International Journal of Fuzzy Logic and Intelligent Systems, vol. 5, no. 2, June 2005 pp. 131-135

Face Recognition by Using FP-ICA Based on Secant Method

Yong-Hyun Cho

School of Computer and Information Comm. Eng., Catholic Univ. of Daegu

Abstract

This paper proposes an efficient face recognition using independent component analysis(ICA) derived from the fixed point(FP) algorithm based on secant method. The secant method can exclude the complex computation of differential process from the FP based on Newton method. The proposed ICA has been applied to recognize the 20 Yale face images of 324x324 pixels. The experimental results show that the proposed ICA is superior to PCA not only in the restoration performance of basis images but also in the recognition performance of the trained images and the test images. Then negative angle as similarity measures has better recognition ratio than city-block and Euclidean.

Key words : Independent Component Analysis, Fixed Point Algorithm, Secant Method, Face Recognition

1. Introduction

A central problem in neural-network research, as well as in statistics and signal processing, are finding a suitable representation or transformation of the data. For computational and conceptual simplicity, the representation is often sought as a linear transformation of the original data. Many methods and principles can be accomplished for suitable linear transformation[1-5]. The principal component analysis (PCA) can be widely used method. PCA is defined by the eigenvectors of the covariance matrix of the input data. Such a representation is adequate for gaussian data[1-3]. However, the input data can contain a nongaussian data for example in various applications of communications, signal and image processing.

Recently, independent component analysis (ICA) has been proposed as an alternative method for PCA[5-11]. It is a signal processing technique whose goal is to express a set of random variables as linear combinations of statistically independent variables. The main applications are in blind source separation (BSS), and feature extractions. Basically PCA considers the second order moments only and it uncorrelates data, while ICA accounts for higher order statistics and it identifies the independent source components from their linear mixtures. ICA thus provides a more powerful data representation than PCA.

The measures of nongaussianity, such as fourth order cumulant(or kurtosis) and negentropy, has been used for estimating the ICA model[5,7]. The minimization of mutual information and the maximum likelihood estimation are also used for ICA estimation. But these numerical methods for ICA need the complex computations for measuring the statistical independence of data, which are probability density function, inverse matrix or higher order statistics.

The neural estimation for ICA has been considered as an alternative of numerical ICA[6-10]. This has been realized using the two layersfeedforward network. Then the prewhitened data has been used for fast convergence and better stability in learning. The recursive least square algorithm, Bell& Sejnowski(BS) algorithm, natural gradient algorithm, and fixed point algorithm have been proposed for neural learning[5-11]]. The gradient-based algorithm has that the convergence is slow, and depends on a good choice of learning rate. A bad choice of learning rate destroys convergence. The fixed point(FP) algorithm can solve this. It has very appealing convergence properties, making them a very interesting alternative to adaptive learning rules.

This paper proposes an efficient ICA of FP(FP-ICA) algorithm based on secant method. The secant method is an alternative of Newton method that is essential to differentiate the function for estimating the root[12]. The performance evaluation of the proposed algorithm has been applied to recognize the 20 facial images of 10 individuals of 324x324 pixels which come from Yale database[13]. The face recognition is compared with PCA. We demonstrate the compression performance due to different numbers of the basis images and the classification by means of distance measures such as city-block, Euclidean, and negative angle.

2. Independent Component Analysis

ICA relies on the independence, which is not as such a computable measure; it is a multi-dimensional extension of PCA, based on high-order statistics.

Denote by $\mathbf{x} = [x(1), x(2),\cdots, x(n)]T$ the n-dimensional data vector made up of the noisy linear mixtures. We can write the ICA model in vector form as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\eta} = \sum_{i=1}^{m} s(i)\,a(i) + \boldsymbol{\eta} \qquad (1)$$

Where, $\mathbf{s} = [s(1), s(2), \cdots, s(m)]^T$ is the source vector or

Manuscript received May. 18, 2005; revised Jun. 4, 2005.

independent component vector consisting of the m source signals. **A**=[a(1), a(2), ⋯, a(m)] is a n x m mixing matrix whose column a(i) are the basis vector of ICA. η denotes possible corrupting additive noise which is often omitted. All we observe is the random vector **x**, and we must estimate both **A** and **s** using it. The important assumption is that the independent component must have nongaussian distributions. Then ICA is simply to obtain the sources by using inverse matrix **W** of **A** ($\mathbf{W}=\mathbf{A}^{-1}$). Fig.1 shows the detailed relation for ICA estimation. From Fig.1, the output signal **y** = **Wx** = **WAs**, finally the united condition of **y** = **s**, that is, satisfying **WA** = **I**, we can estimate the inverse mixing matrix W.

