International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 16, 2017, pp 81-92

A New Outlier Rejection Rule for Robust ICA and Its Application to Image Processing

Md. Shahjaman[1] and Md. Nurul Haque Mollah[2]

[1]Md. Shahjaman, Department of Statistics, Begum Rokeya University, Rangpur-5400

Email: shahjaman_brur@yahoo.com

[2]Department of Statistics, Rajshahi University, Bangladesh

[*]Correspondence should be addressed to Md. Nurul Haque Mollah

(Email: mnhmollah@yahoo.co.in)

[Received April 15, 2017; Accepted September 5, 2017]

Abstract

Independent Component Analysis (ICA) is a powerful statistical method for blind source separation (BSS) from the mixture data. It is widely used in signal processing like audio signal processing, image processing, biomedical signal processing as well as processing any time series data. However, most of the ICA algorithms are not robust against outliers. In this paper we propose anew outlier rejection rule for robustification of ICA algorithms using β-weight function. The values of the tuning parameter β play the key role in the performance of the proposed method. A cross validation technique is used as an adaptive selection procedure for the tuning parameter β. The performance of the proposed method is investigated in a comparison of the popular robust Fast ICA algorithms using natural image signals. Simulation and experimental results show that the proposed method improves the performance over the existing robust Fast ICA algorithms.

Keywords: Independent component analysis (ICA), Minimum β-Divergence Estimator, β-selection, β-Weight function, Outliers and Image signals.

AMS Classification: 54H30.

1. Introduction

ICA aims to maximize the non-goussianity or minimize the dependency among the variables as it seeks to recover the sources that are as independent of each other as possible [5]. The independence is a much stronger property than uncorrelatedness [5]. Thus ICA becomes more superior to the principle component analysis (PCA). There are two popular model based robust ICA algorithms (a) FastICA algorithms [4] and (b) minimum β-divergence method [7]. In the case of minimum β-divergence method for ICA, Gaussianity of source signals should be known in advance, otherwise it may produce misleading results. This method suggests two contrast function, where one works to recover sub-Gaussian signals and the other one works to recover super-Gaussian signals. For example, image signals are sub-Gaussian signals and audio signals are super-Gaussian signals. Therefore, in the case of image processing or audio signal processing, minimum β-divergence method may works well. In other cases where source signals are unknown as sub-Gaussian or super-Gaussian, minimum β-divergence method may gives misleading results. On the other hand, robust FastICA algorithm suffers from the non-robust prewhitening procedure and some outliers those make perpendicular direction with recovering vectors, where non-robust prewhitening can be overcome by β-prewhitening [8], but the later one problem exist yet in the robust FastICA algorithms. Therefore, a user or researcher may feel inconvenience to select an appropriate ICA algorithm in some situations. So, in this paper, our proposal is to use outlier rejection rule [10, 11] for robustification of ICA algorithms based on the β-weight function [9] instead of existing robust ICA algorithms.

2. Independent Component Analysis (ICA)

Independent component analysis (ICA) is a statistical tool for revealing hidden factors that underlie sets of random variables or signals. For given set of observations of random variables, $\boldsymbol{x}(t) = \left(x_1(t), x_2(t), \ldots\ldots, x_m(t)\right)^T$, where $t$is the time or sample index, assume that they are generated as a linear mixture of independent components then ICA model is given by:

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \cdot \\ \cdot \\ \cdot \\ x_m(t) \end{pmatrix} = A \begin{pmatrix} s_1(t) \\ s_2(t) \\ \cdot \\ \cdot \\ \cdot \\ s_q(t) \end{pmatrix}$$

