Underdetermined blind separation of mixtures of an unknown number of sources with additive white and pink noises

Ossama S. Alshabrawy[1,3], Aboul ella Hassanien[2,3]

[1]Faculty of Science (Computer Science), Damietta University, Damietta, Egypt
[2]Faculty of Computers & Information, Cairo University, Cairo, Egypt
[5]Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net

Abstract. In this paper we propose an approach for underdetermined blind separation in the case of additive Gaussian white noise and pink noise in addition to the most challenging case where the number of source signals is unknown. In addition to that, the proposed approach is applicable in the case of separating $I+3$ source signals from I mixtures with an unknown number of source signals and the mixtures have additive two kinds of noises. This situation is more challenging and also more suitable to practical real world problems. Moreover, unlike to some traditional approaches, the sparsity conditions are not imposed. Firstly, the number of source signals is approximated and estimated using multiple source detection, followed by an algorithm for estimating the mixing matrix based on combining short time Fourier transform and rough-fuzzy clustering. Then, the mixed signals are normalized and the source signals are recovered using multi-layer modified Gradient descent Local Hierarchical Alternating Least Squares Algorithm exploiting the number of source signals estimated , and the mixing matrix obtained as an input and initialized by multiplicative algorithm for matrix factorization based on alpha divergence. The computer simulation results show that the proposed approach can separate $I+3$ source signals from I mixed signals, and it has superior evaluation performance compared to some traditional approaches in recent references.

Keywords: Underdetermined Blind Source Separation; Rough Fuzzy clustering; Short Time Fourier transform; Hierarchical Alternating Least Squares; Multi-Layer algorithm

1 Introduction

In recent years, Blind Source Separation (or Blind Signal Separation, BSS) plays an important role in solving some problems in soft computing. Throughout the last two decades, Blind Signal Separation in combination with computer science applications, and information theory, has a variety range of applications in the fields of speech processing, digital communication systems, wireless communications, data mining, water marking, medical imaging, and biomedical engineering

[2],[3],[4],[5],[6]. Blindness or blind separation means that no or very little information is known about the mixing system or the mixing process or the original source signals that need to be separated or recovered [1].

The source signals are assumed to be as statistically independent as possible in most of the traditional BSS approaches given the observable mixed signals or sensors data. Another hypothesis by these approaches is that the mixing matrix is of full column rank. In many real-world situations, however, this hypothesis is not applicable. Consequently, recovering the original source signals by multiplying the observable data mixtures by the pseudo inverse of the mixing matrix cannot be used. This makes recovering the source signals a difficult and very challenging task [7],[12]. In practical terms, the overdetermined mixture assumption does not always hold. For instance, in radio communications the probability of receiving more source signals than sensors data or observed mixed signals increases with increase of reception bandwidth, hence it is urgently necessary to solve the problem of underdetermined blind source separation (UBSS)[8]. The motivation of this paper is to separate sparse, and super and sub-Gaussian signals in the underdetermined case with additive noise such as Gaussian white noise and pink noise in addition to the case that the number of source signals is unknown without imposing any sparsity conditions. Another motivation of this paper and to increase the performance of the separation in the previous cases together.

The rest of the paper is organized as follows. In Section 2, we present the details of the proposed approach. In section 3, we show the analysis of typical experiments and the results obtained by different BSS methods, where the simulation results show the effectiveness and high performance of the proposed algorithm. Finally, a short conclusion and future work are presented in Section 4.

2 The proposed UBSS algorithm

In this section, the proposed approach is presented starting with estimating the number of source signals followed by the mixing matrix estimation knowing only the observable mixtures matrix which contains noise. Also, a method for GMF gradient descent based update rules initialized with matrix factorization multiplicative algorithm based on alpha divergence is introduced followed by the main algorithm for source signals recovery based on multi-layer matrix factorization.