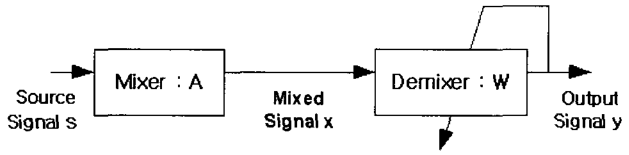

Fig. 1. Detailed relation for ICA estimation

Before applying an ICA algorithm on the data, it is usually very useful to do some preprocessing[5,7]. The most basic and necessary preprocessing is to make **x** a zero mean variables. It is to normalize the data with respect to the first order statistics. This preprocessing is made solely to simplify the ICA algorithm. This can simply be obtained from subtracting the observed vector **x** from the mean **x***. New data vector **x** which got by zero-mean is defined as follow:

$$\mathbf{x} = \mathbf{x} - \mathbf{x}^* \tag{2}$$

Furthermore the effects of second order statistics to the nonlinearities can be removed by whitening the data. We transform the observed vector x linearly so that we obtain a new vector which is whitened, i.e., its components are uncorrelated and their variances equal unity matrix, in other words, E{xxT} = I. Standard PCA is often used for whitening, because one can then simultaneously compress information optimally in the mean square error sense and filter possible gaussian noise. The PCA whitening matrix is given by as follow:

$$\mathbf{V} = \mathbf{D}^{(-1/2)}\mathbf{E}^T \tag{3}$$

Where the diagonal matrix **D** is eigenvalue of the data covariance matrix, and matrix **E** is composed of principal eigenvector. In this paper, the mixing images are generally preprocessed by zero-mean and whitening for using as the data.

3. Fixed Point Algorithm Based on Secant Method

Fixed point algorithm for performing ICA derives from the entropy optimization methods and its speed is about second order function[7-9]. The algorithm converges faster than any other known algorithm. FP algorithm is based on the mutual information, this measure is a kind of distance of independence, and minimizing mutual information leads to ICA solution.

Generally, the mutual information that is a natural measure of the dependence between random variable, can be defined by negentropy. The definition of negentropy J(**y**) for n random vector $\mathbf{y} = (y_1, \cdots, y_n)^T$ is given by as follow:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{4}$$

where ygauss is gaussian random variables of the same covariance matrix as y. Negentropy can also be interpreted as a measure of nongaussianity. It is particularly interesting to express mutual information using negentropy, constrained the variable to be uncorrelated. In this case, we have

$$I(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n) = J(\mathbf{y}) - \sum_{i=1}^{n} J(y_i) \tag{5}$$

Finally, maximization of independence between random variables is minimization of mutual information $I(y_1, y_2, \cdots, y_n)$ defined eq. (5). This can be done by maximization of negentropy in eq. (4). Therefore maximization of negentropy can afford to maximize second term of eq. (5). Using the maximum entropy principle is that,

$$J(\mathbf{y}_i) \cong c[E\{G(\mathbf{y}_i)\} - E\{G(v)\}]^2 \tag{6}$$

where G(.) is any non-quadratic function, c is a constant, and v is a gaussian variable of zero mean and unit variance. The random variable yi is assumed to be of zero mean and unit variance.

The approximation of negentropy given above in eq. (6) gives readily a new objective function for estimating the ICA transform. To find one independent component as $y_i = \mathbf{W}^T\mathbf{x}$, we maximize the function JG(**W**) given by

$$JG(\mathbf{W}) = [E\{G(\mathbf{W}^T\mathbf{x})\} - E\{G(v)\}]^2 \tag{7}$$

Next, using the approach of minimizing mutual information, the above one unit contrast function can be simply extended to compute the whole matrix **W**. To do this, recall from eq. (5) that mutual information is minimized when the sum of the negentropies of the components is maximized. Maximizing the sum of n one unit contrast function, and taking into account the constraint of decorrelation, one obtains the following optimization problem.

