퍼지 및 지능시스템학회 논문지 2004, Vol. 14, No. 5, pp. 623-629

A Efficient Image Separation Scheme Using ICA with New Fast EM algorithm

Bum-Jin Oh*, **Sung-Soo Kim****, **Jee-Hye Kang****

* Kyeryong Technical High School Sansungdong, Donggu, Daejeon, Korea
** School of Electrical and Computer Engineering Chungbuk National University, Cheongju Chungbuk, Republic of Korea

Abstract

In this paper, a Efficient method for the mixed image separation is presented using independent component analysis and the new fast expectation-maximization(EM) algorithm. In general, the independent component analysis (ICA) is one of the widely used statistical signal processing scheme in various applications. However, it has been known that ICA does not establish good performance in source separation by itself. So, Innovation process which is one of the methods that were employed in image separation using ICA, which produces improved the mixed image separation. Unfortunately, the innovation process needs long processing time compared with ICA or EM. Thus, in order to overcome this limitation, we proposed new method which combined ICA with the New fast EM algorithm instead of using the innovation process. Proposed method improves the performance and reduces the total processing time for the Image separation. We compared our proposed method with ICA combined with innovation process. The experimental results show the effectiveness of the proposed method by applying it to image separation problems.

Key Words : Independent Component Analysis, Innovation Process, Expectation-Maximization.

1. Introduction

In this paper, a mixed image separation scheme[1,2] is presented using a new method that combines the properties of the independent component analysis (ICA), and the new fast expectation-maximization (EM). ICA is one of the widely used statistical signal processing tools, which represents the information from observations as a set of random variables in the form of linear combinations of another statistically independent component variables. Technically, ICA is regarded as a useful extension of the principal component analysis (PCA), which was developed to resolve the problems of correlation in the blind separation[3,4] of independent sources from their linear mixtures. Such blind techniques[5] are needed for example in various applications of array processing, communications, medical signal processing, and speech processing. In the similar manner, there are also several applications of ICA[6,7] such as blind source separation, feature extraction, and blind deconvolution, and etc..

In a sense, the fundamental motivation of ICA is to remove the correlatedness property of standard PCA. Roughly, rather than requiring that the coefficients of a linear expansion of the data vectors be uncorrelated, in ICA they transformed to be mutually statistically independent as much as possible, where higher order statistics are required in determining the ICA expansion. It means that some non-linearities must be used in the process of constructing a system. In various useful applications, ICA provides a more meaningful representation of the data than PCA through the transformation of the data to be quasi-orthogonal to each other, which can be utilized in linear projection.

For recent decades, several unsupervised separation algorithms were derived from ICA assuming that the observed data can be transformed into several mutually exclusive data classes whose components are generated by linear mixtures of independent non-Gaussian sources. However, it has been known that ICA does not establish good performance in source separation by itself. In order to overcome this limitation, there have been many techniques that are designed to reinforce the good properties of ICA. Innovation process is one of the methods that were employed in image separation using ICA. Even the innovation process improves the performance of mixed image separation, there are still much work to be done, such as the long processing time compared with ICA or EM needs to be reduced. In this paper, the new fast expectation and maximization method[13] is added to the system with ICA instead of the innovation process. The proposed method yields the improved results that are presented in experiments.

접수일자 : 2004년 4월 19일
완료일자 : 2004년 6월 18일
이 논문의 제 2저자는 2002년도 충북대학교 신진교수연구 지원에 의하여 연구하였음

2. Preliminaries

2.1 Independent Component Analysis

Independent component analysis(ICA) is a statistical technique that transforms the relations between information to be independent to each other. Let m scalar random variables $x_1, x_2, \cdots, x_m$ be linear combination of n unknown mutually statistically independent components with zero mean, denoted by $s_1, s_2, \cdots, s_n$. Then we can put the observed variables x_i into a vector $\mathbf{x} = (x_1, x_2, \cdots, x_m)^T$ and the component variables s_j into a vector $\mathbf{s}$, such that the linear relationship can be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{A}$ is an unknown $m \times n$ matrix of full column rank, called the mixing matrix. The goal of ICA is finding both the mixing matrix $\mathbf{A}$ and the independent components s_j from the observations of the mixtures x_j.

