Blind noisy image separation based on a new robust independent component analysis network

Jian Ji (纪 建)[1] and Zheng Tian (田 铮)[2,3]

[1]Department of Computer Science & Technology, Northwestern Polytechnical University, Xi'an 710072

[2]Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710072

[3]Key Laboratory of Education Ministry for Image Processing and Intelligent Control, Huazhong University of Science & Technology, Wuhan 430074

Received May 23, 2006

The separation of noisy image is a very exciting area of research, especially when no prior information is available about the noisy image. In this paper, we propose a robust independent component analysis (ICA) network for separation images contaminated with high-level additive noise or outliers. We reduce the power of additive noise by adding outlier rejection rule in ICA. Extensive computer simulations confirm robustness and the excellent performance of the resulting algorithms.

OCIS codes: 100.0100, 100.3020, 200.4260.

Blind source separation (BSS) is an important field of research in signal processing and data analysis[1]. Independent component analysis (ICA) is one solution to the problem. The meaning underlying ICA is to separate the original signals into several independent components by selecting a particular criterion and an optimal algorithm. Though the basic knowledge of the sources and the transmitting channels is unknown, the probability to realize this decomposition lies in the principle of statistical independence.

Let us assume that an array of sensors provides a vector of n observed signals $\mathbf{x} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$ that are linear mixtures of $n \geq m$ unobserved random processes $\mathbf{s} = [s_1, s_2, \cdots, s_m]^{\mathrm{T}}$ sources. The basic problem of ICA is defined for the noiseless case, where the sources and observations have the following linear relation[2]

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

where A is an unknown $n \times m$ full-column rank matrix that represents the mixing system.

However, in practical applications, the observed signals tend to be the multidimensional mixture of independent sources and are always corrupted by noise. In these cases, we cannot ignore noise. Therefore, Eq. (1) is not enough to describe the problem. Then we discuss the case where there is additive noise in observations as

$$\mathbf{x} = A\mathbf{s} + \mathbf{e}, \tag{2}$$

here $\mathbf{e} = [e_1, e_2, \cdots, e_n]^{\mathrm{T}}$ is the vector of noise components which are assumed to be Gaussian and statistically independent of the sources.

In order to recover the sources, the observations are processed by a $m \times n$ separating matrix B to produce the vector of outputs or sources estimation

$$\mathbf{y} = B\mathbf{x}. \tag{3}$$

There are a number of efficient adaptive, on-line learning algorithms that have been developed for ICA[3,4]. Although the underlying principles and approaches are different, many of the techniques have very similar forms. Most of these algorithms assume that any measurement noise within the mixed signals can be neglected. Thus, the problem of efficiently reducing the influence of noise on the performance of algorithms for ICA arises, and in particular, methods are desired to reduce noise in the stochastically independent extracted components.

In this paper, we propose a robust ICA network to blind separate images from noisy mixtures. We believe that a way to obtain robust estimates for both the mixing matrix and the sources is to remove the worst outliers in a pre-processing step, before running the ICA algorithm. At the same time we want to be careful not to remove too many points. After a pre-processing stage means of PCA, we remove outliers by applying outlier rejection rules. Then we apply the ICA method on the clean data set. We experimentally verified that the robust ICA network enables us to increase robustness against noise.

Figure 1 shows the three-layer robust ICA network, where the first layer performs pre-whitening (sphering), the second layer is flag noisy using rejection rule, and the third one is that separation of sources uses existing ICA algorithms. We describe these three main steps in detail as follows.

1) Pre-whitening: In ICA algorithms, the data vectors $\mathbf{x}$ are often preprocessed by whitening (sphering): $\mathbf{v} = V\mathbf{x}$. Here $\mathbf{v}$ denotes the whitened vector satisfying $E[\mathbf{v}\mathbf{v}^{\mathrm{T}}] = I$, where I is the unit matrix, and V is a $m \times n$ whitening matrix. In general, principal components

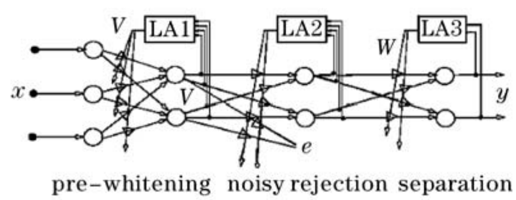

Fig. 1. Three-layer robust ICA network for pre-whitening, noisy rejection, and blind separation.

analysis (PCA) is used to do whitening. After prewhitening the subsequent $m \times m$ separating matrix W can be taken orthogonal, which often improves the convergence. Thus in whitening approaches the total separating matrix is $B = WV$.

