Journal of Computations & Modelling, vol.2, no.1, 2012, 109-122
ISSN: 1792-7625 (print), 1792-8850 (online)
International Scientific Press, 2012

A Regime Switching Independent Component Analysis Method for Temporal Data

Ho-Yin Yue[1,2]

Abstract

A mixture of independent component analysis method for temporal data is presented in this paper. The method is derived by modeling the observations as a mixture of ICA (mICA). mICA model has been applied to data classification and image processing. However, it is hard to use mICA in assigning class memberships of temporal data. In the proposed method, memberships of the data are modified according to its past values in the learning process. It shows that the proposed method is able to detect the switch between mixtures in highly overlapped data, which have smaller error than traditional mICA method.

Mathematics Subject Classification: 68Q32, 68T10

Keywords: automatic context switching, blind source separation, independent component analysis, mixture model

---

[1] Department of Computing, Hang Seng Management College, Hong Kong,
e-mail: willyyue@gmail.com

[2] School of Finance, Shanghai University of Finance and Economics, Shanghai, China

Article Info: *Received* : February 28, 2012. *Revised* : March 27, 2012
Published online : April 10, 2012

1 Introduction

In the last decade, Independent Component Analysis (ICA) [1] [2] became a hot topic in the field of signal processing and data mining. The aim of ICA is decomposing the observations linearly into a set of independent components which are statistically independent. The matrix maps the observations to the independent components is called demixing matrix and its inverse is mixing matrix. In traditional ICA method, both the mixing and demixing matrices are assumed to be constant. This assumption implies that the environment is unchanged throughout the time. However, real environment keeps changing, so the mixing and demixing matrices change with time. To model the dynamic of the mixing matrix, non-stationary independent component analysis [3] [4] and hidden Markov independent component analysis (HMICA) [5] [6] are two commonly used approaches. These two models have different assumptions about how observations are mixed from the independent sources. Assume there are m independent sources whose probability density functions are $p_m(s^m)$. The non-stationary ICA assumes that the sources are mixed linearly by the mixing matrix A_t with observational noise w_t (Equation 1). The mixing matrix A_t is assumed to change with the past observations (X_t) according to Equation 2. The graphical model describes non-stationary ICA is shown in Figure 1.

$$X_t = A_t S_t + w_t \tag{1}$$

$$\alpha_t = vec(A_t)$$

$$\alpha_{t+1} = F\alpha_t + v_t$$

$$P(\alpha_t \mid \Psi_t) = \frac{p(X_t \mid \alpha_t)\, p(\alpha_t \mid \Psi_t)}{p(X_t \mid \Psi_{t-1})} \tag{2}$$

where, v_t is zero-mean Gaussian noise with covariance Q, F is the state transition matrix, and Ψ_t denotes the collection of observations $\{X_1, X_2, \ldots, X_t\}$.

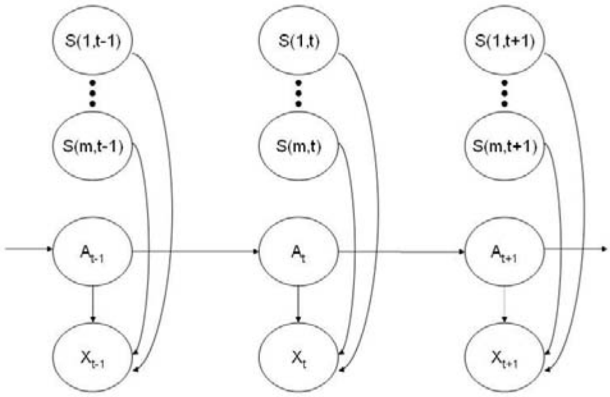

Figure 1: Generative model of the non-stationary ICA model

In the HMICA, every observation at time instant *t* belong to a state $q_t = k$. Each state *k* is associated with the mixing matrix A_k, the demixing matrix W_k and the source parameters vector Θ_k. With the source generated at time *t* is given as $S_t = f(q_t, S_{t-1})$. The observation at time *t* is generated as $X_t = A_k S_t$. Figure 2 shows the general model of the HMICA graphically.