In ICA, the assumption of independence of variables is crucial for the identifiability of the model. However, if we consider the observed model as the time-dependent stochastic processes instead of random variables, the process becomes innovation. The innovation process of a stochastic process is a more general condition than the independence of the components themselves. The mixing matrix $\mathbf{A}$ is estimated by applying ICA on the innovation process. However, unfortunately the estimation done by innovation process is not good enough to separate original images from a set of mixed images. In this paper, we present a new scheme that combines the innovation process and the expectation-maximization to reinforce the weakness of the combined model of ICA and stochastic innovation process.

2.2 ICA Estimation using Innovations [8]

Given a stochastic process $\mathbf{s}(t)$, we define its innovation process $\tilde{\mathbf{s}}(t)$ as the error of the best prediction (i.e. conditional expectation) of $\mathbf{s}(t)$, given its past:

$$\tilde{\mathbf{s}}(t) = \mathbf{s}(t) - E[\, \mathbf{s}(t)|t\ ,\ \mathbf{s}(t-1),\ \mathbf{s}(t-2), \cdots] \tag{2}$$

The expression $\tilde{\mathbf{s}}(t)$ represents the information about the information about the process that can be obtained at time t where he innovation process is uniquely defined by (2). Estimation of the innovation process can be performed by approximating the conditional expectation. i.e. the best prediction of $\mathbf{s}(t)$ given its past (in the least mean-square sense). This is basically a regression problem that can be approximated in many cases by ordinary linear autoregressive models; in the very simplest case, a reasonable approximation of the innovation process may be given by the difference process $\Delta \mathbf{s}(t) = \mathbf{s}(t) - \mathbf{s}(t-1)$. In general , the nonlinear prediction may be approximated, e.g. by multi-layer perceptrons or radial basis function.

Let us consider how the concept of innovation process can be used in the framework of estimation of the ICA data model. Consider a version of the ICA data model in (1) where the observed data is a stochastic process $x(t)$ that is represented as a linear combination of component process:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{3}$$

the concept of innovation can be utilized in the estimation of the data model (3) due to the following. If $\mathbf{x}(t)$ and $\mathbf{s}(t)$ follow the instantaneous mixing model (3), then the innovation processes becomes

$$\tilde{\mathbf{x}}(t) = \mathbf{A}\,\tilde{\mathbf{s}}(t) \tag{4}$$

where

$$\mathbf{A}\,\tilde{\mathbf{s}}(t) = \mathbf{x}(t) - E[\, \mathbf{x}(t)|t\ ,\ \mathbf{s}(t-1),\ \mathbf{s}(t-2), \cdots] \tag{5}$$

since the information contained in $[\, \mathbf{s}(t-1), \mathbf{s}(t-2), \cdots]$ equals the information contained in $[\, \mathbf{x}(t-1), \mathbf{x}(t-2), \cdots]$ due to the invertibility of $\mathbf{A}$, this shows that $\mathbf{A}\,\tilde{\mathbf{s}}(t)$ is the innovation of $\mathbf{x}(t)$. This implies that it is good enough for the model (3) to be identifiable that the innovation process fulfill the identifiability conditions usually required of the random vector $\mathbf{s}(t)$. In particular , it is good enough that $\tilde{\mathbf{s}}(t)$ has independent components and is stationary as well as ergodic [9,10]. This is a generalization of the ordinary identifiability conditions since the independence of the $s_i(t)$ implies the independence of the innovation processes $\tilde{\mathbf{s}}(t)$.

Owing to facts that the independence of the original process implies the independence of the innovations, but not vice versa, and the innovations may correspond to physically independent processes, the innovations are usually more independent from each other than the original process. Thus the accuracy of the estimation of the ICA model increases with increasing independence and nongaussianity of the component s_i[11], such that the innovation process is likely to lead to much better estimates of the mixing matrix.