2) Rejection rules for remove outliers: All of the ICA methods assume that the mixed data are homogeneous, that is free of outliers which is not held in most cases. Data often (always) contain outliers. The contaminated data affect the ICA algorithm and result in an estimated sources, which is different from the original[5]. In order to avoid this problem, we introduce an outlier detection and elimination process. ICA techniques are then applied to the cleaned data set to obtain the estimated sources.

We will try two rules for flagging outliers in the raw data[6]. They are based on different distances or outlying measures computed at each data point. The corresponding rejection rule then flags all points whose outlyingness exceeds a certain cutoff value.

First the data points of the data matrix x_i are projected on a subspace defined by means of a measure of outlyingness. This measure is obtained by projecting the data points on many univariate directions z. For every direction a robust center and scale of the projected data points $x_i'z$ are computed, namely the univariate minimum covariance determinant (MCD) estimator[7] of location $\hat{\mu}^i_{\text{MCD}}$ and scale $\hat{\sigma}^i_{\text{MCD}}$. The outlyingness of a data point x_i is then measured by

$$\text{outl}(x_i) = \max_{z \in B} \frac{|x_i'z - \hat{\mu}^i_{\text{MCD}}|}{\hat{\sigma}^i_{\text{MCD}}}, \tag{4}$$

where B contains all directions (unit length vectors) we search over. Then we obtain a subspace with the smallest outlyinhness that fits the data well. We project the data points on this subspace where we robustly estimate their locations and their scatter matrixes by means of the MCD estimator, of which we compute its m non-zero eigenvalues $l_1, \cdots, l_m$. The corresponding eigenvectors are the m robust principal components. Formally, writing the (column) robust eigenvectors next to each other yields the $n \times m$ matrix P with orthogonal columns. The location estimate is denoted as the column vector $\hat{\mu}$ and called the robust center. Thus, projecting the observations onto this subspace yields the scores t_i, it satisfies

$$t_i = (x_i - \hat{\mu}')P. \tag{5}$$

To distinguish between regular observations and the outliers, we take into account the orthogonal distance OD_i of each observation to the PCA space

$$\text{OD}_i = \|x_i - \hat{\mu} - Pt_i'\|. \tag{6}$$

The first rejection rule flags all points whose robust distance OD_i exceeds a cutoff value.

We also consider the score distance SD_i which represents the distance inside the PCA space taking into account the covariance structure of the data. More formally this distance is defined by

$$\text{SD}_i = \sqrt{t_i^{\text{T}} L^{-1} t_i}, \tag{7}$$

where L is the diagonal matrix with the eigenvalues $l_1, \cdots, l_m$. The corresponding rejection rule flags all points whose outlyingness SD_i exceeds a cutoff.

We can distinguish between four types of observations as follows: regular data (with small SD and small OD) form one homogeneous group that is close to the PCA space; good PCA-leverage points (with large SD and small OD) that lie close to the PCA space but far from the regular observations; orthogonal outliers (with small SD and large OD) whose orthogonal distance to the PCA space is large but which we cannot see when we only look at their projection on the PCA space; and bad PCA-leverage points (with large SD and large OD) that have a large orthogonal distance and whose projection on the PCA subspace is remote from the typical projections. So we can flag three types of outliers and keep the regular data.