$$\text{maximize } \sum_{i=1}^{n} J_G(w_i) \text{ w.r.t. } \mathbf{w}_i,\ i=1,2,\cdots,n$$
$$\text{subject to } E\{(\mathbf{w}_k^T\mathbf{x})(\mathbf{w}_j^T\mathbf{x})\} = \delta_{jk} \tag{8}$$

Where at the maximum, every vector $\mathbf{w}_i$ gives one of the rows of the matrix **W**, and the ICA transform is then given by **s** = **Wx**. Thus we need algorithm for solving the optimization problems in practice. Among this algorithm, the simplest method is based on gradient descent. But this algorithm restricts the convergent speed and doesn't work sometimes. So, FP algorithm has been proposed as alternative[6-9]

If the data vector **x** has been whitened by the correlation matrix $E\{\mathbf{x}\mathbf{x}^T\}=\mathbf{I}$, we shall derive the fixed point algorithm for one unit in eq, (7). The maxima of JG(**W**) are obtained at certain optima of $E\{G(\mathbf{W}^T\mathbf{x})\}$. According to the Kuhn-Tucker conditions, the optima of $E\{G(\mathbf{W}^T\mathbf{x})\}$ under constraint $E\{(\mathbf{W}^T\mathbf{x})2\} = ||\mathbf{W}||^2 = 1$ are obtained at points where

$$E\{\mathbf{x}g(\mathbf{W}^T\mathbf{x})\}\ \beta\mathbf{W} = 0 \qquad (9)$$

Where $g(\cdot)$ is derivative of the non quadratic function G, β is a constant that given by $\beta = E\{\mathbf{W_0}^T\mathbf{x}g(\mathbf{W_0}^T\mathbf{x})\}$, where $\mathbf{W_0}$ is the value of $\mathbf{W}$ at the optimum. This equation can be solved by Newton method[7]. Denoting the function on the left hand side of eq. (9) by F(**W**), we obtain its Jacobian matrix JF(**W**) as

$$JF(\mathbf{W}) = E\{\mathbf{x}\mathbf{x}^T g'(\mathbf{W}^T\mathbf{x})\} - \beta I \qquad (10)$$

Since the data is whitened, to simplify the inversion of this matrix, we decide to approximate the first term in eq. (10). A reasonable approximation seems to be

$$E\{\mathbf{x}\mathbf{x}^T g'(\mathbf{W}^T\mathbf{x})\} \cong E\{\mathbf{x}\mathbf{x}^T\}E\{g'(\mathbf{W}^T\mathbf{x})\} = E\{g'(\mathbf{W}^T\mathbf{x})\}I \qquad (11)$$

Thus the Jacobian matrix becomes diagonal, and can easily be inverted. We also approximate β using the current value of $\mathbf{W}$ instead of $\mathbf{W_0}$.

The secant method is the numerical method that the derivative of function is approximated by using the two consecutive iterative values of function[12]. This eliminates the need to evaluate the derivative function in Newton method. Therefore, the secant method is more efficient, particularly when function is a time-consuming function to evaluate.

We propose a new FP algorithm based on secant method that can be approximately computed by using only the two consecutive values of objective function. Thus we obtain the following approximate derivative function f'(**W**)

$$f'(\mathbf{W}_k) \cong [f(\mathbf{W}_k) - f(\mathbf{W}_{k-1})] / [\mathbf{W}_k - \mathbf{W}_{k-1}] \qquad (12)$$

Where, k, and k-1 is present and prior iteration number, respectively. Using eq. (12), we obtain the updating iteration of inverse mixing matrix W based on the secant iteration

$$\mathbf{W}_{k+1} = \mathbf{W}_k - f(\mathbf{W}_k)(\mathbf{W}_k - \mathbf{W}_{k-1})/[f(\mathbf{W}_k)f(\mathbf{W}_{k-1})] \qquad (13)$$

Eq. (13) needs $\mathbf{W}_k$ and $\mathbf{W}_{k-1}$ as initial value for updating W, but not another derivative function. We can obtain the inverse mixing matrix **W**(solution of eq. (9)) by using eq. (13). Finally we obtain the learning process for inverse mixing matrix **W** as follows:

$$f(\mathbf{W}^\#) = [E\{\mathbf{x}g(\mathbf{W}^{\#T}\mathbf{x})\} - \beta\mathbf{W}^\#]$$
$$f(\mathbf{W}) = [E\{\mathbf{x}g(\mathbf{W}^T\mathbf{x})\} - \beta\mathbf{W}]$$
$$\mathbf{W}+ = \mathbf{W} - f(\mathbf{W})[\{\mathbf{W} - \mathbf{W}^\#\} / \{f(\mathbf{W}) - f(\mathbf{W}^\#)\}]$$
$$\mathbf{W}* = \mathbf{W}+ / ||\mathbf{W}+|| \qquad (14)$$

where $\mathbf{W}^\#$ denotes the previously calculated value of **W**, **W*** denotes the updated value of **W**, $\beta = E\{\mathbf{W}^T\mathbf{x}g(\mathbf{W}^T\mathbf{x})\}$. Eq. (14) is the proposed fixed point algorithm based on secant method for ICA.