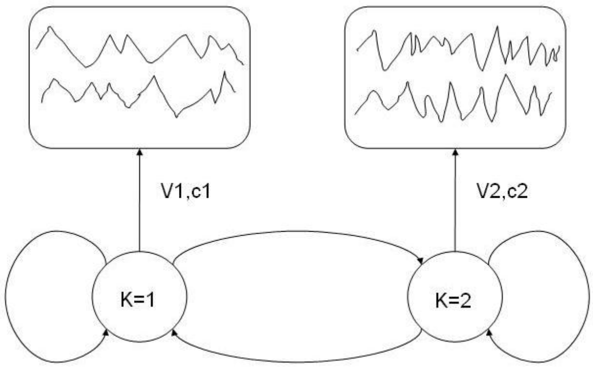

Figure 2: Generative model of the HMICA model

To summarize non-stationary ICA and HMICA, both models have the same general form, $X_t = A_t S_t$, which is assumed that the mixing matrix (A_t) is changing with time. The main difference between them is the assumption on the dynamic of the mixing matrix. The non-stationary ICA assumes that the mixing matrix is a function of past observations. The HMICA assumes that the mixing matrix changes in a form of Hidden Markov Model. However, neither the non-stationary ICA nor the HMICA can model the observations well when observations are generated by a mixture of co-existing systems.

In this paper, a Temporal Mixture of Independent Component Analysis (tmICA) method is suggested to model this kind of mixture system. In the proposed method, memberships of the data are modified according to its past values in the learning process. It shows that the proposed method is able to detect the switch between mixtures in highly overlapped data, which have smaller error than traditional mICA method. This paper is organized as follow: In Section 2, traditional mixture of ICA modeling was presented. An independent component analysis method for temporal data is proposed in section 3. Section 4 and 5 are the experiments and conclusion of the paper respectively.

2 Mixture of ICA Modeling

Traditional ICA [7] allows only one mixing matrix in the system. It is unable for traditional ICA method to decompose the observations correctly if observations are produced by several co-existing mixing/demixing system. mICA [8] relaxed the traditional ICA by assuming that the observations are generated from sources in more than one mixing systems; keeping the sources within the same mixing system statistically independent with others at the same time. Using mICA, some applications have been built in image processing [9] [10] and data clustering [11].

In mICA, the data $X_n = \left[x_{1,n}, \quad x_{2,n}, \quad \cdots \quad , x_{m,n} \right]^T$ is m dimensional observation at time n generated by a mixture density model [12]. The probability of generating a data point, from a *K*-component mixture model is:

$$p(X_n \mid \Theta) = \sum_{k=1}^{K} p(X_n \mid C_c, \theta_k) p(C_k) \tag{3}$$

Also, the probability that X_n is generated from component C_i is:

$$p(C_i \mid X_n) = \frac{p(X_n \mid C_k, \theta_k) p(C_k)}{\sum_{k=1}^{K} p(X_n, C_k) p(C_k)} \tag{4}$$

where, θ_k is the vector of unknown parameters for *k*th mixture. C_k denotes the *kth* mixture and the number of mixture K is assumed to be known in advance. Data within *kth* mixture is described by the standard ICA model:

$$X_n = A_k S_{k,n} \tag{5}$$

where A_k is a $m \times m$ mixing matrix and $S_{k,n} = \left[s_{k,1,n}, \quad s_{k,2,n}, \quad \cdots \quad , s_{k,m,n} \right]^T$ is the m dimensional source for *k*th mixture respectively.