3) ICA algorithms for separation: After rejecting outliers in the raw data, we apply existing ICA algorithms for blind separation. In this paper, we use EICA and FastICA algorithms. Equivariant (EICA) algorithm is a quasi-Newton iteration that will converge to a saddle point with locally isotropic convergence, regardless of the distributions of sources. It has the following equivariant and robust in respect to Gaussian noise algorithm[8]

$$\begin{aligned} \Delta B(l) &= B(l+1) - B(l) \\ &= \eta_l [I - C_{1,q}(y,y) S_{q+1}(y)] B(l), \end{aligned} \tag{8}$$

where $S_{q+1}(y) = \text{sign}(\text{diag}(C_{1,q}(y,y)))$, $C_{p,q}(y,y)$ denotes the cross-cumulant matrix whose elements are $[C_{p,q}(y,y)]_{ij} = \text{Cum}(\underbrace{y_i \quad \cdots \quad y_i}_{p}, \underbrace{y_j \quad \cdots \quad y_j}_{q})$.

One iteration of the generalised fast fixed-point (FastICA) algorithm for finding a row vector $\mathbf{w}_i^{\text{T}}$ of W is[9]

$$\mathbf{w}_i^* = \mathbf{E}\{\mathbf{v} g(\mathbf{w}_i^{\text{T}} \mathbf{v})\} - \mathbf{E}\{g'(\mathbf{w}_i^{\text{T}} \mathbf{v})\} \mathbf{w}_i, \mathbf{w}_i = \mathbf{w}_i^* / \|\mathbf{w}_i^*\|, \tag{9}$$

here $g(t)$ is a suitable nonlinearity, typically $g(t) = t^3$ or $g(t) = \tanh(t)$, and $g'(t)$ is its derivative.

Computer simulations were carried out to verify the performance of the proposed algorithm. Source images include four gray-level figures[10]: Lenna, Baboon, Mandel, and a noise (Fig. 2(a)). The size of each image is 128×128. We add 10-dB Gaussian white noise to the data. Then four noisy mixture images (Fig. 2(b)) were generated from source images.

Now we use robust ICA network to separate noisy mixing images. In order to verify the performance of the proposed algorithm, we separate images using ERICA algorithm, ERICA algorithm with reject rules, FastICA algorithm and FastICA algorithm with reject rules respectively. Figure 2(c) is the results only using ERICA algorithm, Fig. 2(d) is the results using ERICA algorithm with reject rules, Fig. 2(e) is the results only using FastICA algorithm, and Fig. 2(f) is the results using FastICA algorithm with reject rules.

We use the inaccuracy measure[11] to measure whether unmixing matrix B has done a good job

