensity was recorded by a standard black and white CCD camera (Hitachi Denshi KP-M2).

The two neutral-density filters allow the adjustment of the object and the reference intensities.

A square image of area LxL (Sensor size) = 4.83mm X 4.83mm containing N x N= 512 x 512 pixels is acquired in the center of the CCD sensor, and a digital hologram is transmitted to a computer via a frame grabber.

The digital hologram $I_H(k,l)$ results from two-dimensional spatial sampling of $I_H(x,y)$ by the CCD:

$$I_H(k,l) = I_H(x,y)\,rect\left(\frac{x}{L},\frac{y}{L}\right) \times \sum_{k}^{N} \sum_{l}^{N} \delta(x - k\Delta x, y - l\Delta y) \quad \ldots (1.5)$$

Where k and l are integers $(-N/2 \le k,l \le N/2)$ and Δx and Δy are the sampling intervals in the hologram plane i.e. pixel size: $\Delta x = \Delta y$ =L/N

II. NUMERICAL RESULTS

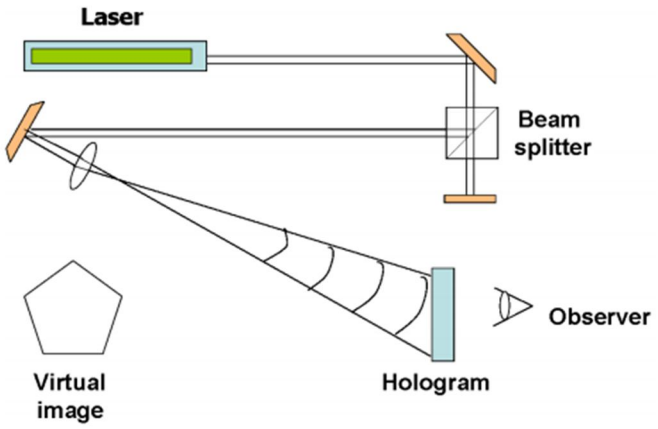


Fig.3 Reconstruction of a hologram

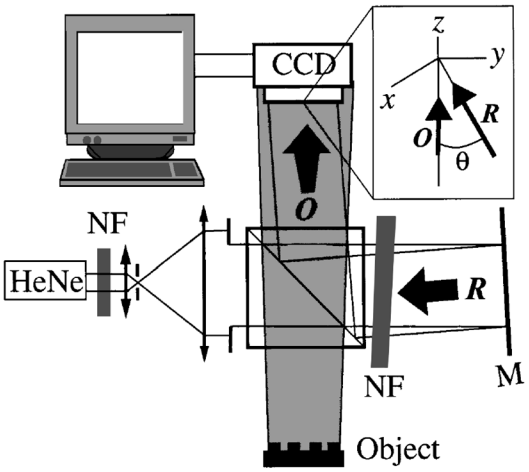


Fig.4 Experimental setup

A wave front $\psi_H(x,y) = R(x,y)I(x,y)$ is transmitted by a hologram and propagates toward an observation plane, where a three-dimensional image of the object can be observed.

For reconstructing a digital hologram, a digital transmitted wave front $\psi_H(k\Delta x, l\Delta y)$ is computed by multiplication of digital hologram $I_H(k,l)$ by a digital computed reference wave, $R_D(k,l)$,called the digital reference wave.

If we assume that mirror M reflects a plane wave of wavelength λ then R_D can be calculated as follows:

$$R_D(k,l) = \underbrace{A_R}_{1} \exp\left[i\frac{2\pi}{\lambda}(k_x k\Delta x + k_y l\Delta y)\right] \quad \text{... (1.6)}$$

Where k_x and k_y are the two components of the wave vector and A_r is the amplitude.
$k_x = -3.12\times10^{-3}$ and $k_y = -5.34\times10^{-3}$

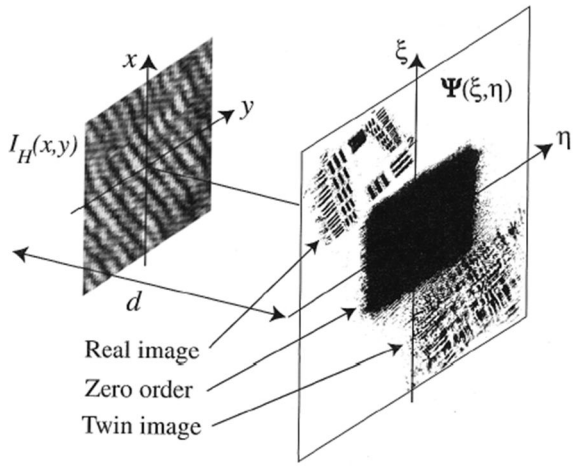
Taking into account the definition of hologram intensity [Eq. 1.4], we have

$$\psi\ (k\Delta x, \Delta y) = R_D(k,l)I_H(k,l) \quad \text{... (1.7)}$$

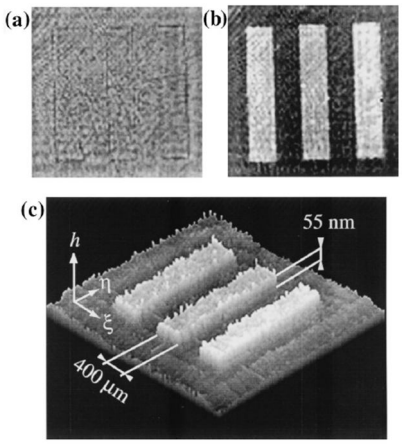
$$= \underbrace{R_D|R|^2 + R_D|O|^2}_{Zero\ order\ of\ diffraction} + \underbrace{R_D R^* O}_{Twin\ image} + \underbrace{R_D R O^*}_{real\ image}$$

To avoid an overlap of these three components of ψ during reconstruction, they recorded the hologram in the so-called off-axis geometry. For this purpose the mirror in the reference arm, M is oriented such that the reference wave **R** reaches the CCD at an incidence angle θ.The value θ must be sufficiently large to ensure separation between the real and the twin images in the observation planes. However, θ must not exceed a given value so that the spatial frequency of the interferogram does not exceed the resolving power of the CCD.

$$\theta \le \theta_{\max} = arc\ \sin\left(\frac{\lambda}{2\Delta x}\right) \qquad \text{... (1.8)}$$



Geometry for hologram reconstruction. *0xy*, hologram plane; *0*ξη, observation plane; *d*, reconstruction distance; **Ψ**(ξ, η), reconstructed wave front.



Reconstructed images obtained with a pure phase object: (a) amplitude contrast, (b) phase contrast, (c) three-dimensional perspective of the reconstructed height distribution (the vertical scale is not equal to the transverse scale).

The reconstructed wave front $\psi(m\Delta\xi, n\Delta\eta)$, at a distance d from the hologram plane, is computed by use of a discrete expression of the Fresnel integral:

$$\psi(m\Delta\xi, n\Delta\eta) = A\exp\left[\frac{i\pi}{\lambda d}(m^2\Delta\xi^2 + n^2\Delta\eta^2)\right] \times FFT\left\{R_D(k,l)I_H(k,l)\exp\left[\frac{i\pi}{\lambda d}(k^2\Delta x^2 + l^2\Delta y^2\right]\right\}_{m,n} \qquad \text{... (1.9)}$$

Where m and n are integers $(-N/2 \le m,n \le N/2)$, FFT is the fast Fourier transform operator, and $A = \exp(i2\pi d/\lambda)/(i\lambda d)$

$\Delta\xi$ and $\Delta\eta$ are the sampling intervals in the observation plane and define the transverse resolution of the reconstructed image.

This transverse resolution is related to the size of the CCD (L) and to the distance d by,

$$\Delta\xi = \Delta\eta = \lambda d / L \qquad \text{... (1.10)}$$

The reconstructed wave front is an array of complex numbers. The amplitude and the phase contrast images can be obtained by calculation of the square modulus.

References

[1] M.Pluta, Non-standard techniques for phase contrast microscopy, Vol. 6 of Advances in Optical and Electron Microscopy

[2] U.Schnars and W. Juptner, Appl. Opt. 33, 179 (1994)

[3] E. Cuche, P. Poscio, and C. Depeursinge, Proc. SPIE 2927, 61 (1996)

Compressed sensing with off-axis frequency-shifting holography

Authors: Marcio M. Marim, Michael Atlan, Elsa Angelini, and Jean-Christophe Olivo-Marin

Publication: Optics letters/ vol. 35, no. 6/ March 2010

Speaker: Hwanchol Jang

Short summary: This work reveals an experimental microscopy acquisition scheme successfully combining compressed sensing (CS) and digital holography in off-axis and frequency-shifting conditions. The authors propose a CS-based imaging scheme for sparse gradient images, acquiring a diffraction map of the optical field with holographic microscopy and recovering the signal from as little as 7% of random measurements.