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t), \quad t = 1,2,\ldots.,n \tag{2.1}$$

where, A is some unknown matrix. The aim of ICA is to estimate both the matrix A and the $\boldsymbol{s}(t)$, when we only observe the $\boldsymbol{x}(t)$. The ICA of a random vector $\boldsymbol{x}(t)$ consists of finding a linear transform

$$\boldsymbol{y}(t) = W\boldsymbol{x}(t), \quad t = 1,2,\ldots.,n \tag{2.2}$$

so that components of $\boldsymbol{y}(t)$ are as mutually independent as possible, where W transformation matrix obtained by ICA algorithm [1-6]. It is also known as recovering matrix or unmixing matrix or pseudo inverse of A or generalized inverse of A. The components of a random vector $\boldsymbol{y}(t)$are said to be independent of each other if and only if the density function of y is factorized as

$$p(\boldsymbol{y}) = \prod_{i=1}^{m} p_i(y_i)$$

where, $p_i(y_i) = \int p(\boldsymbol{y}), \ldots dy_1, \ldots dy_{i-1}\, dy_{i+1}, \ldots\, dy_m$ is the marginal density of y_i, (i= 1, 2, ...,m). If components of y are independent of each other, then most important property of their independence is

$$E\left\{\prod_{i=1}^{m} h_i(y_i)\right\} = \prod_{i=1}^{m} E\{h_i(y_i)\}$$

where, $h_i(y_i)$ is any measurable function of y_i.

3. Outlier Rejection Rule for ICA Based on *β*-weight Function (New Proposal)

Mahalanobis distance (D^2) is a popular measure for detection of multivariate outliers. It works well in presence of few outliers. However, in presence of large number of outliers, it produces misleading results. To overcome this problem, in this paper we propose β-weight function as a new alternative measure for outlier detection.

This β-weight function is originated from the minimum β-divergence estimators for the mean vector $\boldsymbol{\mu}$ and variance- covariance matrix $\mathbf{V}$

$$\boldsymbol{\mu}_{t+1} = \frac{\sum_{i=1}^{n} \varphi_\beta(\boldsymbol{x}_i|\boldsymbol{\mu}_t, \boldsymbol{V}_t)\boldsymbol{x}_i}{\sum_{i=1}^{n} \varphi_\beta(\boldsymbol{x}_i|\boldsymbol{\mu}_t, \boldsymbol{V}_t)} \text{ and}$$

$$\boldsymbol{V}_{t+1} = \frac{\sum_{i=1}^{n} \varphi_\beta(\boldsymbol{x}_i|\boldsymbol{\mu}_t, \boldsymbol{V}_t)(\boldsymbol{x}_i - \boldsymbol{\mu}_t)(\boldsymbol{x}_i - \boldsymbol{\mu}_t)^T}{(1+\beta)^{-1}\sum_{i=1}^{n} \varphi_\beta(\boldsymbol{x}_i|\boldsymbol{\mu}_t, \boldsymbol{V}_t)}$$

$$\text{where, } \varphi_\beta(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{v}) = \exp\left\{-\frac{\beta}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

which is known a β-weight function [9]. It produces smaller weight for each contaminated data vector and larger weight for each uncontaminated data vector. Our intention is to use this weight function to separate the data into two parts bad (outliers/unusual) data points and good (usual) data points. To detect outliers, we compute the β-weight as follows:

$$\varphi_\beta(\boldsymbol{x}|\hat{\mu}_\beta, \hat{V}_\beta) = exp\left\{-\frac{\beta}{2}(\boldsymbol{x}-\hat{\mu}_\beta)^T \hat{V}_\beta^{-1}(\boldsymbol{x}-\hat{\mu}_\beta)\right\}$$

and then we construct a criteria to test the contaminacy of a data vector as follows:

$$\varphi_\beta(\boldsymbol{x}|\hat{\mu}_\beta, \hat{V}_\beta) = \begin{cases} larger\ (but \le 1), & if\ \boldsymbol{x}\ is\ not\ contaminated \\ smaller\ (but \ge 0), & if\ \boldsymbol{x}\ is\ contaminated \end{cases} \tag{2.4}$$

Using the proposed test criteria, we take the decision that a data vector $\boldsymbol{x}$ is said to be contaminated by outliers if

$$\varphi_\beta(\boldsymbol{x}|\hat{\mu}_\beta, \hat{V}_\beta) \le \delta,$$

where we choose the threshold value of $\delta$by

$$\delta\text{=(1-η) min}\varphi_\beta(\boldsymbol{x}|\hat{\mu}_\beta, \hat{V}_\beta) + \text{η max}\, \varphi_\beta(\boldsymbol{x}|\hat{\mu}_\beta, \hat{V}_\beta)$$

$$x\in\mathcal{D}\mathrm{S}\,\boldsymbol{x} \in \mathcal{D}\boldsymbol{S}$$

with heuristically η = 0.10, where $\mathcal{D}$S is the dataset. It was also used in [7,9] for choosing the threshold value. Then we reject or remove the contaminated data points from the dataset and we can apply any ICA algorithms to the clean dataset to recover source signals from the robustness points of view. In this paper we consider Fast ICA algorithm to demonstrate the performance of the proposed rejection rule in a comparison of the classical Mahalanobis distance approach.