2.1 Estimation of number of source signals and the mixing matrix

Traditional approaches require that for each source there exist many TF points of single source occupancy (SSO). However, single source detection (SSD) requires that there exists at least one TF point of SSO and is hence less restrictive than the other approaches [9]. The short time Fourier transform (STFT) of the i^{th} observed signal is defined by the following equation:

$$Y_i^{Fourier}(t,r) = \sum_{l=0}^{\infty} h(l-t)X_i(l)e^{-jrl} \tag{1}$$

At frame t and frequency bin r where $h(l)$ is a window sequence. In equation (1), $i = 1, 2, .., I$; $t = 0, 1, , T-1$ are the sampling points over the time domain and $r = 0, 1, .., T-1$ are the sampling points over the frequency domain. The SSD is based on the ratio of the TF transforms and finds a set of TF points where a single source is active for each source. Consequently, for a given $\varepsilon > 0$ the set that represents the detected points can be obtained by the following equation:

$$\begin{aligned} \chi_F = \{(t,r)| \left\| Im[\frac{Y^{Fo}(t,r)}{Y_1^{Fo}(t,r)}] \right\|_F < \varepsilon, \\ Y_1^{Fo}(t,r) \neq 0\} \end{aligned} \tag{2}$$

where, Y^{Fo} represents the matrix of $Y^{Fourier}$ obtained from Eq. (1), $Im[.]$ denotes the imaginary part. We can choose any of the mixture instead of Y_1.

During clustering the observable mixtures after incorporating STFT, we need to determine the number of source signals. Since there is an overlap between the data objects, estimating the number of sources require an efficient validity index [16] and can be given by the following equation:

$$V(\beta, c) = Scat(c) + \frac{Sep(c)}{Sep(C_{max})} \tag{3}$$

Where β is the cluster centers, c is the number of clusters, and C_max is the chosen maximum number of clusters. Here,

$$\frac{\frac{1}{c}\sum_{i=1}^{c} \|\sigma(\beta_i)\|}{\|\sigma(Y)\|} \tag{4}$$

Also, the value of $Scat(c)$ varies from 0 to 1. The term that represents the separation between clusters is defined by the following equation:

$$\frac{D_{max}^2}{D_{min}^2} \sum_{i=1}^{c} \left(\sum_{j=1}^{c} \|\beta_i - \beta_j\|^2 \right)^{-1} \tag{5}$$

Where $D_{min} = \min_{i \neq j} \|\beta_i - \beta_j\|$, $D_{max} = \max_{i,j} \|\beta_i - \beta_j\|$

After clustering, the ith column vector of A, denoted as $\hat{a}_i$, is estimated as:

$$\hat{a}_i = \frac{1}{|\chi_{C_i}|} \sum_{(t,r) \in \chi_{C_i}} Re[Y^{Fo}(t,r)] \tag{6}$$

Here, χ_{C_i} represents the number of TF points in cluster C_i for $i = 1, 2, \cdots, J$.

2.2 The initialization technique for extracting the source signals

Before The source signals estimation algorithm will be initialized using multiplicative alpha GMF algorithm. This initialization technique will help to improve the results and obtain a better separation performance. The cost function for alpha divergence is outlined in the following equation:

$$COST^{\alpha}(Y||AX) = \frac{1}{\alpha(\alpha-1)} \sum_{it} \left([Y]_{it}[AX]_{it}^{1-\alpha} - \alpha Y_{it} + (\alpha-1)[AX]_{it}\right) \tag{7}$$

The final alpha multiplicative learning algorithm is expressed by the following update rules:

$$\begin{aligned} (x_{jt})_{new} &= (x_{jt})_{old} \left(\frac{\sum_{i=1}^{I} a_{ij}(y_{it}/[AX]_{it})^{\alpha}}{\sum_{i=1}^{I} a_i j} \right)^{\frac{1}{\alpha}}, \\ (a_{ij})_{new} &= (a_{ij})_{old} \left(\frac{\sum_{t=1}^{T} (y_{it}/[AX]_{it})^{\alpha} x_{jt}}{\sum_{t=1}^{T} x_j t} \right)^{\frac{1}{\alpha}} ; \alpha \neq 0 \end{aligned} \tag{8}$$

But here we will make initialization to only the source signals, so we will exploit only the first part of the equation.

2.3 Modified gradient descent local Hierarchical Alternating Least Squares

This section will introduce a quick overview on the analysis and derivation of Hierarchical Alternating Least Squares (HALS). Hierarchical Alternating Least Squares method is suitable for large-scale NMF problems, and it can be applied also for sparse non-negative coding or representation [10]. HALS algorithm can be derived by choosing exploiting a set of local cost functions such as Alpha- and Beta-divergences, and the squared Euclidean distance. Then perform consecutive or simultaneous minimization of these local cost functions. For example, using gradient descent or some nonlinear transformations. The family of HALS algorithms can not only do better for the over-determined case of BSS, but they can also solve underdetermined BSS case under some simple conditions. Especially for the multi-layer technique [11], the extensive experiments and simulation results show the superior performance and validity of the family of HALS algorithms. HALS is used here in this paper in a modified version by relaxing the non-negativity constraints and depending on the gradient descent algorithm.