I. COMPRESSED SENSING

- A signal $g \in R^N$ has a sparse representation if it can be written as a linear com bination of a small set of vectors taken from some basis Ψ, such as $g = \sum_{i=1}^{N} c_i \Psi_i$, with $\|c\|_1 \approx S$ where $S << N$.
- If such a sparsifying transform Ψ exists in the spatial domain, it is possible to reconstruct an image g from partial knowledge of its Fourier spectrum [1].

II. SYSTEM OVERVIEW

- g represents the local optical intensity in the object plane.
- We denote by $f \in C^N$ the associated complex optical field, satisfying $g = |f|^2$.
- The radiation field propagates from the object to the detector plane in Fresnel dif fraction conditions $(d>>(x,y)_{max})$.
- Thus, the optical field in the object plane f is linked to the field F in the detect ion plane by a Fresnel transform, expressed in the discrete case as

$$F = \mathrm{F}(f) : C^N \to C^N,$$
$$F_p = \frac{1}{N}\sum_{n=1}^{N} f_n e^{i(\alpha n^2 - 2\pi np/N)} \quad (1),$$

where p and n denote pixel indices and $\alpha \in R^+$ is the parameter of the quadratic phase factor $e^{i\alpha n^2}$ describing the curvature in the detection plane of a wave emitted by a point source in the object plane.

III. OPTICAL CONFIGURATION [2]

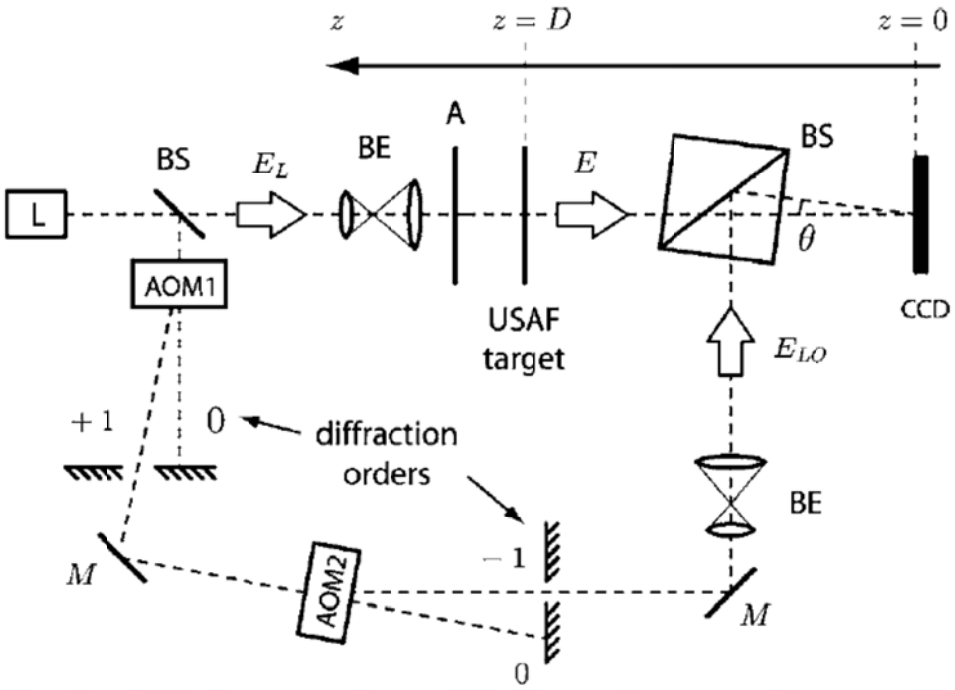


Fig. 1. USAF target digital holography setup: L, main laser; BS, beam splitter; AOM1 and AOM2, acousto-optic modulators; BE, beam expander; M, mirror; A, light attenuator; USAF, transmission target; CCD, CCD camera.

- It consists of an off-axis frequency-shifting digital holography scheme.
- The interfered object beam by the reference beam is recorded at the array detector.
- $F \in C^M$ is calculated from a four-phase measurement.

Phase-shifting digital holography [3].

- The reference beam is dynamically phase shifted with respect to the signal field; the phase shift in the experiment is linear in time (frequency shift).
- It is then possible to obtain the two quadratures of the field in an on-axis configuration even though the conjugate image alias and the true image overlap, because aliases can be removed by taking image differences.
- This shift produces time-varying interferograms on a two dimensional sensor.
- Intensity in the detector plane results from the interference of the signal field with the δf -shifted reference field:

$$I(t) = \left| E_s + E_r \exp i(2\pi \cdot \delta f \cdot t) \right|^2$$

where Es and Er represent the complex amplitudes of the signal and the reference fields, respectively; here E_s=F.

- L intensity $I_l = \left| E_s + E_r \exp i(\frac{2\pi l}{L}) \right|^2$ measurements are performed at $t_l = \frac{l}{\delta f \cdot L}$, $l = 0,...,L-1$; $I_l = |E_s|^2 + |E_r|^2 + E_s^* E_r \exp i(\frac{2\pi l}{L}) + E_s E_r^* \exp -i(\frac{2\pi l}{L})$.
- Obtain Es by demodulating I:

$$E_s = \frac{1}{L \cdot E_r^*} \sum_{l=0}^{L-1} I_l \exp i(\frac{2\pi l}{L}).$$

- The CCD camera records the hologram of the interference with frame rate f_{CCD} =12.5Hz, acquisition time T=1/12.5=80ms. They $\delta f = f_{CCD} / L$ shift the reference beam by combining two acousto-optic modulator (Crystal Technology: $f_{\mathrm{AOM}} \cong 80\mathrm{MHz}$), AOM1 and AOM2, working at $\Delta f + \delta f$ and $-\Delta f$, respectively, with Δf =80MHz. δf =3.125Hz is equal to one quarter of the CCD image frequency (L=4).

The k-space hologram at z=D from the CCD

- $\tilde{E}_s(k_x, k_y, z) = \tilde{K}(k_x, k_y, z)\tilde{E}_s(k_x, k_y, z = 0)$ where $\tilde{K}(k_x, k_y, z)$ is the k-space kernel function that describes the propagation from 0 to z.
- $\tilde{K}(k_x, k_y, z) = e^{iz(k_x^2 + k_y^2)/k}$ where $k = 2\pi / \lambda$ is the optical wave vector.

- Consider the Fresnel propagation of Es from z=0 to z=D, which can be formally expressed as an x and y convolution product (symbol $\otimes$):

$$E_s(x,y,z) = P(x,y,z) \otimes E_s(x,y,z=0),$$

where $P(x,y,z) = \frac{e^{i\cdot kz}}{i\lambda z}\exp\left[i\frac{k}{2z}(x^2+y^2) \right]$.

Combination of "Off axis" and "phase shift".

- With off-axis holography, it is possible to record with a single hologram, the two quadratures of the object complex filed. However, the object field of view is reduced, since one must avoid the overlapping of the image with the conjugate image alias.
- With phase-shifting holography, it is possible to obtain the two quadratures of the field even though the conjugate image alias and the true image overlap, because aliases can be removed by taking image differences.
-

IV. CONVENTIONAL APPROACH

- F can be backpropagated numerically to the target plane with the standard convol ution method when all measurements $F \in C^N$ are available.
- In this case, the complex filed in the object plane f is retrieved from a discrete i nverse Fresnel transform of F; $f = \mathrm{F}^{-1}(F)$,

$$f_p = \frac{1}{N}\sum_{n=1}^{N} F_n e^{-i(\alpha n^2 - 2\pi np/N)} \qquad (2).$$

V. PROPOSED CS APPROACH

- In the CS, the signal reconstruction consists of solving a convex optimization pro blem that finds the candidate $\hat{g}$ of minimal complexity satisfying $\hat{F}|_\Gamma = F|_\Gamma$, whe re $F|_\Gamma \subseteq F$ is a partial subset of measurements in the set Γ.
- We want to recover the intensity image of the object $g = \left\{ |f|^2 : f \in C^N \right\}$ from a small number of measurements $F|_\Gamma \in C^M$, where $M << N$.

- Partial measurements in the detection plane can be written as $F|_\Gamma = \Phi f$, where th e sampling matrix Φ models a discrete Fresnel transformation and a random und ersampling with a flat distribution.
- To find the best estimator $\hat{g}$, we solve the following convex optimization proble m:

$$\hat{g} = \arg\min_{g \in R^N} \|\Psi g\|_1 \text{ subject to } \hat{F}|_\Gamma = F|_\Gamma .$$

- Since the test image is piecewise constant with sharp edges (such as most micros copy images), it can be sparsely represented computing its gradient.
- In image processing, a suitable norm to constrain the gradient of an image was i ntroduced as the total variation (TV) which measures the l-1 norm of the gradien t magnitudes over the whole image, $\|g\|_{TV} = \|\nabla g\|_1$.
- The incoherence property holds for the two bases adopted here which are the Fre snel spectrum and the TV. Moreover, random measurements in the spectral domai n satisfy the RIP condition.
- Hence for an overwhelming percentage of Fresnel coefficients sets Γ with cardin ality obeying $|\Gamma| = M \geq KS \log N$, for some constant *K*, $\hat{g}$ is the unique solution to the problem,

$$\hat{g} = \arg\min_{g \in R^N} \|\nabla g\|_1 \text{ subject to } \hat{F}|_\Gamma = F|_\Gamma .$$

- The reconstruction of g with robustness to noise:

$\hat{g} = \arg\min_{g \in R^N} \|\nabla g\|_1$ subject to $\left\|\hat{F}|_\Gamma - F|_\Gamma\right\|_2 \leq \delta$, for some $\delta \leq C\varepsilon$,
which depends on the noise energy.