4. Simulation Study

To demonstrate the performance of the proposed method in a comparison of some existing methods, we consider the following synthetic data sets.

1. Two-dimensional 1000 random samples were drawn from uniform distribution with mean zero and variance one. Figure 1a represents the scatter plot of this source data points. Then we mixed these data points by a random mixing matrix

$$A = \begin{bmatrix} 1.01 & 0.8 \\ 0.4 & 0.9 \end{bmatrix};$$

Figure 1b represents the scatter plot of this mixed data points.

2. 700 outliers (*) are added to data set 1to make 1700 samples in total. For convenience of presentation, we took last 700 observations as outliers out of 1700. Figure 2a represents the scatter plot of this data set.

Let us consider data set 1 described above. To remove outliers, first we select tuning parameter β using cross validation [8]. Figure 1c shows the plots of $\hat{D}_{\beta_0}(\beta)$. In this plot asterisks (*) are $\hat{D}_{\beta_0}(\beta)$ and a circle outside the asterisk indicates the smallest value. Dotted lines are $\hat{D}_{\beta_0}(\beta) \pm SD_{\beta_0}$. Plot of $\hat{D}_{\beta_0}(\beta)$ shown in the Figure 1c suggest $\beta = 0$ for $\beta_0 = 0.5$ by "One Standard Error" rule. Thus adaptive selection procedure suggests there is no outliers in the data set 1. Figure 1d shows the scatter plot of recovered sources by FastICA. Comparing Figures 1a and 1d, we see that recovered sources are independent with each other with non-Gaussian Structure.

Simulation with Uniformly Distributed Datasetin Absence of Outliers

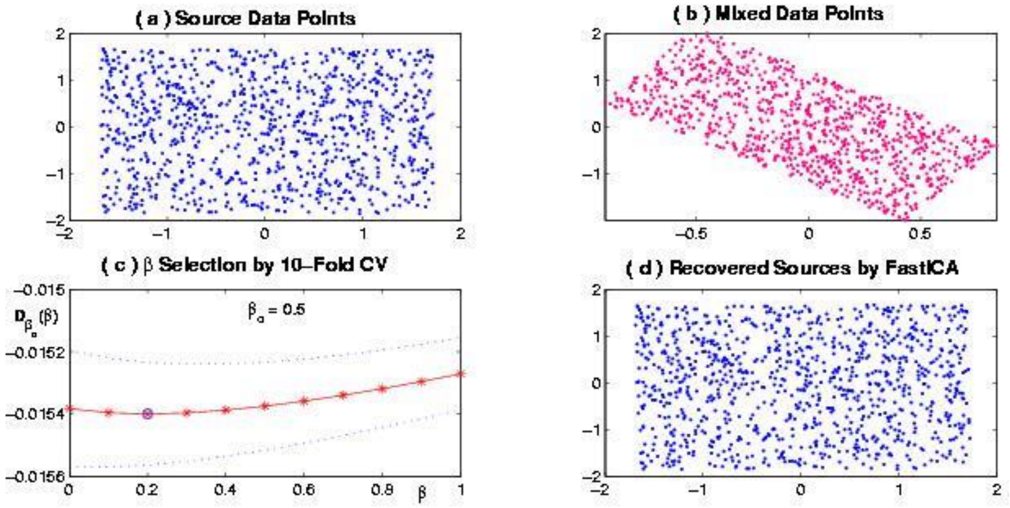

Figure 1: (a) Scatter plot of source signals generated from uniform distribution, (b) Scatter plot of mixed signals, (c) Plots of $\hat{D}_{\beta_0}(\beta)$ for selection of β , (d) Recovered Sources by Fast ICA.