Denote $A = [a_1, a_2, \cdots, a_j]$ and $S = X^T = [s_1, s_2, \cdots, s_j]$ to express the squared Euclidean cost function as:

$$J[a_1,..,a_j,s1,..,s_j] = \frac{1}{2}\|Y - AS^T\|_f^2 = \frac{1}{2}\|Y - \sum_{j=1}^{J} a_j s_j^T\|_{Fro}^2 \tag{9}$$

where, *Fro* refers to the Frobenious norm. The main idea is to define the residues followed by minimizing the set of local cost functions alternatively with respect to the parameters a_i and s_j. The residues can be obtained as:

$$Y^{(j)} = Y - \sum_{p \neq J} a_p s_p^T = Y - AS^T + a_j s_P^T = E + a_j s_j^T (j = [1, 2, ..., J]) \tag{10}$$

Then the alternative minimization of the set of cost functions can be obtained by [12]:

$$Cost_{Fro}^{(j)}(Y^{(j)}||a_j s_j^T) = \frac{1}{2}\|Y(j) - a_j s_j^T\|_{Fro}^2, \quad For j = 1, 2, ..., J \tag{11}$$

The optimality conditions for the set of cost functions (11) can be defined as:

$$a_j \otimes \nabla_{a_j} Cost_{Fro}^{(j)}(Y^{(j)}||a_j s_j^T) = 0 \tag{12}$$

$$s_j \otimes \nabla_{s_j} Cost_{Fro}^{(j)}(Y^{(j)}||a_j s_j^T) = 0 \tag{13}$$

The gradients of the local cost functions in Eq. (11) are computed with respect to the unknown vectors a_j and s_j to obtain the critical or stationary points with the assumption that the other vectors are fixed by the following equation:

$$\nabla_{a_j} Cost_{Fro}^{(j)}(Y^{(j)}||a_j s_j^T) = a_j s_j^T s_j - Y^{(j)} s_j \tag{14}$$

$$\nabla_{s_j} Cost_{Fro}^{(j)}(Y^{(j)}||a_j s_j^T) = a_j^T a_j s_j - Y^{(j)T} a_j \tag{15}$$

Without resorting to any non-negativity constraints on the entries of vectors a_j and s_j $\forall j$, the critical points can be obtained by the following simple update rules:

$$s_j \leftarrow \frac{1}{a_j^T a_j}(Y^{(j)T} a_j) = \frac{1}{a_j^T a_j} Y^{(j)T} a_j. \tag{16}$$

$$a_j \leftarrow \frac{1}{s_j^T s_j}(Y^{(j)} s_j) = \frac{1}{s_j^T s_j} Y^{(j)} s_j, (j = 1, 2, ..., J) \tag{17}$$

2.4 Multi layer matrix factorization algorithm

The multi-layer technique [17], also known hierarchical multi-layer technique. The mixing matrix A is replaced in multi-layer matrix factorization by a set of cascaded matrices as follows:

$$Y = A^{(1)}A^{(2)}A^{(3)}....A^{(L)}X + E \tag{18}$$

where L is the number of layers [18].Since the model is linear, all the matrices can be merged into a single matrix A if no special constraints are imposed upon the individual matrices $A^{l}(l = 1, 2, ..., L)$. The multi-layer technique is used to improve the performance of the NMF. However, in this paper is used to improve GMF algorithm stated in the previous section. In the first step of the multi-layer algorithm the basic approximate decomposition $Y \cong A^{(1)}X^{(1)} \in R^{I\times T}$ can be performed. Then the result obtained from the first step can be used to build up a new assignment, so $X^{(1)} \cong A^{(2)}X^{(2)} \in R^{I\times T}$. The learning update rules are performed hierarchically or sequentially i.e. layer by layer. Repeating these update rules according to the number of layers to get the final form of the multi-layer model as follows:

$$Y \cong A^{(1)}A^{(2)}A^{(3)}....A^{(L)}X^{(L)} \tag{19}$$

where, $A = A^{(1)}A^{(2)}A^{(3)}....A^{(L)}$ and $X = X^{(L)}$. Get the final result for XH which is the final estimation of the source signals, and then compute the signal-to-noise ratio as in the simulation results.

