DIGITAL IMAGE PROCESSING TECHNIQUES FOR DETECTION AND REMOVAL OF NOISE IN IMAGES IMPLEMENTING BLIND SOURCE SEPARATION

K Devaki Devi

M.Tech. – ECE,
Gudlavalleru Engineering College,
Gudlavalleru.
e-mail ID: devakikoduru@yahoo.com
Cell No: 9441362153

Dr GN Swamy

Professor
Dept. of ECE
Gudlavalleru Engineering College
Gudlavalleru
e-mail ID: gavini_s@yahoo.com
Cell No:09849026119

Abstract

In different fields, the problem of Blind Source Separation (BSS) is known as a collection of linear combinations of unknown sources and worse the coefficients of the linear combinations that are unknown. The main problem is to estimate the matrix of the combination coefficients (mixing matrix) and to reconstruct the sources according to it. The quality of separation of sources from mixtures dramatically is done by exploiting the quality of sparsity of sources, whereby the sources are properly represented according to some collection of signals. There are two main approaches to the "blind source separation" problem solution they are clustering and ICA (independent component analysis).
Keywords: BSS, Sparsity, Mixing Matrix, ICA

1.Introduction.

The "Blind Source Separation" (BSS) problem of a vector of N mixtures $X(\xi)$ created by M unknown sources $S(\xi)$ by an N*M sized mixing matrix A that is also unknown can be formulated by the equation:

$$X(\xi) = AS(\xi) + N(\xi) \qquad 1.1$$

when $N(\xi)$ is an additive noise that may be present in the problem. From here and until stated otherwise we will assume that $N(\xi) = 0$ (no noise). The coordinate ξ can be a time coordinate (for audio signals) or a space coordinate (for images**)**.

2. ICA

One possible approach is to assume the independence of the sources with a method called Independent Component Analysis (ICA).

Another assumption is the sources' sparsity assumption when there are properly represented according to a group of functions $\{\varphi_k(\xi)\}$, meaning:

$$s_m = \sum_k c_{mk}\varphi_k(\xi) \qquad 2.1$$

The functions $\varphi_k(\xi)$ do not have to be linearly independent and can form an over complete set of functions like a Wavelets family. Assuming sparsity gives even better performance than standard ICA and even allows more sources than mixtures.

2.1 ICA based Algorithms

ICA is presented in Three forms based on the coordinates. The first one integrates clustering and ICA algorithms and the other two are based only on ICA. For simplicity the algorithms are assumed with equal number of mixtures and sources. Thus the problem of estimating the number of sources is bypassed. However the ICA approach allows dealing with the case of the number of mixtures smaller than the number of sources also.

2.1.1 ICA Algorithm 1

This algorithm applies at the first stage the clustering algorithm . After this stage is finished , instead of separating the sources according to the centers of the clusters returned by the FCM algorithm, the unmixing matrix directly according to the Maximum Likelihood criterion is estimated, by finding the minimum of the function:

$$L_W(Y) = k*\ln(|\det W|) - \sum_{m=1..M}\sum_{k=1..K} v((WY)_{mk}) \qquad 2.2$$

This formula is received under the assumption of the independence of sources (the ICA assumption). The probability density function for the coefficients c_{mk} is modeled as:

$$\rho(c_{mk}) \propto \exp\{-v(c_{mk})\} \qquad 2.3$$

when v(x) is the absolute value function |x| with smoothing.

The matrix Y appearing in the expression $L_W(Y)$ is the data matrix when in each row there are the DWP coefficients of one of the mixtures. In addition to that, k is the number of coefficients (called the number of features), and M=N is the number of sources and also the number of mixtures.

The maximization of $L_W(Y)$ is performed effectively by the Natural Gradient algorithm. Using the estimated W the sources are exactly reconstructed as in the case of square A in the clustering based algorithm.

2.1.2 ICA Algorithm 2

This algorithm starts with building the tree of DWP coefficients of mixtures exactly. After that a pass on all the subsets in the tree is performed and for each subset the following stages are performed:

1. Estimating the unmixing matrix W according to the ICA algorithm based on the current subset coefficients only.
2. After that the sources are reconstructed using the unmixing matrix of the previous stage. To these sources the DWP transform is applied, and their entropy is calculated in the Wavelet Transform domain. The entropy of a series of coefficients “coefficient” is:
   $\text{Entropy} = \sum_{i=1..length(coeff)}[|coeff[i]|]^{0.5}$ 2.4

   Normalizing the coefficients to have a square norm of 1:

   $[\sum_{I=1..LENGTH(COEFF)} |COEFF[I]|^2]^{0.5} = 1$ 2.5

   The entropy of the sources is calculated as the sum of entropies of each source separately. Finally, the “best” set as the set with the minimal entropy is chosen. This best set is chosen for reconstruction.

   2.1.3 ICA Algorithm 3:

   With the difference, after the tree of DWP coefficients is built sources are reconstructed according to all the coefficients .The reconstructed sources are transformed to the Wavelet Transform domain. Then, they are passed on all the sets in the tree and for each set the entropy of its coefficients alone on all the sources is calculated.