VI. RESULTS

Standard convolution method (eq. 2) and CS approach.

- For the CS approach, Fresnel coefficients are undersampled randomly
- Figure 3(b) shows the CS reconstruction result from only 7% of the pixels used in the standard approach.
- Figure 3(d) illustrates the residual (Euclidean distance $|\hat{g} - g|$) from standard holographic and CS reconstructions. The global normalized error is $\|\hat{g} - g\|_2$ =0.005.

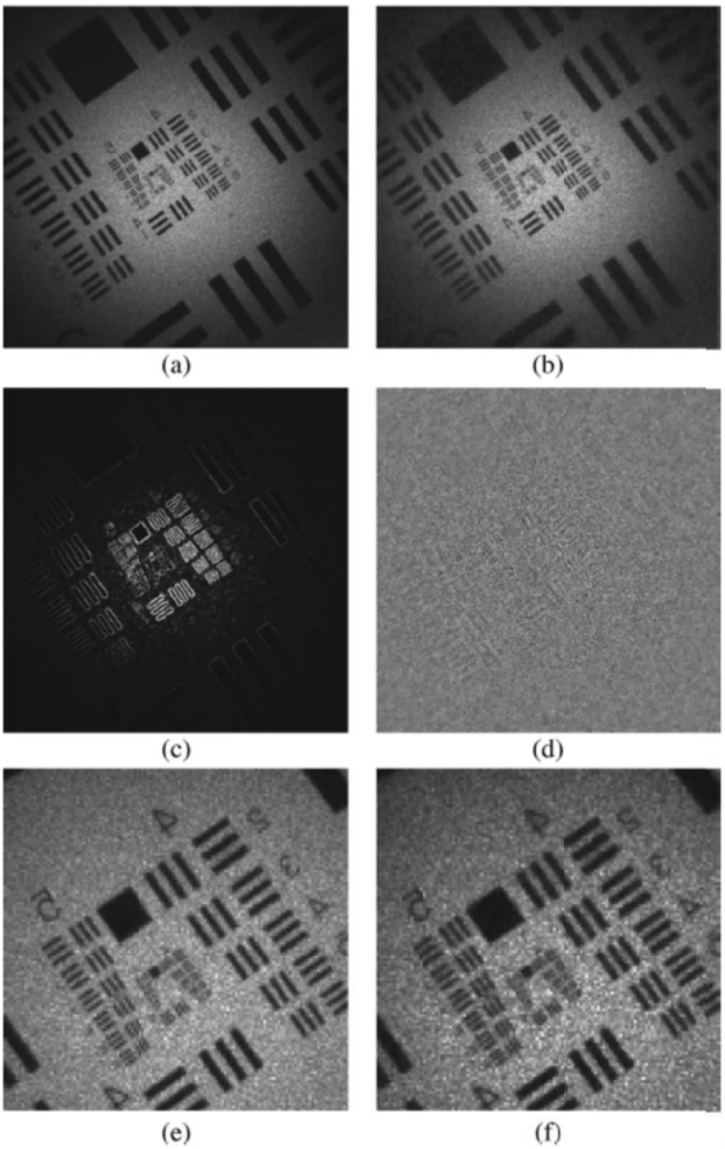


Fig. 3. (a) Standard holography, as described in Eq. (1). (b) CS reconstruction, using 7% of the Fresnel coefficients. (c) Gradient of g. (d) Residual from (a) and (b). (e), (f) Magnified views from (a) and (b).

Reference

[1] E. Candes and J. Romberg, “Practical signal recovery from random projections,” Proc. SPIE, 5674, Jan. 2005.
[2] M. Gross and M. Atlan, “Digital holography with ultimate sensitivity,” Opt. Lett., Vol. 32, no. 8, Apr. 2007.
[3] F. Le Clerc and L. Collot, “Numerical heterodyne holography with two-dimensional photodetector arrays,” Opt. Lett., vol. 25, no. 10, May 2000.

VII. FURTHER DISCUSSION

- How do they check $\hat{F}|_\Gamma = F|_\Gamma$ where $\hat{F}|_\Gamma = \Phi\hat{f}$ and $F|_\Gamma = \Phi f$; it looks difficult as they estimate $g = |f|^2$ not f.

An EEG-Based BCI System for 2-D Cursor Control by Combining Mu/Beta Rhythm and P300 Potential

Yuanqing Li et al.

IEEE Trans. Biomedical Eng. (2010.10)

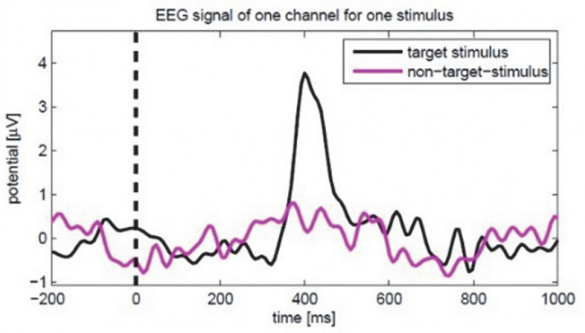
Presenter : Younghak Shin

GIST, Dept. of Information and Communication, INFONET Lab.

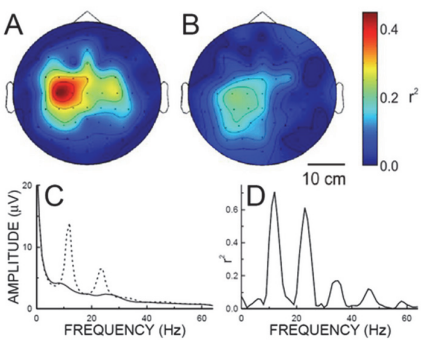
Gwangju Institute of Science and Technology

Background

- P300



- Motor imagery (Mu/Beta Rhythm)



Introduction & Motivation

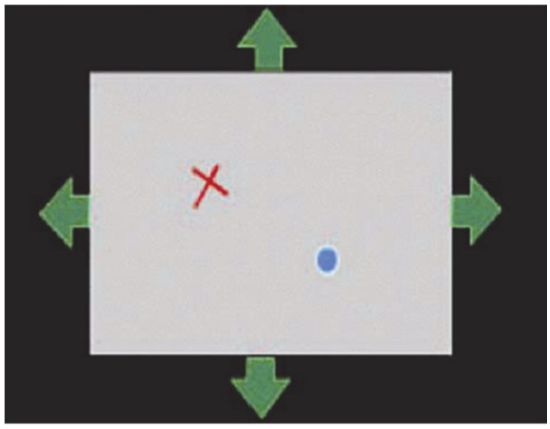
- Brain computer interface is a communication and control pathway to directly translate brain activities into computer control signals.
- An important issue in BCI research is cursor control: to map brain signals to movements of a cursor on a computer screen.
- Cursor control is a basic and necessary application in BCIs.
- In the EEG-based BCI field, most related studies were focused on 1-D cursor control using motor imagery feature. (our experiment)
- Multidimensional cursor control enables enhanced interfacing between the user and the machine (much wider range of applications)

Introduction & Motivation

- The development of 2-D control BCI is impeded by the difficulty in obtaining two independent control signals from the noisy EEG data of poor spatial specificity.
- The first EEG-based 2-D cursor control BCI is remarkable [Wolpaw 2004]: two independent control signals could be derived from combinations of the rhythmic powers(mu and beta).
- However, the downside of this approach is the required intensive user training (some weeks or months).
- Also other forms of 2-D BCI were reported that adopted a discrete control paradigm using the steady-state visual evoked potential (SSVEP) and P300.

Introduction & Motivation

- P300 and SSVEP based BCI cursor control system



- For each update of the position, the cursor takes only one of a few fixed directions (discrete).
- This leads to unsmooth, zigzag like moves of the cursor.

Introduction & Motivation

- More efficient 2-D control is required for real world application
- This could be implemented through the combination of two independent control signal for vertical and horizontal movements
- In this paper, they propose a new paradigm for 2-D cursor control by simultaneously detecting two brain signals; P300 and motor imagery (hybrid BCIs)
- The user control the vertical movement by using a P300.
- At the same time, the user also use a motor imagery to control the horizontal movement.

Introduction & Motivation

- The P300 allows to select one of the three states: moving upwards, moving downwards, or no vertical movement.
- The motor imagery translates motor imagery into a continuous value that determine the direction and the velocity of the horizontal movement.
- Through selecting the vertical movement state and manipulating the horizontal speed with direction, users will move the cursor along any direction in a self-paced manner.
- Since the neural mechanism of motor imagery differs largely from that of P300, the two signals can be independently controlled by the user.