2.3 General Statement of the EM Algorithm [12]

Let Y denote the sample space of the observations, and let $\mathbf{y} \in \mathbf{R}^m$ denote an observation from Y. Let x denote the underlying space and let $\mathbf{x} \in \mathbf{R}^n$ be an outcome from, with $m < n$. The data x is not observed directly, but only by means of $\mathbf{y}$, where $\mathbf{y} = \mathbf{y}(\mathbf{x})$, and $\mathbf{y}(\mathbf{x})$ is a many-to-one mapping. An observation determines a subset of $\mathbf{x}$, which is denoted as $x(y)$ as

Fig. 1 illustrates the mapping.

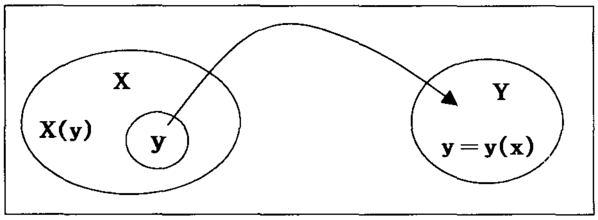


Fig.1 Illustration of many-to-one mapping

The probability density function (pdf) of the complete data is $f_X(\mathbf{x}|\theta)=f(\mathbf{x}|\theta)$, where $\theta\in\Theta\subset \boldsymbol{R}^r$ is the set of parameters of the density. Then pdf of the incomplete data is

$$g(\mathbf{y}|\theta)=\int_{x(y)} f(\mathbf{x}|\theta)d\,\mathbf{x} \quad (6)$$

Let

$$L_y(\theta)=\log \quad g(\mathbf{y}|\theta) \quad (7)$$

denote the log-likelihood function. The algorithm finds θ to maximize $\log f(\mathbf{y}|\theta)$, by maximizing the expectation of $\log f(\mathbf{y}|\theta)$ given the data $\mathbf{y}$ and estimate of θ. Let $\theta^{[K]}$ be our estimate of the parameters at the k-th iteration. Then we compute

$$Q(\theta|\theta^{[k]})=E[\log \quad f(\mathbf{x}|\theta)|\,\mathbf{y}\,,\theta^{[k]}] \quad (8)$$

Then, we compute $\theta^{[K+1]}$ where the value of θ maximizes $Q(\theta|\theta^{[k]})$:

$$\theta^{[k+1]}=\arg \max_\theta Q(\theta|\theta^{[k]}) \quad (9)$$

The EM algorithm consists of choosing an initial $\theta^{[K]}$, then performing the E-step (8) and the M-step (9) successively. The pdf is of the form

$$f(\mathbf{x}|\theta)=b(\mathbf{x})\exp[\,\mathbf{c}(\theta)^T\mathbf{t(x)}]\,/\,a(\theta) \quad (10)$$

where θ is a vector of parameters for the family[10,11]. The function $\mathbf{t(x)}$ is called the sufficient statistic of the family. Members of the exponential family include most distributions of engineering interest, including Gaussian, Poisson, binomial, uniform, Rayleigh, and other. For exponential families, the E-step can be written as

$$Q(\theta|\theta^{[k]})=E[\log \quad b(\mathbf{x})|\,\mathbf{y}\,,\theta^{[k]}] + \mathbf{c}(\theta)^T E[\,\mathbf{t(x)}|\,\mathbf{y},\,\theta^{[k]}]-\log a(\theta) \quad (11)$$

Let $t^{[k+1]}=E[t(x)|y,\theta^{[k]}]$. As a conditional expectation is an estimator, $t^{[k+1]}$ is an estimate of the sufficient statistic of the sufficient statistic. In light of the fact that M-step will be maximizing the term of the right side of equation (11) with respect to θ and that $E[\log \quad b(\mathbf{x})|\,\mathbf{y}\,,\mathbf{q}^{[k]}]$ does not depend upon $\boldsymbol{\theta}$, it is sufficient to write:

E-step Compute :

$$\mathbf{t}^{[k+1]}=E[\quad \mathbf{t(x)}|\,\mathbf{y},\theta^{[k]}]$$

M-step Compute :

$$\theta^{[k+1]}=\arg \max_\theta \mathbf{c}(\theta)^T\,\mathbf{t}^{[k+1]}-\log a(\theta)$$

Every iteration of the EM algorithm increases the likelihood function until a point of (local) maximum is reached.