It is shown that the model parameters can be estimated by maximizing the sum of the log-likelihood of the data (Equation 6) through Expectation-Maximization (EM) algorithm [11].

$$L = \sum_{n=1}^{N} \sum_{k=1}^{K} \log\left(p(X_n \mid C_k, \theta_k) p(C_k) \right) \tag{6}$$

A survey of mICA was given in [13]. The main difficulty for applying mICA on temporal data is that a single data does not contain enough information for assigning the class membership. This problem is the most serious when observations having similar membership values among different mixture. In this paper, a temporal mICA method, tmICA hereafter, is proposed. In tmICA, memberships for different ICA mixtures are used to model the changes of system gating. Take cocktail party as an example, when the microphone was moved from one position to another position at time *t*, the mixing matrix would be changed at time *t* with the position of the microphone for the same sources. In order to model such a case with mICA, two ICA mixtures were used. Suppose A_1 and A_2 are mixing matrices used in mICA. Given any observation, its membership of A_1 is greater than membership of A_2 before time *t*, and the membership of A_1 becomes lower than A_2 after time *t*. In other words, the change of the mixing matrix can be indicated by the change of membership value.

3 Temporal Mixture of ICA Modeling

In this section, description on the learning process of tmICA is presented. Given a m dimensional observations, $X_n = \left[x_{1,n}, \quad x_{2,n}, \quad \ldots \quad , x_{m,n} \right]^T$, which are generated by a K-component mixtures density model. Probability of X_t being generated from mixture C_k is given by Equation 4. For the *k*th mixture, data is described by the standard ICA model:

$$X_t = A_k S_{k,t} \tag{7}$$

In tmICA, $p(C_k \mid X_t)$ is used to represent the membership of the observation X_t to the *k*th ICA mixture at time *t*. The details of the source model used to calculate $p(C_k \mid X_t)$ is given in [7]. Memberships are assumed to change smoothly across time.

Therefore, after obtained $p(C_k \mid X_t)$, we smooth the value by:

$$p(C_k \mid X_t)_{\text{smooth}} = \sum_{i=1}^{L} \gamma(i) p(C_k \mid X_{t-i}) \quad . (8)$$

where, $\gamma(i) > 0$ is a constant that represents the importance for affecting, $p(C_k \mid X_{t-i})$ and $p(C_k \mid X_t)$ with the constraint $\sum_{i=1}^{L} \gamma(i) = 1$.

However, smoothing the membership probability is not enough for obtaining a stable result. In tmICA, Smoothed probabilities are further modified to decrease the effect of ambiguity $p(C_k \mid X_t)$ in the membership assignment.

After the memberships are smoothed by Equation 8, memberships are modified as follow:

if $p(C_k \mid X_t)_{\text{smooth}} \geq p(C_i \mid X_t)_{\text{smooth}}$, $\forall i \neq k$ then,

$$p(C_k \mid X_t)_{\text{modified}} = \alpha p(C_k \mid X_t)_{\text{smooth}} .$$

if $p(C_k \mid X_t)_{\text{smooth}} < p(C_i \mid X_t)_{\text{smooth}}$, $\forall i \neq k$ then,

$$p(C_k \mid X_t)_{\text{modified}} = \frac{1}{\alpha} p(C_k \mid X_t)_{\text{smooth}}$$

where, $\alpha > 1$ is modification factor.

Then, normalization is performed according to Equation 9 in order to keep the constraint $\sum_{i=1}^{K} p(C_i \mid X_t) = 1$.

$$p(C_k \mid X_t)_{\text{norm}} = \frac{p(C_k \mid X_t)_{\text{modified}}}{\sum_{i=1}^{K} p(C_i \mid X_t)_{\text{modified}}} \tag{9}$$

After assigned the membership values, these modified memberships are used together with FastICA [15][16] to calculate the mixing matrix in each mixture. The algorithm of tmICA is outlined in Algorithm 1. In the next section, tmICA are tested with sources which are highly overlapped.

Algorithm 1: Learning algorithm for tmICA.

Input: A *M*-dimensional temporal data X_t, and the number of ICA mixtures *K*.