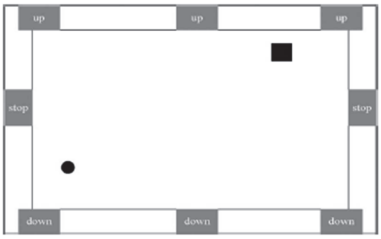
Method

- Experimental Procedure

- To evaluate the proposed method, they conduct an online experiment involving six subjects.
- After a short calibration(training) session for subject-specific P300 and motor imagery modeling,
- the subjects undergo a few sessions of feedback training using the GUI, and subsequently perform cursor control tasks in test sessions.
- In particular, the feedback training emphasizes on learning of motor imagery control.

Method

- GUI and Control Mechanism



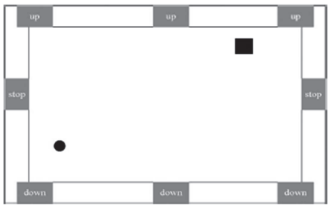
- The ball and the square, respectively, represent a cursor and a target.
- the GUI has a dimension of 1166 pixels ×721pixels.

- In each trial, the initial cursor and target randomly appear inside a rectangular workspace.
- Eight flashing buttons are located at the side, which would generate P300 potentials when the user focus attention on one of them.
- The main objective for the arrangement of three “up” and “down” buttons is for user’s convenient use.
- For example, if the cursor is now in the right-hand side, then the user can choose the “up” or “down” button located in the same side
- The horizontal movement of the cursor is controlled by motor imagery.

Method

- Test trial

- A trial begins when a target and a cursor appear.
- At 100 ms later, the eight buttons begin to flash alternately in a random order.
- Each button is intensified for 100 ms, while the time interval between two consecutive button flashes is 120 ms.
- The trial ends when the cursor hits the target or when the 60-s timeout limit occurs.
- The interval between two consecutive trials is 2 s.
- The cursor’s position is updated every 200 ms; the user is able to control the cursor’s movement at any time at will (self-paced)



Method

- Control of Vertical Movement Based on P300 Potential

- Let $c(k) \in \{1, -1, 0\}$ represents the output: 1 for "down," −1 for "up," and 0 for "stop" (applied to vertical movement only).
- The vertical movement model: $y(k + 1) = y(k) + c(k)v_0$
- where $y(k)$ represents the vertical position at the k-th update, at a fixed interval of 200 ms and v_0 is a positive speed constant.
- If the output $c(k)$ is 0, the cursor stops vertical movement; if the output is 1 or −1, the cursor moves up or down at a speed of 10 pixels every 200 ms.

Method

- Algorithm (vertical)

1. Feature extraction
   I. 30 channel EEG signals are filtered in the range of 0.1–20 Hz
   II. Extract a segment from 0 to 600 ms after a button flashes
   III. The obtained data vector is denoted as $P_{ai,j,q}$, where $i, j,$ and q represent the i th channel, the j th button associated with this flash, and the q th round,
   IV. Construct a feature vector corresponding to the j th button and the q th round by concatenating the vectors $P_{ai,j,q}$ from all channels, i.e.,

$$F_{ej,q} = [P_{a1,j,q}, \ldots, P_{a30,j,q}].$$

2. Train a support vector machine (SVM) classifier.
   I. Using the feature vectors of the training dataset $\{F_{ej,q}, j = 1, \ldots, 8; q = 1, \ldots, 64\}$ and their corresponding labels, we train an SVM classifier.
   II. If the subject is focusing on the j th button in the q th round,
   then the label of $F_{ej,q}$ is 1 (p300 presence),
   while the labels of $F_{eu,q}$ ($u \neq j$) is −1 (p300 absence).

Method

3. P300 detection with SVM scores

I. For the l th round, we extract feature vectors $Fe_{j,l}$ (j = 1, . . . , 8)

II. Applying the trained SVM to $Fe_{j,l}$, we obtain eight scores denoted as $s_{j,l}$

III. P300 detection requires discriminating between non-P300(idle state) and P300 rounds ➡ self-paced

IV. The detection of P300 is accomplished by a threshold mechanism

- Calculate the sum of SVM scores for each button obtained from l accumulated rounds
- Find maximum and the second maximum of the eight summed scores

$$ss_j = s_{j,1} + \cdots + s_{j,l},\ j = 1, \ldots, 8$$
$$ss_{j_0} = \max\{ss_1, \ldots, ss_8\}$$
$$ss_{j_1} = \max\{\{ss_1, \ldots, ss_8\} \backslash \{ss_{j_0}\}\}$$

- where the j_1 th button is obviously different from the j_0 th button
- define a threshold condition : $1 - \frac{ss_{j_1}}{ss_{j_0}} > \theta_0$
- where the threshold θ_0 is a predefined positive constant, and is empirically set at 0.3
- If the threshold condition is satisfied, then the system makes a decision that P300 potential occurs at the j_0 th button

Method

4. Movement Decision

I. when P300 is detected for the l th round at the j_0 button

II. if the j_0 button is one of the three “up” buttons, set c(k) = −1 (the cursor will go up);

III. if the j_0 button is one of the two “stop” buttons, set c(k) = 0 (the cursor will have no vertical movement).

IV. If P300 is not detected for the l th round, i.e., the threshold condition is not satisfied,

V. then the system does not change the direction of the vertical movement of the cursor, i.e., c(k) = c(k − 1).

Method

- Control of the Horizontal Movement Through Motor Imagery

- The horizontal movement model:

$$x(k+1) = x(k) + \frac{a}{3}(f(k-2) + f(k-1) + f(k)) + b$$

- where *k* represents the *k* th update of the cursor position, *x*(*k*) is the horizontal coordinate of the cursor, *f*(*k*) is the continuous output (score) of the SVM classifier, and *a* and *b* are two constants.
- They introduce delays into the control model to make the cursor move smoothly
- The score *f*(*k*) is generated by an SVM classifier at every 200 ms.
- pre-processing steps
  - spatial filtering with common average reference (CAR)
  - bandpass filtering in specific mu rhythm band (8–13 Hz)
  - spatial filtering based on a common spatial pattern (CSP) matrix W.
- The feature vector is fed into an SVM classifier that is trained using the training data to separate the two classes of feature vectors

Method

- In motor imagery, the cursor shall not move, if the user is in idle state.
- This is accomplished by the introduction of the two parameters: *a* and *b*
  - An EEG dataset is collected when the user is in idle state. This dataset contains N time segments of 200 ms (N = 600), i.e., lasts 2 min.
  - They calculate the SVM scores f(1), . . . , f(N).
  - Set

$$m = \frac{1}{N}\sum_{k=1}^{N} f(k)$$

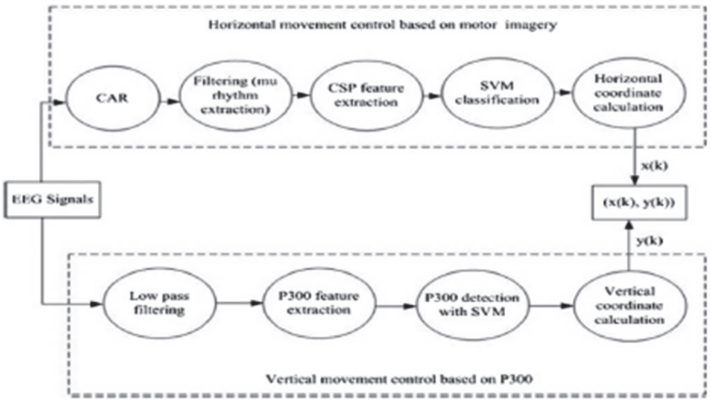
$$mi = \min\{f(k),\ k = 1, \ldots, N\}$$

$$ma = \max\{f(k),\ k = 1, \ldots, N\}.$$

  - Then we calculate *a* and *b* as $a = \dfrac{h}{\max\{ma - m, m - mi\}}$ $b = -am.$
  - The parameter *h* is used for adjusting the velocity of the cursor. *h* is fixed to 8 for all the subjects.
  - Considering the model : $x(k+1) = x(k) + \frac{a}{3}(f(k-2) + f(k-1) + f(k)) + b$
  - and the parameter setting of *a* and *b*, the average horizontal movement during the idle state $\frac{1}{N}\sum_{k=1}^{N-1}(x(k+1) - x(k))$ Is close to zero.
  - Thus, the cursor almost does not change its horizontal position, if the subject is in the idle state. This has been demonstrated during the online experiments.

Method

- Combining the algorithms for vertical and horizontal movement control, they obtain the algorithm for 2-D cursor control



[Diagram of the algorithm for 2-D cursor control]

- The horizontal movement control based on motor imagery and the vertical movement control based on P300 are performed simultaneously

results

	Number of trials	Hit rate (%)	Average time (s)
Subject A	80	97.5	24.8
Subject B	80	91.25	25.0
Subject C	80	86.25	30.7
Subject D	80	84.5	28
Subject E	80	92.5	31.8
Subject F	80	92.5	25.6

- we can find out that the accuracy rates of all the six subjects are satisfactory. However, the control time of each trial was not short (about 28 s in average).
  - The relative small size of the cursor and the target.
  - The ratios of the size of the cursor, target, and the workspace are fixed to be 0.00084:0.003:1.
  - Triggering and effectively detecting P300 are time consuming to some degree (repeated round).