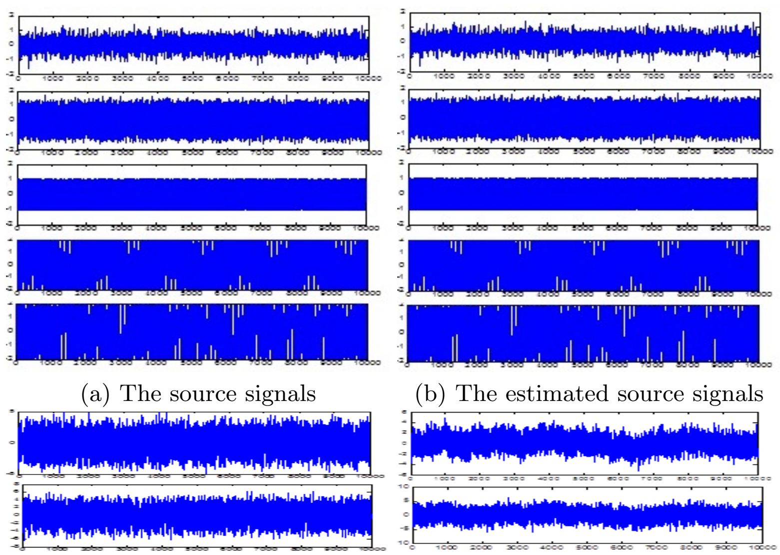
3 Simulation Results

In this section, the performance and effectiveness of the proposed approach will be discussed by comparing results of experiments and stimulations. Simulations were performed on synthetically generated signals using the proposed approach and some other approaches. In the simulations, sparse, super- and sub-Gaussian signals were separated from the underdetermined noisy mixtures in the challenging case where the true number of source signals is unknown. The types of noise that is considered in this paper is the white noise and pink noise. The parameter inputs of the modified Modified gradient descent local Hierarchical alternating least squares algorithm are the observable mixtures matrix Y, and the mixing matrix A obtained by the method stated above. We choose the maximum number of iterations to be only 50 iterations. We investigate the performance of the proposed UBSS approach in the above mentioned cases by comparing its results with the results of approaches in Snoussi and Idier (2006) [13], Peng and Xiang (2010) [14], and S. Sun et al. (2012) [15]. Here, the simulation of the separation of a variety of sparse, non-sparse, and super- and sub-Gaussian signals are stated.

3.1 Separation of synthetic signals with additive noise

Here, the simulation of the separation of a variety of sparse, non-sparse, and super- and sub-Gaussian signals are stated. All these cases are in the presence of above mentioned kinds of noise.

Sparse, non-sparse, and super- and sub-Gaussian signals The effectiveness of the proposed UBSS approach is investigated by comparing the results of the proposed approach with the methods mentioned above. We chose the number of mixtures to be only 2 and the number of sources to be 5 to create a more challenging case and to prove that the proposed approach can separate $I + 3$ source signals from I mixtures. The five source signals, two observable mixtures that contains additive white noise, and pink noise, and the estimated source signals are plotted in Figs. 1. The number of sampling time points is 10,000. The simulation results of the proposed approach in addition to those of the five different UBSS methods are shown in Fig. 2.

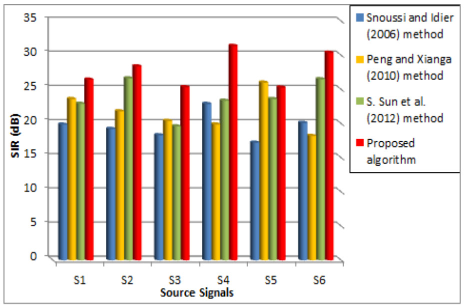


(a) The source signals (b) The estimated source signals

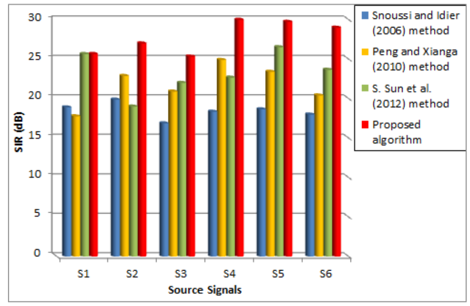
(c) The mixed signals with additive white noise (d) The mixed signals with additive pink noise

Fig. 1. Source and estimated source signals

We note from Fig. 2 that the proposed approach achieves about 4 dB higher SNR for J=7 sources with only two mixtures than the highest performance algorithm among the other five approaches. Likewise, the proposed approach achieves higher performance in case of Pink noise. From the results in Figs. 1, and 2, we can conclude that the separation performance of the proposed approach is very high, has faster convergence, and can separate sparse, non-sparse, and super- and sub-Gaussian signals in addition to sparse and non-sparse signals when compared with the other approaches.