Output: A set *N*-dimensional of estimated sources, $N \times N$ mixing matrices of each ICA mixture, and membership probabilities for every observation in each ICA mixture.

Step 1:

Choose an initial (e.g. random) demixing matrix $B(k)$ for all mixtures.

Step 2:

Compute the independent components (S_k) and the membership probability ($p(C_k \mid X_t)$) for each pair of the *kth* mixture and observation at time *t*.

2.1. for all mixtures $S_{k,t} = B_k X_t$

2.2. for all pairs of mixture *k* and time *t*.

Step 3:

Modify the membership probability for each pair of the *kth* mixture and observation at time *t*.

3.1. $p(C_k \mid X_t) \leftarrow \sum_{i=1}^{L} \gamma(i) p(C_k \mid X_{t-i})$

3.2. if $p(C_k \mid X_t) \geq p(C_i \mid X_t), \quad \forall i \neq k$ then,

$$p(C_k \mid X_t) \leftarrow \alpha p(C_k \mid X_t).$$

if $p(C_k \mid X_t) < p(C_i \mid X_t), \quad \forall i \neq k$ then,

$$p(C_k \mid X_t) \leftarrow \frac{1}{\alpha} p(C_k \mid X_t)$$

3.3 $p(C_k \mid X_t) \leftarrow \dfrac{p(C_k \mid X_t)}{\sum_{i=1}^{K} p(C_i \mid X_t)}$

Step 4:

Perform FastICA algorithm for each mixture *k*.

Center the data to make its mean zero.

Compute the weighted correlation matrix (C_k) for each mixture.

4.1 $\beta_{k,i} = -E\{s_{k,i,t} \tanh(s_{k,i,t})\} \quad \forall\ i = 1, 2, \ldots, M$.

4.2 $\alpha_{k,i} = -1/(\beta_{k,i} + E\{1 - \tanh(s_{k,i,t})^2\}) \quad \forall\ i = 1, 2, \ldots, M$.

Update the separating matrix by:

4.3. $B_k \leftarrow B_k + diag(\alpha_{k,i})[diag(\beta_{k,i}) + E\{\tanh(s_{k,i,t}) s_{k,i,t}^T\} B_k]$

Decorrelate and normalize the separating matrix by:

4.4 $B_k \leftarrow (B_k C_k B_k^T)^{-1/2} B_k$

Step 5:

If not converged, go back to Step 2.

4 Experimental Results

In this section, tmICA was applied on a set of temporal data from ICA mixture with two components. The same 2D sources are used in both mixtures. One dimension of the sources is triangular wave while another is sine wave. The sources were shown in Figure 3.

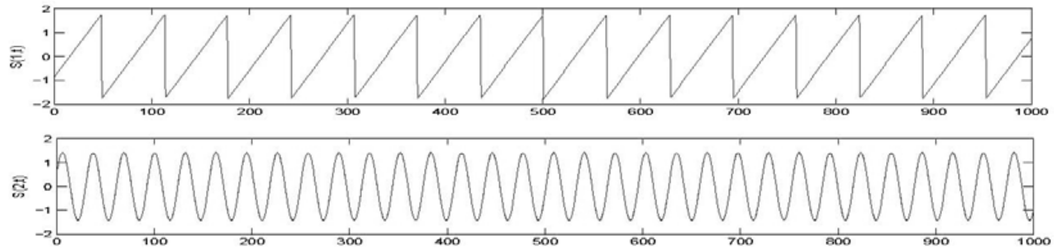

Figure 3: The sources used to generate the observations

1000 observations were generated in the sense that the first 500 observations were generated from the first mixture while the remaining 500 observations were generated from another mixture. The observations from the mixture were shown in Figure 4.