Data Analysis and Discussions

- For efficient 2-D control, the two control signals need to be as independent as possible.
- First, analyze the data collected from online experiment and assess the independence between the two control signals by correlation method.
- There exist horizontal and vertical control variables with values $f(k)$, $c(k)$
- Here, $f(k)$ is an SVM score from motor imagery, while $c(k)$ with value of 1, −1, or 0 is determined by the button
- The button at which P300 occurs is determined by $ss_{j0}(k)$, the maximum of 8 SVM scores corresponding to the eight buttons.
- Since $ss_{j0}(k)$ is generally positive, they let $c(k)ss_{j0}(k)$ to be a new vertical control variable with signs and amplitude.
- For each trial, they also define two position variables representing the relative position of the cursor and the target

$$\bar{x}(k) = x_t - x_c, \quad \bar{y}(k) = y_t - y_c$$

- where t and c means the target and the cursor (at its initial position)

Data Analysis and Discussions

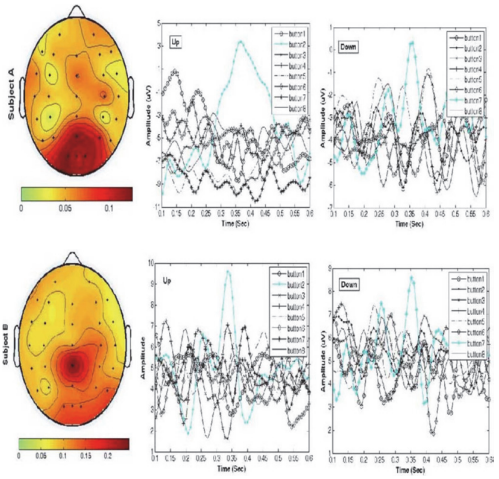
- Calculate the correlation coefficients for each pair of control variable ($f(k)$ or $c(k)ss_{j0}$ (k)) and position variable ($\bar{x}(k)$ or $\bar{y}(k)$)
- and the correlation coefficient for the two control variables.

Subject	$C_{mu,\bar{x}}$	$C_{mu,\bar{y}}$	$C_{P300,\bar{x}}$	$C_{P300,\bar{y}}$	$C_{P300,mu}$
A	0.685	0.016	0.038	0.653	0.042
B	0.521	0.008	0.044	0.566	0.004
C	0.432	0.048	0.013	0.407	0.028
D	0.443	0.019	0.041	0.506	0.021
E	0.501	0.001	0.024	0.321	0.001
F	0.411	0.035	0.008	0.417	0.012

- From the result, each control variable correlates strongly with its own dimension of target cursor position
- and does not correlate with the other variables' dimension
- and also does not correlate with the other variable.
- Hence, such results support the independence of the horizontal and vertical control variables.

Data Analysis and Discussions

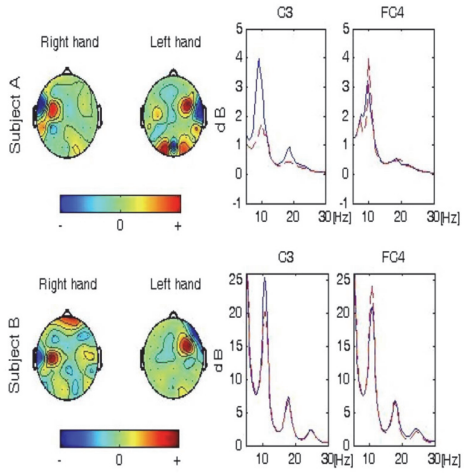
- Next, they present their data analysis results to show that the subjects truly used P300 and motor imagery in the 2-D cursor control.



- P300 detection is mainly based on those channels located in the occipital and parietal areas rather than the frontal area (considered EOG).
- we can see P300 potentials elicited by the two chosen buttons and no P300 potentials elicited by other buttons.

Data Analysis and Discussions

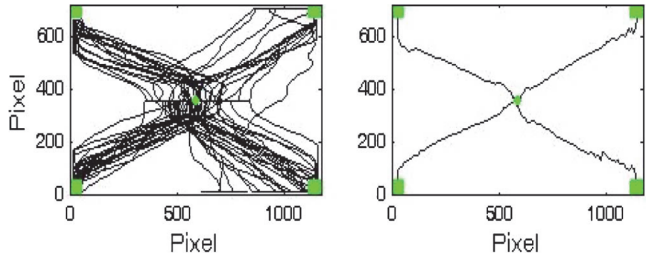
- They also plot the topographies of CSP filters and the power spectra of two channels of raw EEG signals calculated based on the training dataset.



- CSP filters (the first and the last rows of **W**) are displayed which are easily related to the motor imageries of right and left hands
- The discriminability of the brain signals corresponding to the right and left hands is demonstrated.

TWO SUPPLEMENTARY EXPERIMENTS

- In order to show the average trajectory of the cursor based on desired movements, the trajectories of the cursor starting at the center of the GUI and moving to one of the four fixed targets at four corners

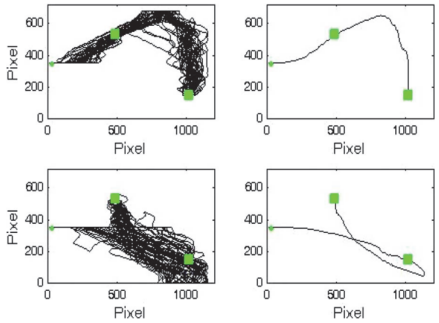


Left: trajectories of the cursor obtained in 80 trials. Right: four average trajectories

- The hit rate and the average control time are 97.5% and 24.7 s
- The trajectories of single trials and the average trajectories for four target positions are quite smooth.

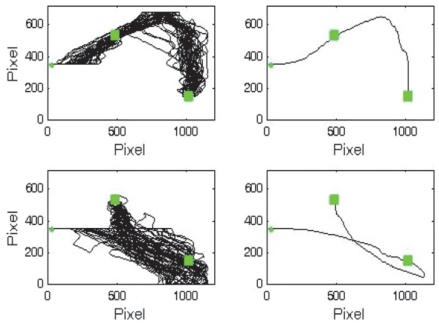
TWO SUPPLEMENTARY EXPERIMENTS

- Second experiment, to show P300 and motor imagery control work simultaneously.



- fix two targets, and fix the initial position of the cursor at the middle left
- In each trial, the subject needs to move the cursor to hit these two targets sequentially without stop. (two classes of trial)
- first class, the subject move the cursor to first hit the target at the middle of upper part and then hit the target at down-right corner
- second class, the subject move the cursor to first hit the target at down right corner and then hit the target at the middle of upper part.

TWO SUPPLEMENTARY EXPERIMENTS



- 80 trials were collected for two classes of trials (40 trials for each class)
- The hit rate and the average control time are 98.75% and 49.1 s
- for each trial, the direction of vertical movement of the cursor is changed while the cursor simultaneously moves to the right.
- Thus, the cursor can be simultaneously controlled (no zigzag) by the two control signals of P300 and motor imagery.

Conclusion

- They have presented a new BCI and its implementation for 2-D cursor-control by combining the P300 potential and motor imagery.
- Two independent signals have been obtained for controlling two degrees of movements of a cursor simultaneously.
- Through an online experiment involving six subjects performed 2-D cursor control tasks.
- The six subjects successfully carried out 2-D cursor control with satisfactory accuracies (>80%).

Thank you

Robust Face Recognition via Sparse Representation

Authors:	J. Wright, Allen, Y. Yang, A. Ganesh, S. Shankar, and Yi Ma
Publication:	IEEE Trans. On Pattern Analysis and Machine Intelligence, Feb.2009
Speaker:	Woongbi Lee

Short summary: In this paper, sparse signal representation is investigated for recognizing human faces from frontal views with varying expression and illumination, as well as occlusion and disguise. Based on a sparse representation computed by l^1-minimization, this face recognition problem is cast as a general classification among multiple linear regression models. Even with severe occlusion and corruption, their algorithms show high performance classification of high dimensional data.

I. INTRODUCTION

In this paper, the discriminative nature of sparse representation for classification is exploited. Instead of using the generic dictionaries, the test sample is represented in an overcomplete dictionary whose base elements are the training samples themselves. If sufficient training samples are available from each class, it will be possible to represent the test samples as a linear combination of just those training samples from the same class. This representation is naturally sparse, involving only a small fraction of the overall training database. In many problems of interest, it is actually the sparsest linear representation of the test sample in terms of this dictionary and can be recovered efficiently via l^1-minimization.