3. Proposed Efficient Algorithm for Image Separation

As mentioned before, the method of the ICA estimating using the innovation process for the separation problems has some limitations. In order to overcome we proposed the new method which is combined with the new fast EM algorithm instead of the innovation process. The new fast EM algorithm has shorting convergence speed compared with conventional EM by reduction of the total iterations. There are two efficient methods to improve the EM algorithm. At first, we proposed the new K-menas algorithm using the uniform partition method. In general, conventional K-means used the random choice for the initialization. Which causes slow convergence speed and unwanted clustering result. Fig.2 is shown that proposed K-means procedure [13].

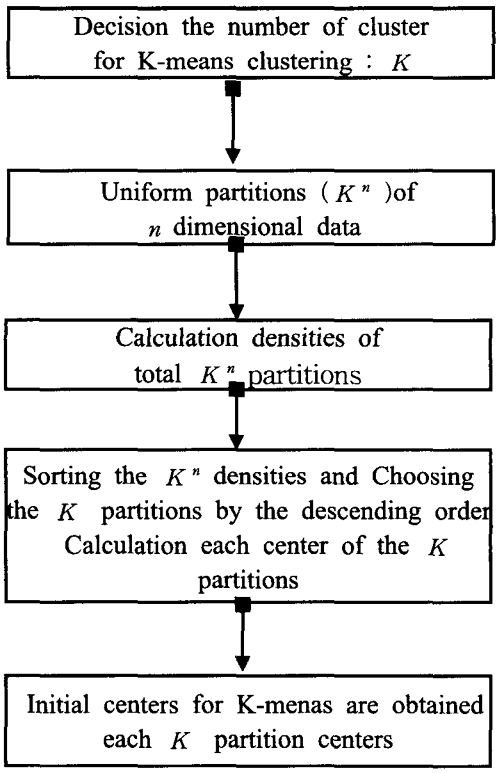


Fig.2 Proposed K-means Initialization using the Uniform partitions method

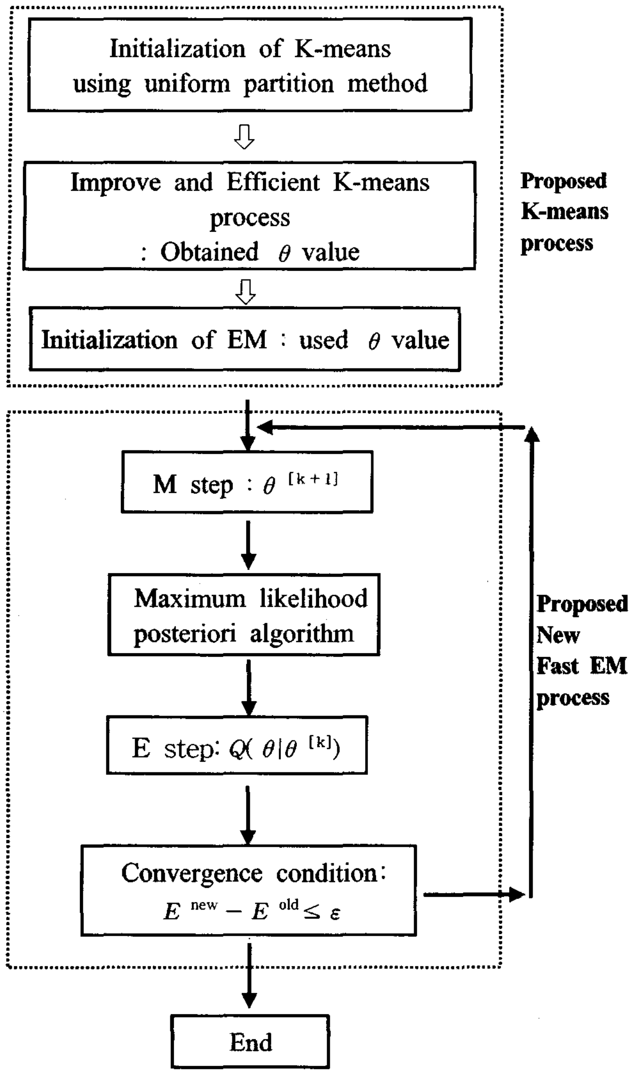


Fig.3 Proposed the EM initialization and the new fast EM procedure

This new K-menas method is applied to the EM initialization in order to obtain proper initial estimate. The secondly proposed the new EM method is maximum likelihood posteriori (MLP) which is mapped the maximum posteriori to unit value and the others are assigned zeros. It is reasonable that the posteriori in the M-step is obtained the conditional expectation of the parameter in the E-step by the good initial estimate. Thus this method reinforce the posteriori which is very important factor in the EM algorithm. Fig.3 is shown the new fast EM procedure. Proposed method improve the performance and reduce the convergence speed extremely.

We applied the new EM algorithm combined with ICA for the image separation. Instead of using the innovation process to improve the ICA performance[14], new EM algorithm is attached as the pre-process of the proposed ICA method.