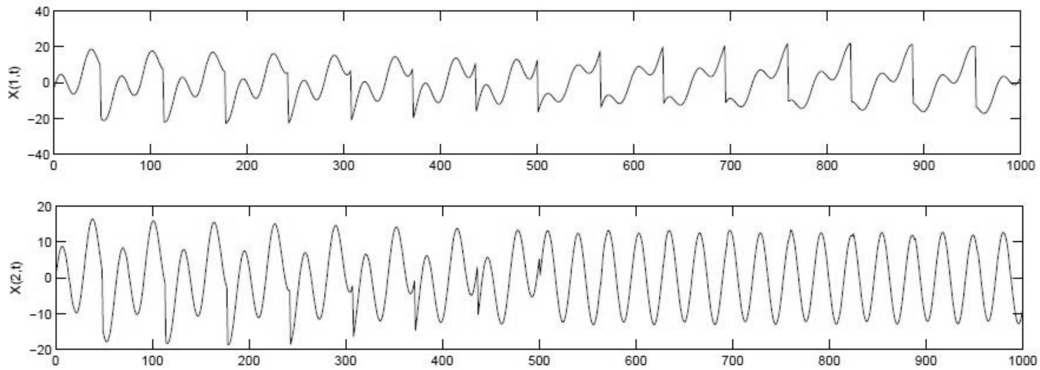

Figure 4: The observations generated from the ICA mixture

From the scatter plot of the observations (Figure 5), it shows that the observations from the mixture are highly overlapped. So, to identify the truth memberships and the sources from the observations are difficult.

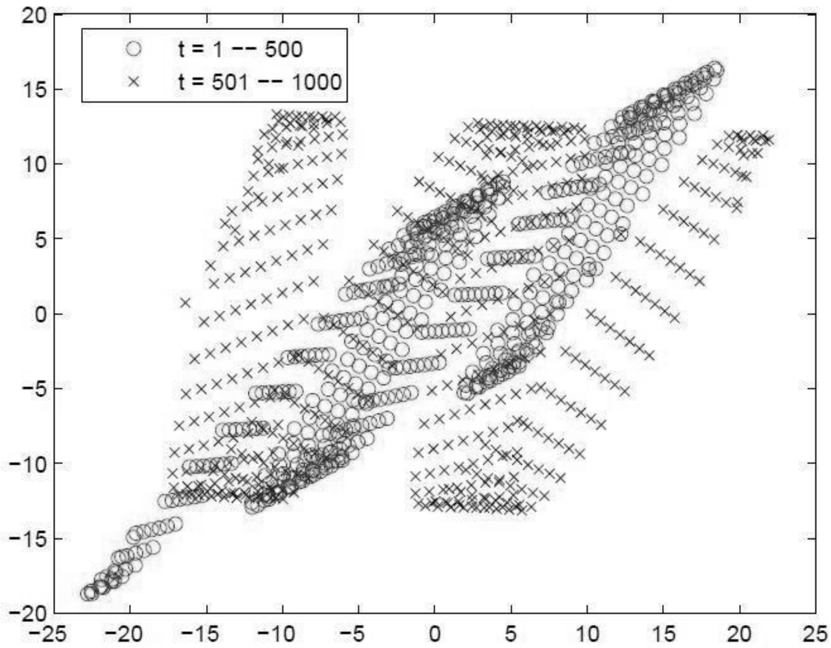

Figure 5: The scatter plot of the observations

The sources of the mixtures which recovered from tmICA were shown in Figure 6 and Figure 7. Membership results show that the first ICA mixture dominates to the first 500 observations and the second ICA mixture dominates to the remaining 500 observations. The sources recovered from tmICA are very close to the sources used to generate the observations. The results of tmICA are compared with mICA method [13]. Results of mICA are shown in Figure 8 and Figure 9. It is found that although both methods are capable to decide the membership probability correctly. tmICA produces a more accurate recovered sources than mICA method. So, it is concluded that tmICA has comparatively better sources recovery power in data which observations generated from highly overlapped mixtures.