To investigate the performance of the proposed method in presence of outliers (*), we consider data set 2 shown in figure 2a. To remove outliers by the proposed method, we select the values of the tuning parameter β by *K*-fold CV (k=10) as before. We computed $\hat{D}_{\beta_0}(\beta)$ for β varying from 0 to 1 by 0.5 with β_0 = 0.5. Figure 2b show the plots of $\hat{D}_{\beta_0}(\beta)$. In this plot the asterisk (*) are $\hat{D}_{\beta_0}(\beta)$ and the smallest value is indicated by a circle outside the asterisk. Dotted lines are $\hat{D}_{\beta_0}(\beta) \pm SD_{\beta_0}$ .Plots of $\hat{D}_{\beta_0}(\beta)$ shown in the Figure 2b have elbow shape and suggested β = 0.8 for β_0 = 0.5 by ”One Standard Error” rule. Thus adaptive selection procedure suggests that outliers corrupt the data set 2, which is true as shown in figure 2a. Figure 2c shows the recovered sources by Fast ICA using dataset 2 before removing the outliers. Clearly we see that recovered sources are not similar to the original sources. Then we remove outliers from dataset 2 using Mahalanobis distance. Figure 2d shows Mahalanobis D^2 for each data points. Figure 2e shows the mixed data after removing the outliers by Mahalanobis D^2

with $\chi^2_{1,0.95} = 3.84$. Clearly we see that all outliers are not removed based on Mahalanobis distance. Figure 2f shows the recovered sources by FastICA after removing outliers by Mahalanobis D^2. Clearly we see that recovered sources are not similar to the original sources as shown in 1a.