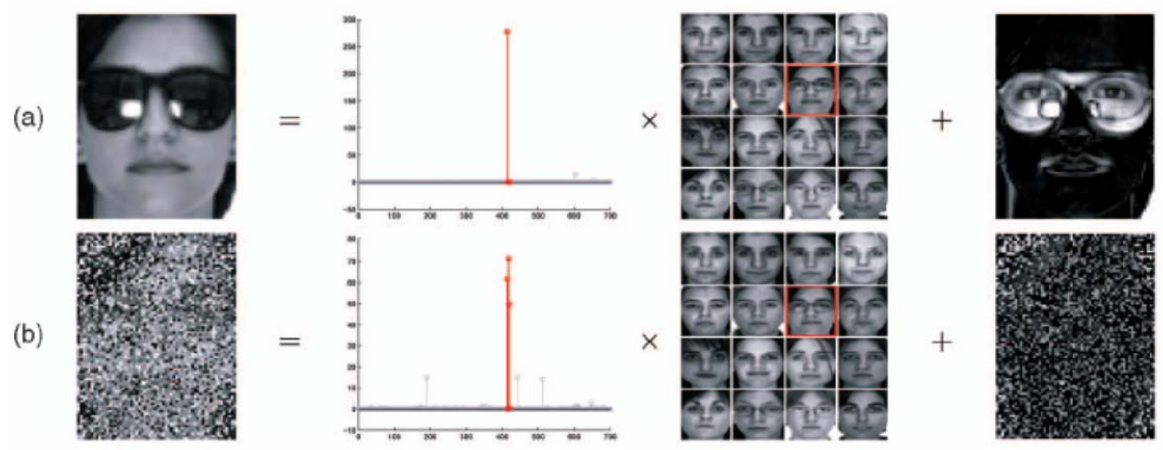


Fig. 1 Simple Example of face recognition: (a) occluded (b) corrupted

There are 700 training images of 100 individuals (7 each). A test image (left) is represented as a sparse linear combination of all the training images plus sparse errors due to occlusion or corruption. Red coefficients correspond to training images of the correct individual.

The theory of sparse representation and compressed sensing yields new insights into two crucial issues in automatic face recognition: the role of feature extraction and the difficulty due to occlusion.

The role of feature extraction: Which low-dimensional features of an object image are the most relevant or informative for classification is a central issue in face recognition. As conventional methods, there exist Eigenfaces (PCA), Fisherfaces (LDA), Laplacianfaces, and a host of variants. The theory of compressed sensing implies that the precise choice of feature space is no longer critical: Even random features contain enough information to recover the sparse representation and hence correctly classify any test image. What is critical is that the dimension of the feature space is sufficiently large and that the sparse representation is correctly computed.

Robustness to occlusion: Occlusion brings about significant troubles in face recognition. This is due to the unpredictable nature of the error occurred by occlusion. Typically, this error corrupts only a small part of the image pixels and therefore is sparse in the standard basis given by individual pixels. The sparse representation of an occluded test image naturally separates the component of the test image arising due to occlusion from the component arising from the identity of the test subject.

II. Classification Based on Sparse Representation

In this section, they introduce classification using sparse representation and show that the sparse representation can be computed by l^1-minimization and can be used for classifying and validating any given test sample.

Object recognition aims to use labeled training samples from *k* distinct object classes to correctly determine the class to which a new test sample belongs. n_i training samples from the *i*-th class form a matrix $A_i \doteq \left[v_{i,1}, v_{i,2}, \cdots, v_{i,n_i} \right] \in \mathbb{R}^{m \times n_i}$ whose columns are the training face images of the *i*-th subject. The image is represented by $w \times h$ gray scale with the vector $v \in \mathbb{R}^m$ ($m = wh$).

A. Test Sample as a Sparse Linear Combination of Training Samples

For the structure of the A_i for recognition, one particularly simple and effective approach models the samples from a single class as lying on a linear subspace. Given sufficient training samples of the *i*-th object class, $A_i = \left[v_{i,1}, v_{i,2}, \cdots, v_{i,n_i} \right] \in \mathbb{R}^{m \times n_i}$, any new test sample $y \in \mathbb{R}^m$ from the same class will approximately lie in the linear span of the training samples associated with object *i*:

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \cdots + \alpha_{i,n_i} v_{i,n_i}$$

for some scalars, $\alpha_{i,j} \in \mathbb{R}$, $j = 1, 2, \ldots, n_i$.

A new matrix **A** is defined for entire training set as the concatenation of the *n* training samples of all *k* object classes:

$$A_i \doteq \left[A_1, A_2, \cdots, A_k \right] = \left[v_{i,1}, v_{i,2}, \cdots, v_{i,n_k} \right]$$

Then, the linear representation of y can be written as

$$y = A x_0 \in \mathbb{R}^m$$

where $x_0 = \left[0, \cdots, 0, \alpha_{i,1}, \alpha_{i,2}, \cdots, \alpha_{i,n_i}, 0, \cdots, 0 \right]^T \in \mathbb{R}^n$ is a coefficient vector whose entries are zero except those associated with the *i*-th class.

Obviously, if $m > n$, the system of equations $y = Ax$ is over-determined, and the correct x_0 can usually be found as its unique solution. In robust face recognition, the system $y = Ax$ is

typically under-determined, and so, its solution is not unique. Conventionally, this difficulty is resolved by choosing the minimum l^2-norm solution.

$$\left(l^2\right): \hat{x}_2 = \arg\min \|x\|_2 \quad \text{subject to} \quad Ax = y$$

While this optimization problem can be easily solved by the pseudo-inverse of A, the solution $\hat{x}_2$ is not especially informative for recognizing the test sample y. To resolve this difficulty, they instead exploit the following optimization problem:

$$\left(l^0\right): \hat{x}_0 = \arg\min \|x\|_0 \quad \text{subject to} \quad Ax = y$$

where $\|\cdot\|_0$ denotes the l^0-norm, which counts the number of nonzero entries in a vector. However, the problem of finding the sparsest solution of an underdetermined system of linear equations is NP-hard and difficult even to approximate.

B. Sparse Solution via l^1-Minimization

If the solution x_0 sought is sparse enough, the solution of the l^0-minimization problem is equal to the solution to the following l^1-minimization problem.

$$\left(l^1\right): \hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to} \quad Ax = y$$

This problem can be solved in polynomial time by standard linear programming methods.

1) Dealing with Small Dense Noise

Since real data are noisy,

$$y = Ax_0 + z$$

where $z \in \mathbb{R}^m$ is a noise term with bounded energy $\|z\|_2 < \varepsilon$. The sparse solution x_0 can be approximately recovered by solving the following stable l^1-minimization problem:

$$\left(l_s^1\right): \hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to} \quad \|Ax - y\|_2 \le \varepsilon$$

This convex optimization problem can be efficiently solved via second-order cone programming.

C. Classification Based on Sparse Representation

Given a test sample y from one of the classes in the training set, the sparse representation $\hat{x}_1$ is computed. Noise and modeling error may lead to small nonzero entries associated with

multiple object classes. Based on the global sparse representation, one can design many possible classifiers to resolve this. By using linear structure, we instead classify *y* based on how well the coefficients associated with all training samples of each object reproduce y. For each class *i*, let $\delta_i : \mathbb{R}^n \to \mathbb{R}^n$ be the characteristic function that selects the coefficients associated with the *i*-th class. For $x \in \mathbb{R}^n$, $\delta_i(x) \in \mathbb{R}^n$ is a new vector whose only nonzero entries are the entries in *x* that are associated with class *i*. Using only the coefficients associated with the *i*-th class, one can approximate the given test sample y as $\hat{y}_i = A\delta_i(\hat{x}_1)$. We then classify y based on the these approximations by assigning it to the object class that minimizes the residual between y and $\hat{y}_i$:

$$\min_i r_i(y) \doteq \left\| y - A\delta_i(\hat{x}_1) \right\|_2$$

The recognition procedure is summarized in Algorithm 1. Implementation is done by minimizing the l^1-norm via a primal-dual algorithm for linear programming.

Algorithm 1: Sparse Representation-based Classification (SRC)
1: **Input**: a matrix of training samples $A = [A_1, A_2, \cdots, A_k] \in \mathbb{R}^{m \times n}$ for *k* classes, a test sample $y \in \mathbb{R}^m$, (and an optional error tolerance $\varepsilon > 0$.)
2: Normalize the columns of A to have unit l^2-norm.
3: Solve the l^1-minimization problem: $\hat{x}_1 = \arg\min_x \|x\|_1$ subject to $Ax = y$. (Or alternatively, solve $\hat{x}_1 = \arg\min_x \|x\|_1$ subject to $\|Ax - y\|_2 \le \varepsilon$.)
4: Compute the residuals $r_i(y) \doteq \left\| y - A\delta_i(\hat{x}_1) \right\|_2$ for $i = 1, \cdots, k$.
5: **Output**: identity $(y) = \arg\min_i r_i(y)$.