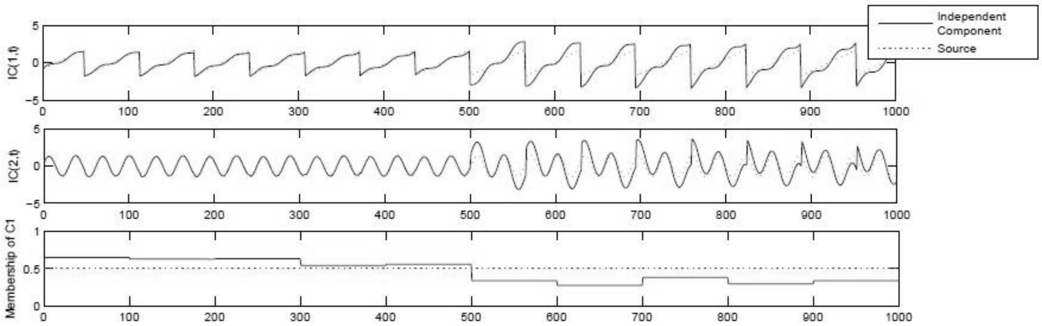

Figure 6: The independent components and the membership for the first mixture with tmICA

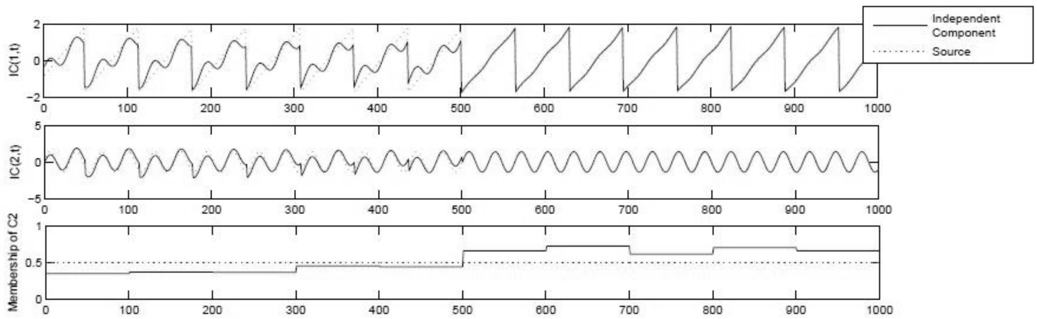

Figure 7: The independent components and the membership for the second mixture with tmICA

5 Conclusion

An independent component analysis method for temporal data, tmICA, is presented in this paper. In tmICA, observations are modeled as a mixture of systems. Each system is further described by an ICA model. Memberships of observations are used to decide the degree of influence for a mixture towards the observations. When compared with mICA model, memberships in tmICA are modified in the estimation process. This modification provides a better assignment in the memberships. According to the experimental results, tmICA shows a excellent power in discovering the context switch of observations automatically in highly overlapping data. The estimated sources are closer to the true sources when compare with those estimated sources from ICA mixture model [8].

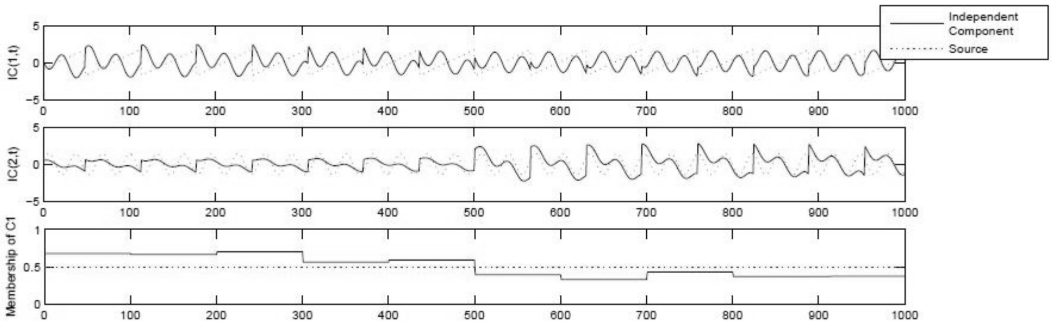

Figure 8: The independent components and the membership for the first mixture without tmICA