Simulation with Uniformly Distributed Dataset in Presence of Outliers

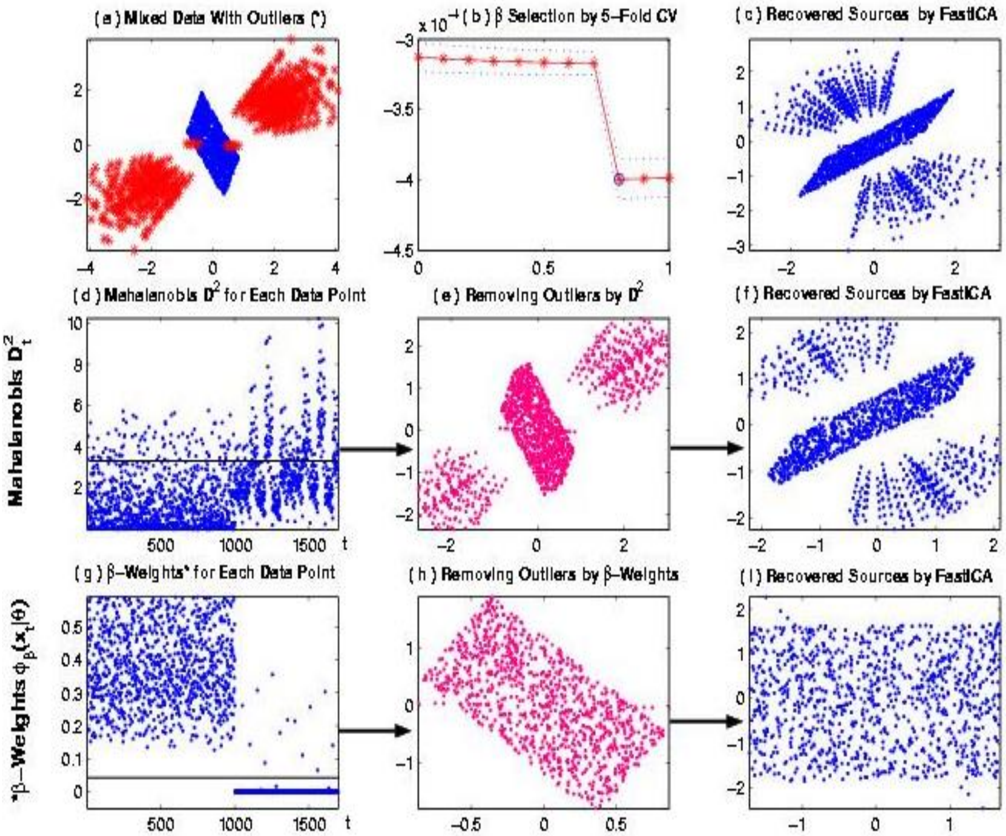

Figure 2: (a) Scatter plot of mixed signals with outliers (*), (b) Plots of $\hat{D}_{\beta_0}(\beta)$ for selection of β ,(c) Recovered Sources by FastICA before removing outliers, (d) Mahalanobis distance for each data points with cut-off point $\chi^2_{1,0.95} = 3.84$, (e) Mixed data after removing outliers using Mahalanobis distance, (f) Recovered sources by FastICA after removing outliers by Mahalanobis distance, (g)β -weight for each data point, (h) Mixed data after removing outliers using β-weight

function , (i)Recovered sources by FastICA after removing outliers by β-weight function.

Then we remove outliers from data set 2 using our proposed method. The Figure 2g shows the β-weight for each data points. Figure 2h shows the mixed data points after removing outliers by proposed method. Clearly we see that almost all outliers are removed by our proposed method. Figure 2i shows the recovered sources by FastICA after removing outliers by our proposed method. Clearly we see that recovered sources are similar to the originals sources as shown in 1a.

4.1 Images Processing

To demonstrate the performance of the proposed method for image processing, we considered two 256×256 pixels original images of flower and Gaussian noise as shown in figure (3a-3b), respectively. Figure 3c represents the scatter plot of these two original images. Then we mixed these two original images using the linear ICA model. Figures (3d-3e) represent the mixture of flower image and Gaussian noise image respectively. Figure 3f represents the scatter plot of these two mixed images. To recover original images from the mixture, we first apply well known FastICA algorithm. Figure (3g-3h) shows the recovered images of flower and Gaussian noise, respectively by FastICA. Figure 3i represents the scatter plot of these two separated images. Then we apply FastICA algorithm to recover the original images using our proposed method again from the same mixture. Figure (3j-3k) shows the recovered images of flower and Gaussian noise, respectively by our proposed method. In absence of outliers, clearly we see that performance of robust FastICA and our proposed method based FastICA is good and recovers almost similar images.