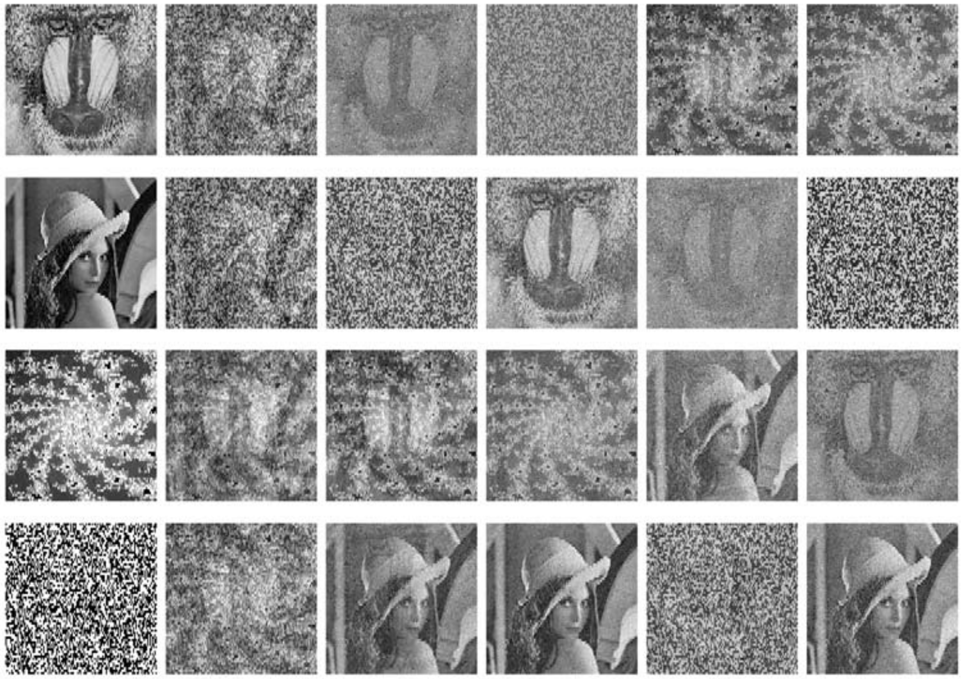

Fig. 2. (a) Source images $\mathbf{s}$; (b) noisy mixtured images $\mathbf{x}$; (c) only use EICA; (d) use EICA with rejection rules for remove outliers; (e) only use FastICA; (f) use FastICA with rejection rules for remove outliers.

$$\text{INACC} = \frac{\sum_{i=1}^{M}\left(\sum_{j=1}^{m}\frac{|\mathbf{Q}_{ij}|}{\max_k|\mathbf{Q}_{ik}|}-1\right)+\sum_{j=1}^{m}\left(\sum_{i=1}^{m}\frac{|\mathbf{Q}_{ij}|}{\max_k|\mathbf{Q}_{kj}|}-1\right)}{2m(m-1)}, \tag{10}$$

where $Q = BA = (Q_{ij})_{i,j=1,\cdots,m}$. In the ideal case, INACC = 0. At the other extreme, the worst case is when all $|Q_{ij}|$ are equal, and then INACC = 1. The ability of separation with different algorithms is listed as follows. The INACC for the separation images of only using RICA and FastICA are found to be 17.78% and 17.72%, respectively. But the INACC for the separation images of use RICA and FastICA with reject rules are found to be 4.91% and 5.75%, respectively. We note that results of the two separations with reject rules have lower INACC than results of only using ICA.

A robust ICA network for blind images separation is proposed under the conditions that the sensor signals are contaminated with a high-level additive noise and outliers. The method is to preprocess the data by rejecting outliers based on orthogonal distance and score distance outlyingness measure, using a high enough cutoff value. As always, it is good to compare the ICA result with the robustified one.

Although the outlier sensitivity of ICA algorithms and rejection rules for remove outliers were demonstrated in this paper, two important considerations must be kept in mind. First, the ICA algorithms were designed to perform optimally for the cases considered, and may not be optimal for other cases. Second, the computational complexity of the method requires considerate amount of time. The interplay between the efficiency of computational processing and the sensitivity to outliers should be investigated further to improve the evaluation of the outlier robustness of ICA algorithms.

This work was supported by the National Natural Science Foundation of China (No. 60375003), the Aeronautics and Astronautics Basal Science Foundation of China (No. 03I53059). J. Jian' e-mail address is jijiangao@126.com.

References

1. X. Chai, C. Zhou, Z. Feng, Y. Wang, and Y. Zuo, Chin. Opt. Lett. **4,** 11 (2006).
2. P. Comon, Signal Processing **36,** 287 (1994).
3. S. Amari, A. Cichocki, and H. H. Yang, *Proc. Int. Symp. on Nonlinear Theory and its Applications* 37 (1995).
4. S. Amari, A. Cichocki and H. H. Yang, *Advances in Neural Information Processing Systems* 757 (1996).
5. S. Pandey, N. Billor, and A. Turkmen, *Int. Conf. on Statistics, Combinatorics, Mathematics and Applications* (2005).
6. M. Hubert, P. J. Rousseeuw, and K. Vanden Branden, Technometrics **47,** 64 (2005).
7. P. J. Rousseeuw and K. Van Driessen, Technometrics **41,** 212 (1999).
8. S. Cruces, L. Castedo, and A. Cichocki, Neurocomputing **49,** 87 (2002).
9. A. Hyvarinen, IEEE Trans. NN **10,** 626 (1999).
10. A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing* (John Wiley & Sons, Cichester, 2002).
11. X. Giannakopoulos, J. Karhunen, and E. Oja, Int. J. Neural Systems **9,** 99 (1999).