4. Experimental Results

The results of a scheme that employs ICA using the new EM pre-process are better than ones by ICA with innovation. However, there is a weak point of this algorithm too, which ICA still is very much affected by the degree of correlation between mixed images whose correlation coefficients are presented in Table 1.

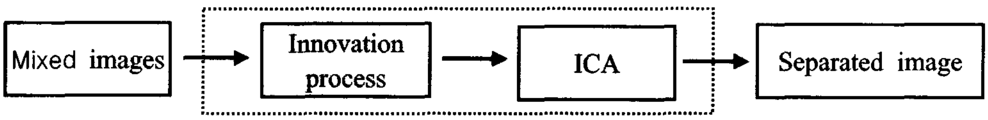
Table 1. Correlation between Original Images

Corr	Objection 1	Objection 2	Objection 3	Objection 4
Objection 1	1.0000	0.7542	0.6763	0.6495
Objection 2	0.7542	1.0000	0.7148	0.7167
Objection 3	0.6763	0.7148	1.0000	0.8463
Objection 4	0.6495	0.7167	0.8463	1.0000

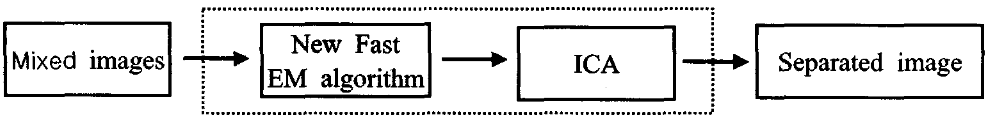
In Table 2. the weights from the new fast Expectation and Maximization(EM) are presented. The weights are used as a preprocessing of ICA.

Table 2. Weights from the new EM

Objection 1	0.6290	0.3815	0.3255	0.2971
Objection 2	0.1831	0.1071	0.0948	0.0936
Objection 3	0.3480	0.2099	0.1796	0.1747
Objection 4	0.4625	0.2780	0.2410	0.2216



(a) ICA estimation with Innovation process



(b) Proposed new pre-process of the ICA estimation

Fig.4 Comparison the conventional ICA and the proposed method

In order to strengthen the uncorrelatedness of the mixed images, a new scheme for image separation using ICA and EM without innovation process will be presented.

At first we presented the superiority of the new fast EM with respect to the convergence speed. Fig. 5 is shown that the average of the total number of iterations between the conventional EM and the proposed new fast EM [13]. The proposed EM algorithm yields fast convergence where the average of the total number of iterations is only six as shown in Fig.5 while the conventional's one converges after longer iterations. Thus, we applied the new fast EM algorithm to ICA.