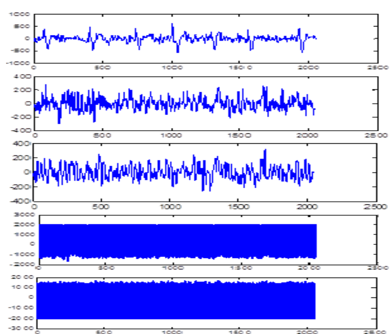
(a) Performance estimation of the source signals from 3 observable mixtures with additive white noise



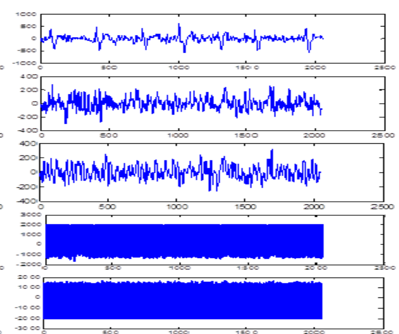
(b) Performance estimation of the source signals from 3 observable mixtures with additive pink noise

Fig. 2. Performance estimation of the synthetic source signals with additive white and pink noises

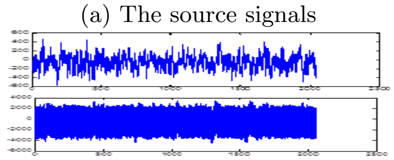
3.2 Separation of real-world signals



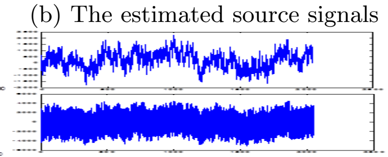
(a) The source signals



(b) The estimated source signals



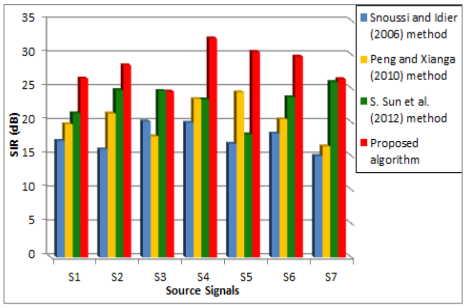
(c) The mixed signals with additive white noise



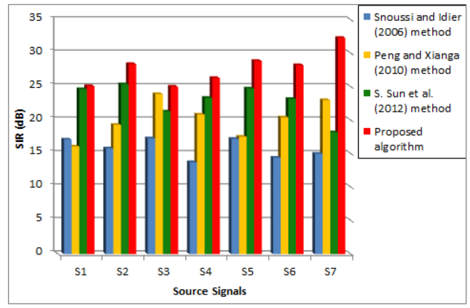
(d) The mixed signals with additive pink noise

Fig. 3. Source and estimated source EEG signals

To further measure the estimation performance of the source recovery approach in the previous Subsection, other comparisons are performed with the three other approaches using a dataset of real-world signals that are available to



(a) Performance estimation of the seven EEG19 source signals from 3 observable mixtures with additive white noise



(b) Performance estimation of the seven EEG19 source signals from 3 observable mixtures with additive pink noise

Fig. 4. Performance estimation of the seven EEG19 source signals from 3 observable mixtures

download from `http://www.bsp.brain.riken.jp/ICALAB/ICALABSignalProc/benchmarks`. This benchmark is EEG19, which contains 19 electroencephalogram (EEG) signals with clear heart, eye movement, and eye blinking artifacts. Only five signals are chosen as shown in Fig. 3 (a). Note that the first four signals X_1, X_2, and X_3 are super-Gaussian while the last two signals X_4 and X_5 are sub-Gaussian. Fig. 3 (c) and (d) shows the observable mixtures signals. The estimated source signals are shown in Fig. 3 (b). The mixing matrix A is the same as in the previous experiments. Figs. 4 shows the SIR results for all methods tested.