The updating process for inverse mixing matrix **W** can be summarized as follow:

Step 1 : $\mathbf{W}(0) = \text{rand}(\cdot)$.
Step 2 : $\mathbf{W} = \mathbf{W}(0) / ||\mathbf{W}(0)||$
Step 3 : $\mathbf{W}^\# = 0$
Step 4 : While $||\mathbf{W} - \mathbf{W}^\#|| > \varepsilon \cap ||\mathbf{W} + \mathbf{W}^\#|| > \varepsilon$
  (1) β, $f(\mathbf{W}^\#)$, and $f(\mathbf{W})$
  (2) $\mathbf{W}+ = \mathbf{W} - f(\mathbf{W})[\{\mathbf{W} - \mathbf{W}^\#\} / \{f(\mathbf{W}) - f(\mathbf{W}^\#)\}]$
  (3) $\mathbf{W}^\# = \mathbf{W}$
  (4) $\mathbf{W} = \mathbf{W}+ / ||\mathbf{W}+||$
  end

where vector **x** is whitened, ε is a deviation of **W** and smaller than 1, $g(\cdot)$ is nonlinear function and is generally used function of $(\cdot)^3$ and function of $\tanh(\cdot)$[7-10]. The $(\cdot)^3$ function has been used in this paper.

4. Simulations and experimental results

The performance evaluation of the proposed algorithm has been performed to recognize the 20 facial images (2images * 10 persons) of 324x324 pixels, which comes from Yale database[13].

To evaluate the performance of proposed algorithm, we define an absolute mean sum error (AMSE) for comparing the accuracy of matching between original images and restored images. Then AMSE = $\frac{1}{N}\sum_{i=1}^{N}|x_i - y_i|$, where x_i and y_i represent the i-th pixel values for original and restored image, respectively, and N is the total numbers of pixel. The experiments were performed by Matlab 5.1 under Pentium IV 1.5G.

Fig.2 shows normal facial images for learning ICA used in experiments. Upper left shows 10 original facial images of 10 individuals. Upper right shows the mean image of those, and lower also shows the zero mean images that are subtracted mean image from original image.

Fig. 3 shows the 10 face images (lower) restored from 6 basis images (upper). This represents the compression efficiency by reducing the dimension, which are the results of extracting the 6 global features from 10 faces. But with the naked eye, the restored images including the noise, are different from the original images more or less. It is due to decrease of the numbers of basis vector.

On the one hand, Table 1 shows AMSE of PCA and ICA for evaluating the performance of restoration according to the number of basis images. Here, we can see that AMSE is gradually decreased in proportion to the number of basis images, and since then increased in reverse. It means that the restoration performance is decreased by over-learning if the number of feature vectors, which are the basis images, is increased excessively. Also, Table 1 shows the fact that AMSE of ICA is superior to that of PCA about 105 times. We made a recognition experiment using 9 basis images, which gives the best result in restoration performance.

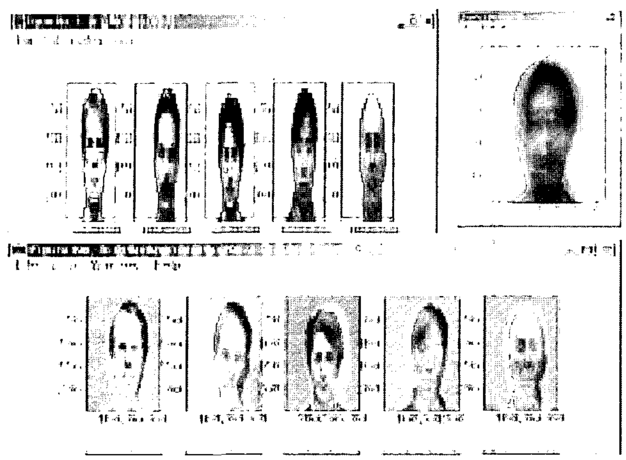

Fig. 2. Face images for learning. (upper left) 10 original facial images corresponding to 10 persons, (upper right) mean image of those, (lower) zero-mean images that is subtracted the mean image from the original image.