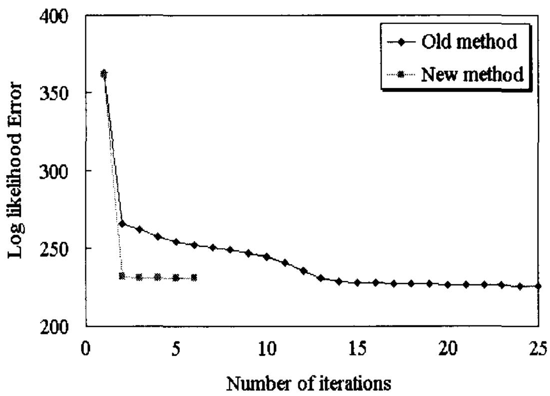


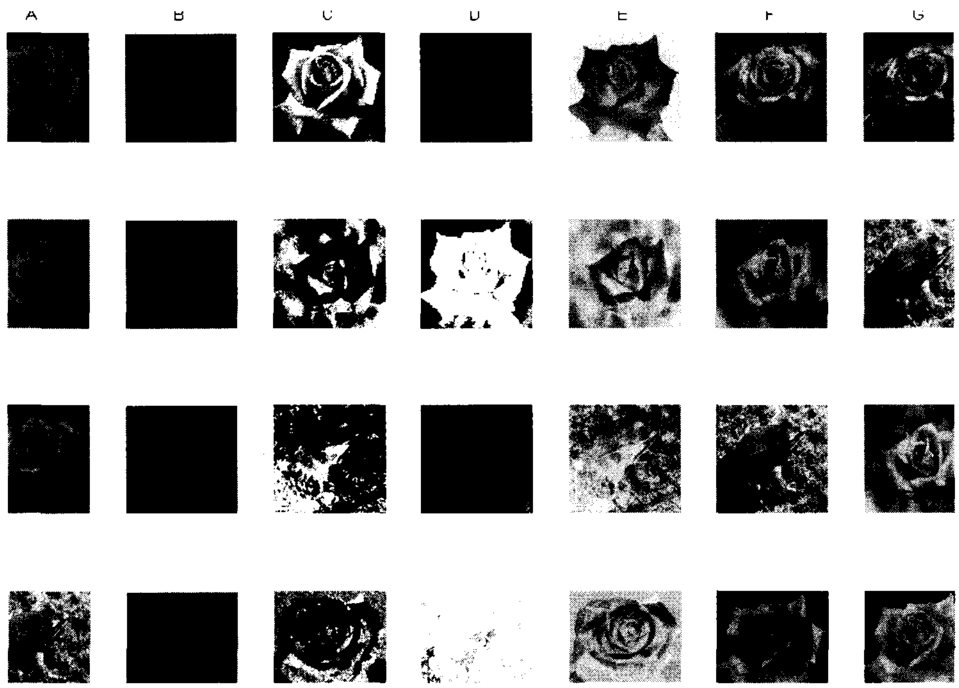
Fig.5 Comparison of the convergence speed

In general, ICA is combined with the innovation process for the separation image. However, this method needs a long processing time compared with the proposed algorithm. Table.3 is presented the difference of the processing time for each part of the image separation. The general ICA and Innovation methods required the total processing time about three times of the ICA with new fast EM.

Table 3. Comparison each processing time [sec]

	Total processing time	ICA	Innovation process	New fast EM
ICA and Innovation	37.5930	2.2980	35.2950	
ICA with new fast EM	12.2960	3.2340		9.06200

One of examples of the experimental results is demonstrated in Fig. 6. It demonstrates the results of each different method. The objections in the first column is the original images that are used for experiments. The second column represents the mixed images from which we desire to separate the faces in the first column. The third and fourth columns are the results of ICA and the innovation process without scaling respectively. The fifth column shows the results from ICA with scaling, which is inferior to the results of the sixth column that are



(A) Original images, (B) Mixed images, (C) Results of ICA, (D) Results of Innovation ,
(E) ICA with scaling, (F) ICA and Innovation with scaling, (G) ICA and new EM with scaling.

Fig. 6 Original image and comparison each process experimental results from the innovation process with scaling. However, the sixth column performed less effectively than the seventh column that are from the proposed method. As the Fig.6 shows, the proposed method performs better than the previously discussed methods.

6. Conclusion

In this paper, the new fast expectation and maximization (EM) is considered to reinforce the scheme of using independent component analysis(ICA) and time-dependent stochastic process. Innovation process performs a more accurate estimation of the ICA data model, but it has still many cases that fail to separate images[15] and requires much processing time. The proposed EM is used to remove the defects of these limitations. As experiments show, the new fast EM in the proposed scheme turns out to be a possible solution to the problems, but further studies on this subject are needed.