Image Source Separation in Absence of Outliers

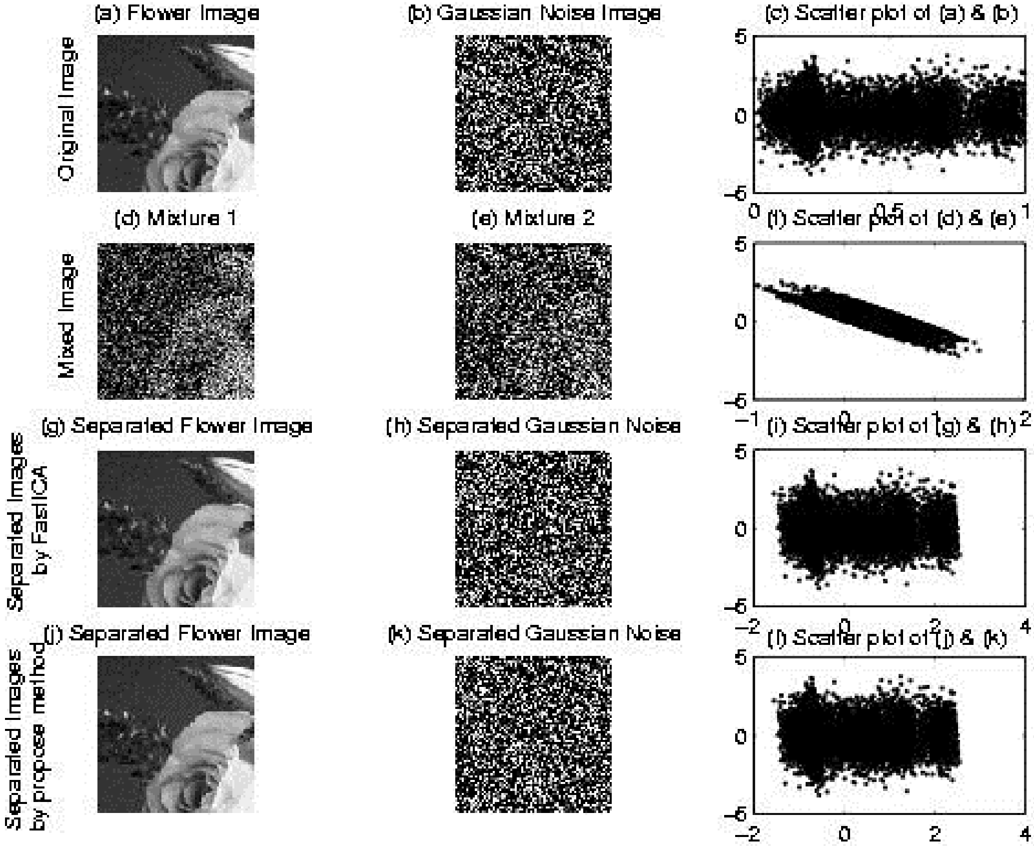

Figure 3: (a) Original flower image, (b) Original Gaussian noise image, (c) Scatter plot of original images, (d) Mixture of flower image, (e) Mixture of Gaussian noise image, (f) Scatter plot of two mixed images, (g) Flower image recovered from the mixed images by FastICA, (h) Gaussian noisy image recovered from the mixed images by FastICA, (i) Scatter plot of two recovered images by FastICA, (j) Flower image recovered from the mixed images by proposed method, (k) Gaussian noisy image recovered from the mixed images by proposed method, (l) Scatter plot of two recovered images by proposed method.

To investigate the performance of FastICA in presence of outliers in a comparison of Mahalanobis distance, we added 256×6 pixel values with each mixed images from Gaussian distribution as outliers. Figures (4a-4b) represent the mixture of flower image and Gaussian noise image in presence of outlying pixel values (+), respectively. Figure 4c shows scatter plot of two mixed images in presence of outliers. To recover original images from the mixture in presence of outliers, we first apply FastICA algorithm after removing outliers by Mahalanobis distance. Figure (4d-4e) shows the recovered images by MahalanobisD^2. Figure 4f shows scatter plot of two recovered images by Mahalanobis distance. Clearly we see that performance of FastICA using Mahalanobis distance is not good in presence of outliers and recovered images cannot recognize the original images of flower and Gaussian noise. Then we apply the FastICA algorithm after removing outliers by our proposed method with $\beta = 0.1$ to recover the original images again from the same mixture. Figure (4g-4h) shows the recovered images of flower and Gaussian noise, respectively. Obviously, it is seen that the performance of robust FastICA using our proposed method is good and recovered images can easily recognize the original image of flower and Gaussian noise.