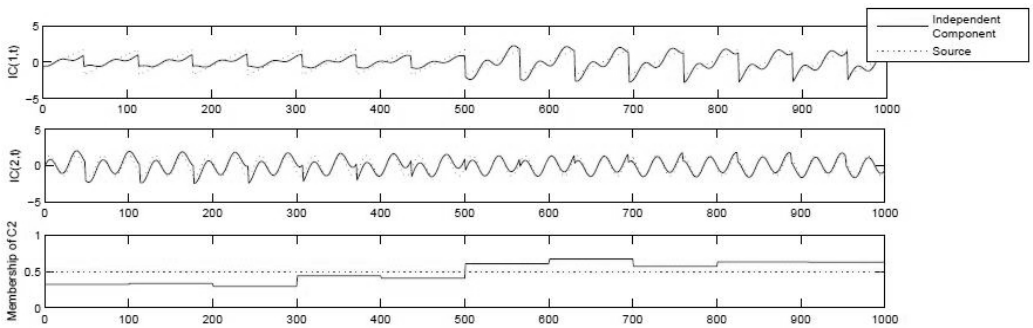

Figure 9: The independent components and the membership for the second mixture without tmICA

References

[1] S. Amari and J. Cardoso, Blind source separation – semiparametric statistical approach, *IEEE Transactions on Signal Processing*, **5**(11), (1997), 2692-2700.

[2] A. Bell and T. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, *Neural Computation*, **7**, (1995), 1129-1159.

[3] R. Everson and S. Roberts, Non-stationary independent component analysis, *ICANN-99*, (1999).

[4] R. Everson and S. Roberts, Particle filters for non-stationary ICA, *Advances in independent components analysis*, M. Girolami, Ed. Springer, 2000.

[5] W. Penny, R. Everson and S. Roberts, Hidden Markov independent components analysis, *Advances in independent components analysis*, M. Girolami, Ed. Springer, 2000.

[6] W. Penny, S. Roberts and R. Everson, Hidden Markov independent components for biosignal analysis, *MEDSIP-2000*, *International Conference on Advances in Medical Signal and Information Processing* 2000, Bristol, UK, (September 2000).

[7] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 204 – 207, 2001.

[8] T. Lee, M. Lewicki and T. Sejnowski, Unsupervised classification with non-gaussian mixture models using ica, *Proceedings of the 1998 conference on Advances in neural information processing systems II*, (1999), 508-514.

[9] C. Shah, M. Arora, S. Robila and P. Varshney, ICA mixture model based unsupervised classification of hyperspectral imagery, *Proceedings of the 31st Applied Imagery Pattern Recognition Workshop (AIPR'02)*, (2002), 29-35.

[10] T. Lee and M. Lewicki, Unsupervised image classification, segmentation, and enhancement using ica mixtures models, *IEEE Transacations on Image Processing*, **11**(3), (2002), 270-279.

[11] S. Roberts and W. Penny, Mixtures of independent component analysers, *Proceedings of the International Conference on Artificial Neural Networks*, (2001), 527-534.

[12] R. Duda and P. Hart, *Pattern classification and scene analysis*, Wiley, 1973.

[13] T. Lee, M. Lewicki and T. Sejnowski, ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation, *IEEE Transacations on Pattern Analysis and Machine Intelligence*, **22**(10), (October, 2000), 1078-1089.

[14] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, *IEEE Transacations on Neural Networks*, **10**(3), (1999), 626-634.

[15] A. Hyvärinen, The fixed-point algorithm and maximum likelihood estimation for independent component analysis, *Neural Processing Letters*, **10**(1), (1999), 1-5.