Reference

[1] N. Delfosse and P. Loubaton, "Adaptive blind separation of convolutive mixtures", In Proc. ICASSP'96, pp.2940-2943, 1996.

[2] C. Jutten and J. Herault, "Blind separation of sources, part Ⅰ: An adaptive algorithm based on neuromimetic architecture", Signal Processing, vol.24, pp.1-10, 1991.

[3] P. Pajunen, "Blind source separation using algorithmic information theory", Neurocomputing, 1998.

[4] H.-L. Nguyen Thi and C. Jutten, "Blind source separation for convolutive mixtures", Signal Processing, vol.45, pp.209-229, 1995.

[5] **S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures," *IEE Electronics Letters,* vol. 36, no. 9, pp. 848-849, April 27, 2000**

[6] A. Hyv rinen and E. Oja, "A fast fixed-point algorithm for independent component analysis", Neural Computation, vol.9, no.7, pp.1483-1492, 1997.

[7] **A. Hyvarinen, "Fast and robust fixed-point algorithms for Independent Component Analysis," *IEEE Trans. Neural Networks,* vol. 10, no. 3, pp. 626-634, May 1999.**

[8] A. Hyv rinen. "Independent component analysis by minimization of mutual information", Technical Report A46, Helsinki University of Technology, Laboratory of Computer and Information Science, 1997.

[9] J.-F. Cardoso and B. Hvam Laheld, "quivariant adaptive source separation", IEEE Trans. on Signal Processing, vol.44 no.12, pp.3017-3030, 1996.

[10] P. Comon, "Independent component analysis-a new concept?", Signal Processing, 36, pp.2940-2943, 1996..

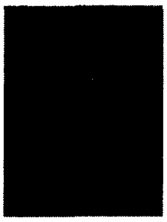
[11] A Hyvarinen, "Independent Component Analysis for Time-dependent Stochastic Processes", ICANN'98, pp. 541-546, 1998.

[12] Todd K. Moon, "The Expectation-Maximization Algorithm", IEEE Signal Processing Magazine, pp 47-60, Nov. 1996.

[13] Sung-Soo Kim and Jee-Hye Kang, "Improved Expectation and Maximization via a New Method for Initial Values", vol.13, No.4, pp. 416-426, Aug. 2003.

[14] **T.-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski, "A unifying information-theoretic framework for Independent Component Analysis," *Computers & Mathematics with Applications,* vol. 31, no. 11, pp. 1-21, March 2000**

저 자 소 개



Bum-Jin Oh

He received the B. S., M. S.degrees in electrical engineering from the Soongsil University, Korea, in 1982, 1984, respectively, and the Ph. D. degree from Chungbuk National University in 2001. He has been an Adjunct Professor at Hanbat National University, since 1984. He is working as the research head of the Inc. B. J. Information communication. His current research interests are Artificial Intelligence, Pattern Recognition, Microprocessor.

Phone : +82-42-273-7460
E-mail : akdang@chol.n

Sung-Soo Kim

He received the B. S. degree from Chungbuk National University, Chungbuk, Korea in 1983, the M. S. degree from University of Arkansas-Fayetteville in 1989, and the Ph. D. degree from University of Central Florida in 1997. He was a researcher at the Electronics and Telecommunications Research Institute in Korea form 1998 to 1999. From 1999 to 2001, he was an Assistant Professor at the Woosuk university. At present, he is an Associate Professor at the Department of Electrical and Computer

Engineering, Chungbuk National University, since 2001. His current research interests include Neural Networks, Fuzzy Theory, Artificial Intelligence, Intelligent Systems.

Phone : +82-43-261-2421
Fax : +82-43-261-3191
E-mail : Sungkim@chungbuk.ac.kr

Jee-Hye Kang
She received the B. S. degree from Chungbuk National University, Chungbuk, Korea in 2003. She is currently working toward the M. S. degree in the department of Electrical and Computer Engineering, Chungbuk National University. Her research interests are Statistical Signal Processing, Pattern Recognition, Artificial Intelligence.

Phone : +82-43-261-2421
Fax : +82-43-261-3191
E-mail : jeehye_kang@yahoo.co.kr