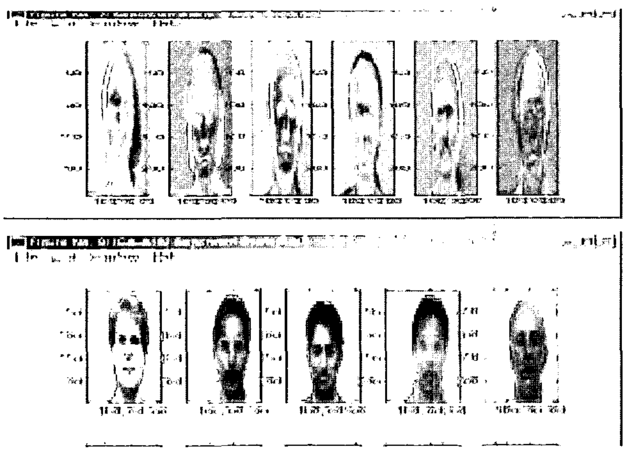

Fig. 3. (upper) 6 basis images, (lower) 10 restored face images.

Table 1. AMSE of PCA and ICA according to the number of basis images

The number of basis images		1	2	3	4	5	6	7	8	9	10
AMSE	PCA	21.74	19.16	16.15	14.11	11.41	8.48	6.04	3.24	1.36e-012	6.22
	ICA	4.64e-005	4.25e-005	3.50e-005	3.40e-005	3.23e-005	2.65e-005	2.49e-005	2.95e-006	3.45e-018	2.17e-006

Fig. 4 is the 10 face images for tests, which include a various facial expression and lighting effects - surprise, sadness, happy, nightwear, wink, glasses, center lighting, right lighting, left lighting etc.. Upper is the original face images for tests, and lower is those zero-mean face images. The test images are different from learning images in facial expression and lighting etc., and the eigenvectors of those were driven by using zero-mean images.

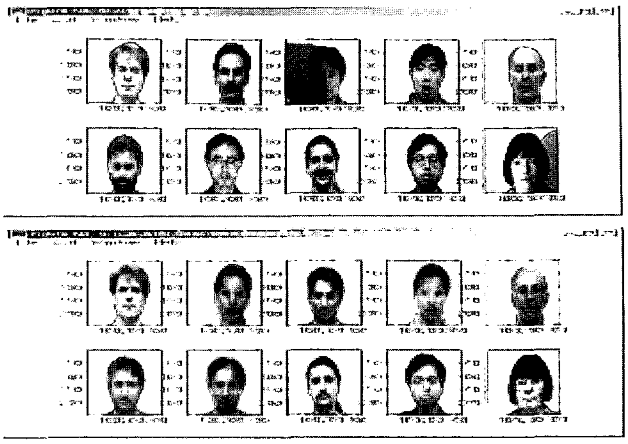

Fig. 4. Face images for testing. (upper) 10 original images, (lower) 10 zero-mean images.

Table 2 shows the recognition results of the 10 learning and test faces by using PCA and proposed ICA, respectively. Here, we used 3 distances to measure the similarity between the trained image and the test images, such as city block (L1-norm), Euclidean (L2-norm), and negative angle (cosine). Both PCA and the proposed ICA had a ratio of recognition 100% and 80% for the trained and test images, respectively. Using city block and Euclidean, the learning face images had error values of less than about 10-9 by PCA and 10-14 by the proposed ICA, respectively. The other hands, the 10 test images had an error value of less than about 105 by PCA and 102 by the proposed ICA, respectively. Thus, we can see that the proposed ICA has about 103 times to the trained images and 105 to the test images better performance than PCA. Using the negative angle, neither PCA nor ICA had an error value for the 10 trained images, but the 10 test images had the error value of about 1.2352 by PCA and 1.1746 by ICA, respectively. Then, the proposed ICA took a similar performance with PCA. The negative angle as the similarity measure has better recognition ratio than the city block and Euclidean measures. However, No. 3 test image with glasses on is misidentified as No. 9 image with lighting, and No. 6 as No. 3.