Example 1. Original image: 192 x 168, downsampled: 12 x 10, size of matrix *A* is 120 x 1207, 38 classes

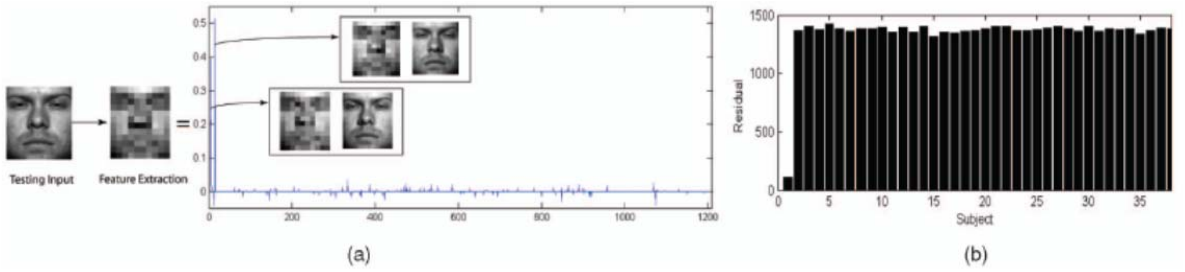


Fig. 2 A valid test image: (a) coefficients (b) residuals

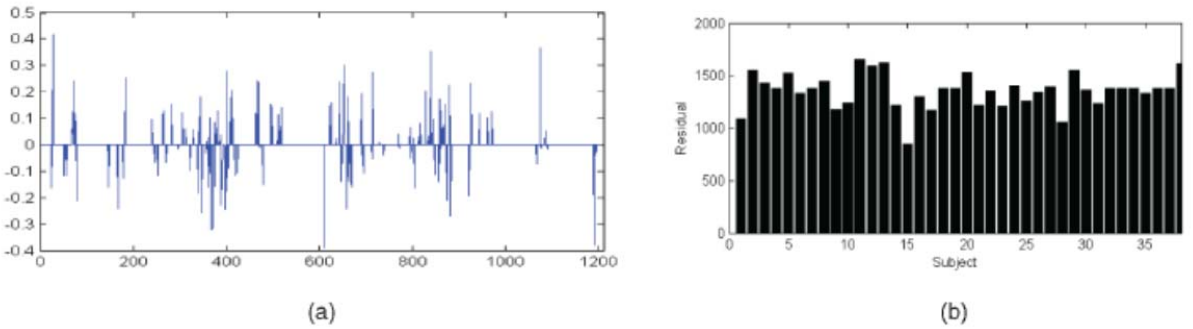


Fig. 3 Non-sparsity of the l^2-minimizer: (a) coefficients (b) residuals

D. Validation Based on Sparse Representation

Before classifying a given test sample, test for validity of the sample is necessary.

Example 2. Randomly selecting an irrelevant image from Google and downsample it to 12 x 10 in the example 1

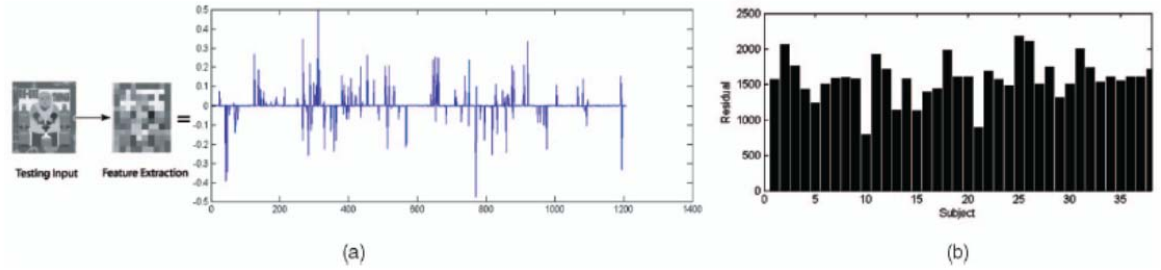


Fig. 4 Example of an invalid test: (a) coefficients (b) residuals

Definition 1 (sparsity concentration index (SCI))

$$SCI(x) \doteq \frac{k \cdot \max_i \left\|\delta_i(x)\right\|_1 / \|x\|_1 - 1}{k-1} \in [0,1]$$

If $SCI(\hat{x}) = 1$, the test image is represented using only images from a single object, and if $SCI(\hat{x}) = 0$, the sparse coefficients are spread evenly over all classes. So, we can choose a threshold $\tau \in (0,1)$ and do the validity test.

If $SCI(\hat{x}) \geq \tau$, then the test image is valid.

III. Two Fundamental Issues in Face Recognition

A. The Role of Feature Extraction

One benefit of feature extraction, which carries over to the proposed sparse representation framework, is reduced data dimension and computational cost. Conventionally, on class of methods extract holistic face features such as Eignefaces, Fisherfaces, and Laplacianfaces. Another class of methods tries to extract meaningful partial facial feature such as patches around eyes or nose.

Since most feature transformations involve only linear operations, the projection from the image space to the feature space can be represented as a matrix $R \in \mathbb{R}^{d \times m}$ with $d \ll m$.

$$\tilde{y} \doteq Ry = RAx_0 \in \mathbb{R}^d$$

The system of equations $\tilde{y} = RAx \in \mathbb{R}^d$ is underdetermined in the unknown $x \in \mathbb{R}^n$. By solving the following l^1-minimization problem,

$$(l_r^1): \hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to} \quad \left\|RAx - \tilde{y}\right\|_2 \leq \varepsilon$$

for a given error tolerance $\varepsilon > 0$.

For the sparse representation approach to recognition, it is important how the choice of the feature extraction R affects the ability of the l^1-minimization to recover the correct sparse solution x_0. There is remarkable analysis: if the solution x_0 is sparse enough, then with overwhelming probability, it can be correctly recovered via l^1-minimization from any sufficiently

large number d of linear measurements $\tilde{y} = RAx_0$. In other words, if x_0 has $t << n$ nonzeros, then with overwhelming probability, $d\left(\geq 2t\log\left(n/d\right)\right)$ random linear measurements are sufficient for l^1-minimization to recover the correct sparse solution x_0. Random features can be regarded as a less-structured counterpart to classical face features such as Eignefaces or Fisherfaces. The linear projection generated by a Gaussian random matrix is called Randomfaces.

Definition 2. Randomfaces.

- a transform matrix $R \in \mathbb{R}^{d\times m}$ whose entries are independently sampled from a zero mean normal distribution, and each row is normalized to unit length.

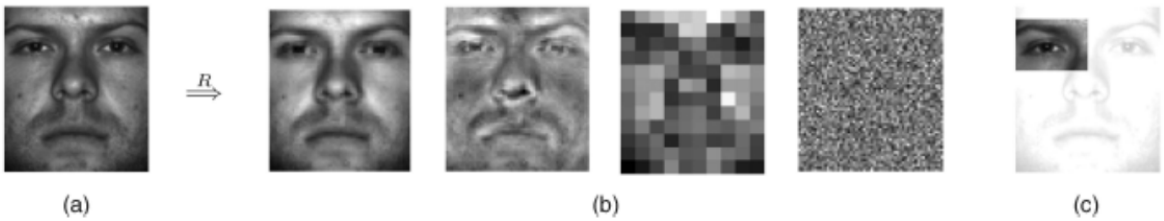


Fig. 5 Examples of feature extraction (a) Original face image (b) 120D representations in terms of four different features: Eigenfaces, Laplacianfaces, downsampled (12 x 10 pixel) image, and random projection (c) The eye is a popular choice of feature for face recognition. In this case, the feature matrix R is simply a binary mask.

B. Robustness to Occlusion or Corruption

In many practical face recognition scenarios, the test image y could be partially corrupted or occluded. The linear model can be modified as

$$y = y_0 + e_0 = Ax_0 + e_0$$

where $e_0 \in \mathbb{R}^m$ is a vector of errors – a fraction, ρ_r of its entries are nonzero. The nonzero entries of e_0 model which pixels in y are corrupted or occluded.

Let us assume that the corrupted pixels are a relatively small portion of the image. Then, the above equation can be written as

$$y = \begin{bmatrix} A & I \end{bmatrix} \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} \doteq Bw_0$$

where $B = \begin{bmatrix} A & I \end{bmatrix} \in \mathbb{R}^{m \times (n+m)}$, so the system $y = Bw$ is always underdetermined and does not have a unique solution for w. From the analysis of sparsity of x_0 and e_0, the correct generating w_0 has at most $n_i + \rho m$ nonzeros. We want to recover w_0 as the sparsest solution to the system $y = Bw_0$.

$$\left(l_e^1\right): \hat{w}_1 = \arg\min \|w\|_1 \text{ subject to } Bw = y$$

Once the sparse solution $\hat{w}_1 = \begin{bmatrix} \hat{x}_1 & \hat{e}_1 \end{bmatrix}$ is computed, setting $y_r \doteq y - \hat{e}_1$ recovers a clean image of the subject with occlusion or corruption compensated for. To identify the subject, we slightly modify the residual $r_i(y)$, computing it against the recovered image y_r

$$r_i(y) = \left\| y_r - A\delta_i\left(\hat{x}_1\right) \right\|_2 = \left\| y - \hat{e}_1 - A\delta_i\left(\hat{x}_1\right) \right\|_2$$

IV. Experimental Verification

A. *Feature Extraction and Classification Methods*

SRC algorithm using several conventional holistic face features, namely, Eigenfaces, Laplacianfaces, and Fisherfaces, and compare their performance with two unconventional features: randomfaces and downsampled images. They compare the SRC algorithm with three classical algorithms, namely, NN, and NS, discussed in the previous section, as well as linear SVM (support vector machine).