4 Conclusions

In this paper, we addressed the problem of underdetermined blind source separation with the challenging case that to separate I+3 source signals from I mixtures with additive white and pink noises. A new two-step approach for optimum estimation of the source signals. In this approach, STFT is combined with rough fuzzy c-means clustering to estimate the mixing matrix. Then the source signals are estimated by a modified gradient descent local Hierarchical alternating least squares based general matrix factorization. Simulation experiments demonstrated the validity and superior performance of the proposed approach. Our future work will include more types of noise in the case of the number of source signals is unknown.

References

1. Ossama S. Alshabrawy, M. E. Ghoniem, W. A. Awad and Aboul ella hassanien, "Underdetermined Blind Source Separation based on Fuzzy C-Means and Semi-Nonnegative Matrix Factorization", IEEE Federated Conference on Computer

Science and Iinformation Systems (FedCSIS), Wroclaw, Poland, pp. 695-700, 9-12 September, 2012

2. Yadong, Liu, Zongtan, Zhou, Dewen, Hu, “A novel method for spatio temporal pattern analysis of brain fMRI data”, Science in China Series F: Information Sciences, vol. 48, no. 2, pp. 151-160, 2005.
3. Araki, S., Makino, S., Blin, A., “Underdetermined blind separation for speech in real environment with sparseness and ICA”, In Proceedings of the ICASSP04, Montreal, Canada, pp.881-884, 2004.
4. Ohnishi, Naoya, Imiya, Atsushi, Independent component analysis of optical flow for robot navigation, Neurocomputing, vol.7, nos. 10-12, pp. 21402163, 2008.
5. Tonazzini, Anna, Bedini, Luigi, Salerno, Emanuele, A Markov model for blind image separation by a mean-field EM algorithm, IEEE Transactions on Image Processing, vol. 15, no.2, pp.473-482, 2005.
6. Er-Wei, Bai, QingYu, Li, Zhiyong ,Zhang, Blind source separation channel equalization of nonlinear channels with binary inputs, IEEE Transactions on Signal Processing, vol. 53, no. 7, pp.2315-2323, 2005.
7. Dezhong Peng, Yong Xiang, Underdetermined blind separation of nonsparse sources using spatial time-frequency distributions, Digital Signal Processing, vol. 20, pp. 581-596, 2010.
8. Fengbo Lu, Zhitao Huang,Wenli Jiang, Underdetermined blind separation of nondisjoint signals in timefrequency domain based on matrix diagonalization, Signal Processing, vol. 91, pp. 1568-1577, January 2011.
9. SangGyun Kim, Chang D. Yoo, Underdetermined Blind Source Separation Based on Subspace Representation, IEEE Transaction on Signal Processing, vol. 57, no. 7, July 2009.
10. A. Cichocki, A.H. Phan, C. Caiafa, FlexibleHALSalgorithms for sparse nonnegative matrix/tensor factorization, In Proc. of 18th IEEE workshops on Machine Learning for Signal Processing, Cancun, Mexico, 16-19, 2008.
11. A. Cichocki, R. Zdunek, and S.-I. Amari. Hierarchical ALS algorithms for nonnegative matrix and 3-D tensor factorization. Springer, Lecture Notes on Computer Science, LNCS-4666, pages 169-176, 2007.
12. Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, Shun-ichi Amari, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, John Wiley, 2009
13. Hichem Snoussi, Jerome Idier. Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures, IEEE Transactions on Signal Processing, vol. 54 no. 9, pp. 3257-3269, 2006.
14. Dezhong Peng, Yong Xiang. Underdetermined blind separation of non-sparse sources using spatial time-frequency distributions, Digital Signal Processing, vol. 20, pp. 581-596, 2010.
15. Shijun Sun, Chenglin Peng, Wensheng Hou, Jun Zheng, Yingtao Jiang, Xiaolin Zheng, Blind source separation with time series variational Bayes expectation maximization algorithm, Digital Signal Processing, vol. 22, pp. 17-33, 2012.
16. Sun, Haojun, Wang, Shengrui, Jiang, Qiangshan, “FCM-based model selection algorithms for determining the number of clusters”, Pattern Recognition, vol. 37 (10), pp. 2027-2037, 2004.
17. A. Cichocki and R. Zdunek, “Multilayer nonnegative matrix factorization”, Electronics Letters, vol. 42(16), pp. 947-948, 2006
18. A. Cichocki and R. Zdunek, “Multilayer nonnegative matrix factorization using projected gradient approaches”, In The 13th International Conference on Neural Information Processing (ICONIP06), Hong Kong, October 3-6 2006