Table. 2. Recognition results of the 10 trained face images and the test face images.

Distance Measure	Input face images	Recognition rate(%)		Error Sum		Incorrect faces	
		PCA	ICA	PCA	ICA	PCA	ICA
City Block (L1-norm)	Trained	100	100	6.16e-009	4.34e-014	-	-
	Test	80	80	1.23e+005	1.76e+002	3(6), 9(3)	3(6), 9(3)
Euclidean (L2-norm)	Trained	100	100	3.61e-009	2.81e-014	-	-
	Test	80	80	5.17e+004	7.99e+001	3(6), 9(3)	3(6), 9(3)
Negative Angle (L3-norm)	Trained	100	100	0.0	0.0	-	-
	Test	80	80	1.24	1.17	3(6), 9(3)	3(6), 9(3)

* i(j) : (i) the number of test face to be correctly identified, and (j) the number of incorrectly identified test face.

5. Conclusions

This paper proposed an efficient ICA of FP algorithm based on secant method, which can be approximately computed by only the values of function for estimating the root of objective function for optimizing entropy. The secant method is an alternative of Newton method that is essential to differentiate the function for estimating the root. The performance evaluation of the proposed algorithm has been performed to recognize the 20 facial images of 10 individuals of 324x324 pixels.

In the experimental results, both PCA and the proposed ICA had the recognition ratio of 100% for the trained images and 80% for the test images. But, the proposed ICA had the better performance than PCA not only in the restoration by basis images but also in the recognition ratio of the trained images and the test images. Especially, the restoration performance of the proposed ICA was superior to that of PCA about 105 times. Then negative angle as similarity measures for face recognition had better recognition performance than city block and Euclidean.

References

[1] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications, Adaptive and learning Systems for Signal Processing, Communications, and Control*, John Wiley & Sons, Inc., 1996

[2] S. Haykin, Neural Networks: *A Comprehensive Foundation*, Prentice Hall, 2ed, London, 1999

[3] J. Karhunen and J. Joutsensalo,"Generation of Principal Component Analysis, Optimization Problems, and Neural Networks," *Neural Networks*, Vol. 8, No. 4, pp. 549 562, 1995

[4] P. Comon, "Independent Component Analysis A New Concept?," *Signal Processing*, vol.36, No.3, pp.287 314, Apr. 1994

[5] T. W. Lee, *Independent Component Analysis: Theory and Applications*, Kluwer Academic Pub., Boston, 1998

[6] J. Karhunen, "Neural Approaches to Independent Component Analysis and Source Separation," *4th European Symp., Artificial Neural Network*, ESANN96, Burges, Belgium, pp. 249 266, Apr. 1996

[7] A. Hyvaerinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., New York, 2001

[8] A. Hyvaerinen and E. Oja, "A Fast Fixed Point Algorithms for Independent Component Analysis", *Neural Computation*, 9(7), pp. 1483 1492, Oct.1997

[9] A. Hyvaerinen, "Fast & Robust Fixed Point Algorithms for Independent Component Analysis", *IEEE Trans. on Neural Networks*, Vol. 10, No. 3, pp.626 634, May 1997

[10] A. Hyvaerinen and E. Oja, "Independent Component Analysis: Algorithms and Applications", *Neural Networks*, Vol. 13, No. 4 5, pp. 411 430, June 2000

[11] A. Cichocki and R. Unbehauen, "Robust Neural Networks with On Line Learning for Blind Identification and Blind Separation of Sources", *IEEE Trans. on Circuits & Systems*, Vol. 43, No. 11, pp. 894-906, Nov. 1996

[12] K. Atkinson, *Elementary Numerical Analysis*, John Wiley & Sons, Inc., New York, 1993

[13] "Yale Face Databases," *http://cvc.yale.edu/projects/ yalefaces/ yalefaces.html*

Yong-Hyun Cho
received the B.S, M.S. and Ph.D. degree in electronics from Kyungpuk National Univ. Korea, in 1979, 1981, and 1993, respectively. From 1983 to 1987, he worked in exchange system lab. of ETRI. Since 1987, he has been a Professor in the School of Computer and Information Comm. Eng., of Catholic Univ. of Daegu. His research interests include neural networks, image and biological signal processing, and pattern recognition, etc..

Phone : +82-53-850-2747
Fax : +82-53-850-2740
E-mail : yhcho@cu.ac.kr