Solving $\left(l_r^1\right): \hat{x}_1 = \arg\min \|x\|_1$ subject to $\left\| RAx - \tilde{y} \right\|_2 \le \varepsilon$ with the error tolerance $\varepsilon = 0.05$.

1) Extended Yale B Database

The Extended Yale B database consists of 2,414 frontal-face images of 38 individuals. The copped and normalized 192 x 168 face images were captured under various laboratory-controlled lighting conditions. For each subject, they randomly select half of the images for training (about

32 images per subject) and the other half for testing. They compute the recognition rates with the feature space dimensions 30, 56, 120, and 504, whose numbers corresponding to downsampling ratios of 1/32, 1/24, 1/16, and 1/8, respectively.

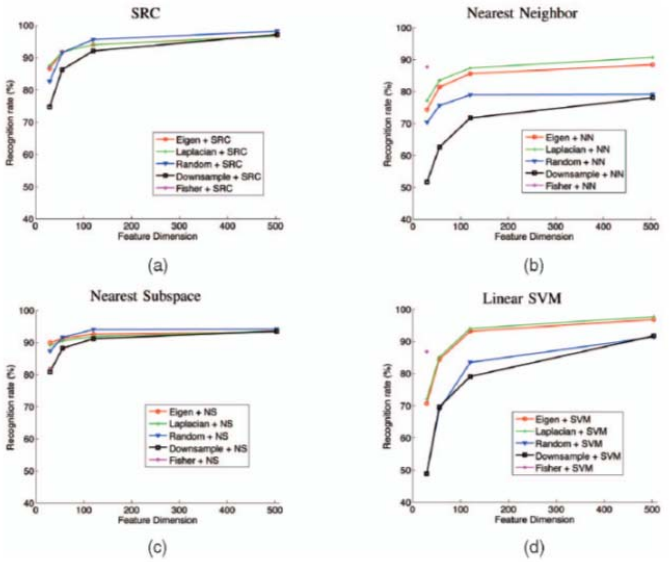


Fig. 6 Recognition rates on Extended Yale B database, for various feature transformation and classifiers. (a) SRC (b) NN (c) NS (d) SVM

2) AR Database

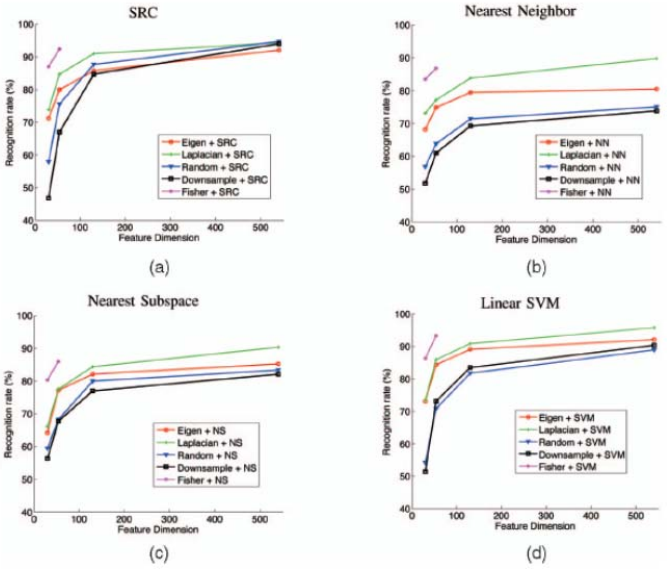


Fig. 7 Recognition rates on AR database, for various feature transformation and classifiers. (a) SRC (b) NN (c) NS (d) SVM

B. Partial Face Features

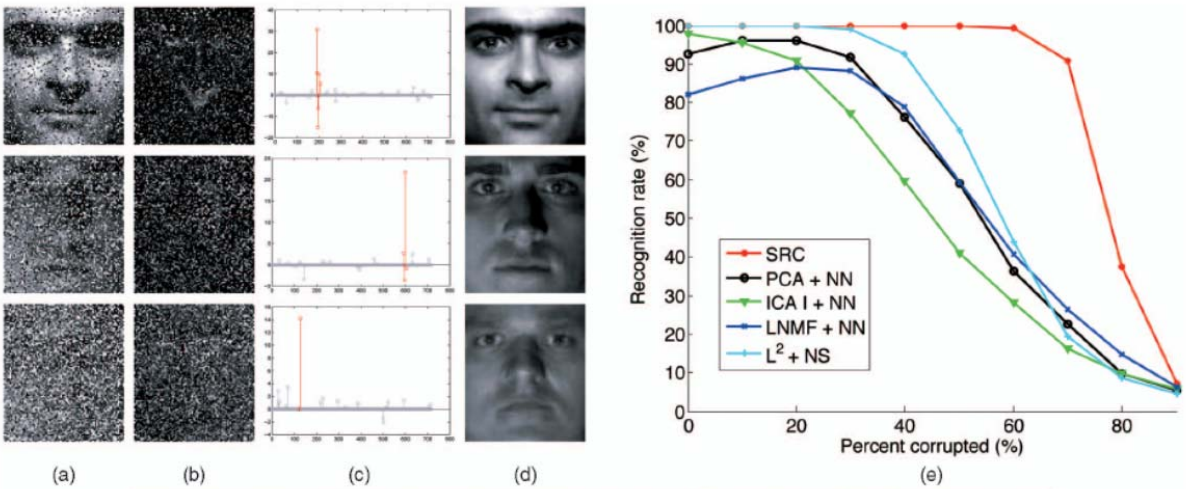
(a)

(b)

Features	Nose	Right Eye	Mouth & Chin
Dimension (d)	4,270	5,040	12,936
SRC	**87.3%**	**93.7%**	**98.3%**
NN	49.2%	68.8%	72.7%
NS	83.7%	78.6%	94.4%
SVM	70.8%	85.8%	95.3%

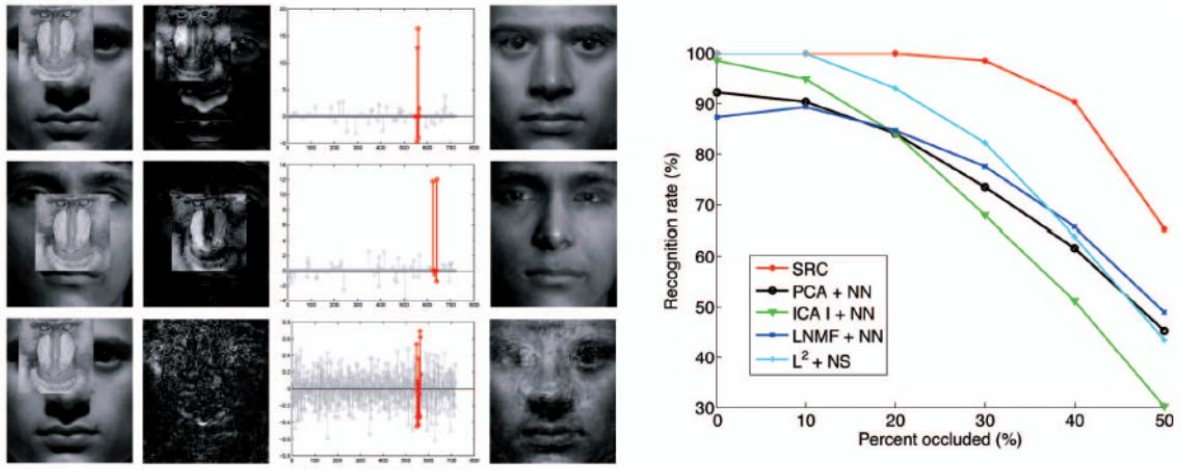
Fig. 8 Recognition with partial face features (a) example features, (b) recognition rates of SRC, NN, NS, and SVM on the Extended Yale B database.

C. Recognition Despite Random Pixel Corruption



Percent corrupted	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Recognition rate	100%	100%	100%	100%	100%	100%	99.3%	90.7%	37.5%	7.1%

D. Recognition Despite Random Block Occlusion



Percent occluded	0%	10%	20%	30%	40%	50%
Recognition rate	100%	100%	99.8%	98.5%	90.3%	65.3%

V. CONCLUSIONS

In this paper, exploiting sparsity can be used for the high-performance classification of high-dimensional data such as face images. The number of features is more important than the choice of features. With occluded and corrupted images, their classification algorithm shows still high-performance.

VI. DISCUSSION

※ Sparse error expression (Occlusion or corruption)

- Can we utilize the sparse error expression of this paper to hologram, microscopy, or spectrometer experiments?

※ BCI experiments

- need to increase classes (multi-class classification)

- applying Gaussian random matrix