Image Source Separation in Presence of Outliers

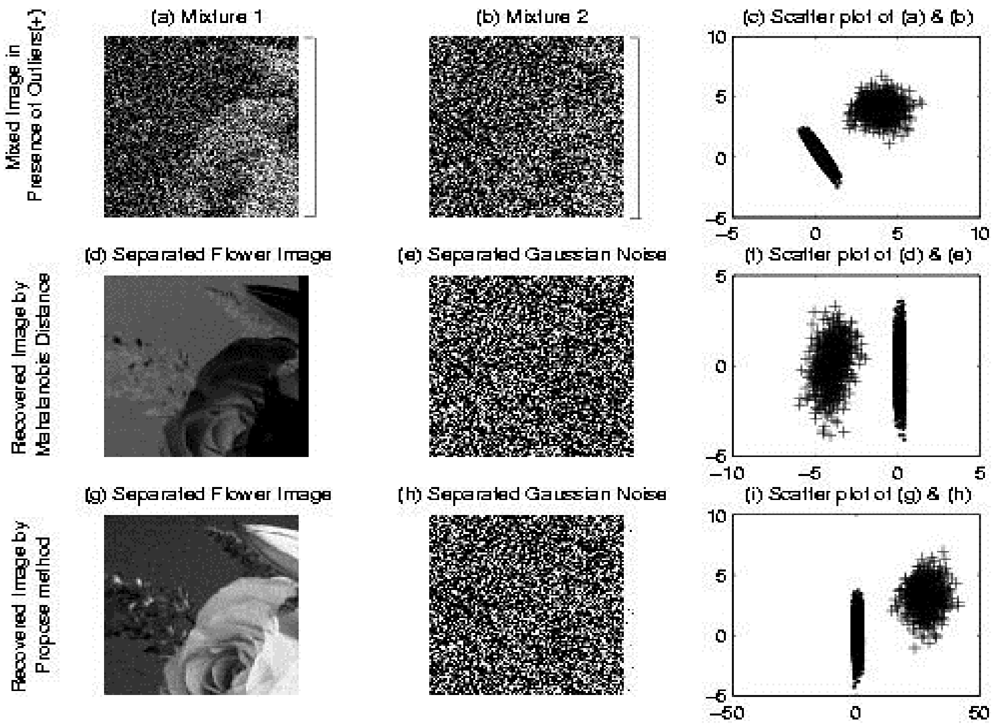

Figure 4. (a) Mixture 1 with 256×6 pixels outliers, (b) Mixture 2 with 256×6 pixels outliers, (c) Scatter plot of two mixed images in presence of outliers (+), (d) Flower image recovered by FastICA using MahalanobisD^2, (e) Gaussian noise image recovered by FastICA using MahalanobisD^2, (f) Scatter plot of two recovered images by FastICA using MahalanobisD^2, (g) Flower image recovered from the mixed images by FastICA using proposed method, (h) Gaussian noisy image recovered from the mixed images by FastICA using proposed method, (i) Scatter plot of two recovered images by FastICA using proposed method.

5. Conclusion

The two popular robust ICA algorithms are minimum β divergence method [7] and FastICA algorithm [5]. But when the data contains outlier they do not work well as usual. In this paper we discuss the robustification of ICA using an outlier rejection rule [10]-[11]based on β-weight function [9]. The values of the tuning parameter β play the key role in the performance of the proposed method. A cross validation technique is discussed as an adaptive selection procedure for the tuning parameter β[9]. If adaptive selection procedures produce $\beta > 0$, then data set is corrupted by outliers. If adaptive selection procedure produce $\beta = 0$, then data set is not corrupted by outliers. Both simulation and real image data results show that FastICA algorithm is able to recover all hidden signals properly after removing outliers by our proposed method. But FastICA algorithm is not able to recover all hidden signals properly after removing outliers by Mahalanobis distance.

References

[1] Cichoki, A. and Amari, S. (2002). Adaptive Blind Signal and Image Processing. Wiley, New York.

[2] Comon, P. (1994). Independent component analysis - a ew concept? Signal Processing,Vol.36, pp. 287-314.

[3] Hyvärinen , A. (1999). Fast and robust fixed-pointalgorithms for independent component analysis. IEEE Trans. on Neural Network, 10(3), 626-34.

[4] Hyvärinen , A. Karunen, J. and Oja, E. (2001). Independent Component Analysis, New York: Wiley.

[5] Lee, T.-W. (2001). Independent Component Analysis: Theory and applications, Kluwer Academic Publishers.

[6] Minami, M. and Eguchi, S. (2002). Robust Blind Sourceeparation by beta-Divergence. Neural Computation 14, 1859-1886.

[7] Mollah, M.N.H., Minami, M. and Eguchi, S. (2006). Exploring latent structure of mixture ICA models by the minimum β-divergence method, Neural Computation, 18(1), pp.166-190.

[8] Mollah, M.N.H., Minami, M. and Eguchi, S. (2007). Robust prewhitening for ICA by minimizing β-divergence and its application to FastICA. Neural Processing Letters, 25(2), 91-110.

[9] Mollah, M.N.H., Sultana, N., Minami, M. and Eguchi, S., Robust Extraction of Local Structures by the Minimum β-Divergence method. Neural Network, Vol. 23, (2010) , pp. 226-238.

[10] Shahjaman, Md. and Mollah, M.N.H., (2010). Arobustification of Independent Component Analysis and Its Application, The First International Conference on Theory and Applications of Statistics, 2010 (DUSDAA), abstract booklet, pp.82.

[11] Shahjaman, Md. and Mollah, M.N.H., (2010). OutlierRejection Rule for Independent Component Analysis and Its Application, 13th International Conference on Computer and Information Technology, 2010 (ICCIT2010), program book, pp.72.