Finally the set chosen is the one with minimal entropy. Based on this set, the final reconstruction of the sources is calculated.

3. Clustering

Another approach to the sources separation, again under the assumption of sparsity is the clustering approach, to separate the coefficient data, generated after using a transform on the mixtures into several clusters where cluster is a group of condensed points. The data to separate will be collection of points of the form (d_{1k}, d_{2k}, ..., d_{Nk}) when N is the number of mixtures and k runs on the indices of the family of functions $\varphi_k(\xi)$.

for example in the simple case of two sources and two mixtures when the sources are already sparse in the original domain.

$$x_1 = a_{11}s_1(t) + a_{12}s_2(t)$$
$$x_2 = a_{21}s_1(t) + a_{22}s_2(t) \qquad 3.1$$

Since the sources are sparse in most cases when $s_1(t)$ is different than 0, $s_2(t)$ is equal to 0 and therefore:

$$x_1 = a_{11}s_1(t)$$
$$x_2 = a_{21}s_1(t) \qquad 3.2$$

Plotting the values of $x_1(t)$ v.s. the values of $x_2(t)$ will give straight line passing through the origin of the axes with a slope of a_{11}/a_{21}. Similarly when $s_2(t)$ is different from 0 and $s_1(t)$ is 0 (or close enough to 0) we get:

$$x_1 = a_{12}s_2(t)$$
$$x_2 = a_{22}s_2(t) \qquad 3.3$$

that in the same way is a description of a straight line passing through the origin of the axes with the slope of a_{12}/a_{22}. After finding the two slopes, we can reconstruct the mixing matrix A as it appears in the equation:

$$X = AS \qquad 3.4$$

when

$$X = (x_1(t)\ x_2(t))^T \qquad 3.5$$

and

$$S = (s_1(t)\ s_2(t))^T \qquad 3.6$$

assuming the mixing matrix is normalized, for example each column norm is 1.

As it turns out the sources are rarely sparse in their original domain, but their decomposition coefficients according to a set of functions that is wisely chosen are usually sparse. Thus, the mixtures are not plotted by themselves, but the coefficients produced by a proper transform activated on them. Also it is clear that the lines obtained, not always will be clear ones and therefore an algorithm is needed to estimate their slopes.

4. Clustering Based Algorithm:

Before using this algorithm the Discrete Wavelet Packet (DWP) transform is applied to the mixtures received as input. As a result a tree of subsets of the transform coefficients is obtained.

The following algorithm is applied to all the subsets of coefficients including the set of all the coefficients together.

4.1 The Clustering Algorithm :

1. All the coefficients of the subset for all the mixtures in the form of points in the mixtures space are taken and casted on the upper half of an N dimensional sphere with a radius of 1 (N is the number of mixtures).

2. The Fuzzy C-Means (FCM) clustering algorithm is executed on the resulting data with the number of clustering changing from 1 to a predefined maximum .
3. The output of the FCM algorithm is the centers of the clusters the data was separated.
4. Based on the clusters, centers from the previous stage optimal number of clusters are found and then the global distance measure for each subset are calculated:

{GLOBAL DISTANCE} =
{$\sum_{ALL\ POINTS}$ distance from center of point's cluster} / {num of points} 4.1

After those three stages is completed the "best set" is chosen according to minimum global distance. The assumption for the chosen set is that all the sources are "present" in it.

For the chosen set the clusters' centers are chosen and build from them the estimated mixing matrix when each center is one of its columns. Using that matrix the sources from mixtures are separated. In order to describe, the 2 cases are:

1. the number of mixtures is equal to the number of sources (n=m). the number of sources are estimated are exactly, therefore the estimated mixing matrix A is square. as stated before, the mixtures X for the noiseless case are given by:
   $X = AS$ 4.2

   when s is the vector of sources, and so
   $S = A^{-1}X = WX$ 4.3

(W is called the unmixing matrix and the reconstruction of the sources is done the same way the construction of mixtures was done, but with the W matrix instead of A.)

2. The number of mixtures is different from the number of sources (N≠M).
   Here, assuming that the estimated number of sources are correct, A is not square (it is of size N*M), and therefore, it has no inverse. The separation algorithm is:
   The chosen set (or sets) is taken and pass on all its clusters. For each cluster its coefficients are put into the distinct set. The coefficients of the rest of the sets and the other clusters are zero. Finally, each source is reconstructed according to (the corresponding coefficients plugged into) the DWP tree.

Description of the algorithm for estimating the number of clusters using the Hartman statistics:

For a collection of n points clustering is estimated to get k clusters
$C_1, C_2, \ldots, C_K$, when in each cluster $N_R = |C_R|$ points $(1 \leq R \leq K)$.

$D_R = \sum_{I,J \in CR}(D_{IJ})$ 4.4

when d_{ij} is the distance between points i and j (use the Euclidian distance $\sum_k(X_{ik}-X_{jk})^2$ while k runs on all the dimensions of the points).
Based on the Dr Parameter we define:
$W(k) = \sum_{r=1..k}(1/2n_r)D_r$ 4.5

that is actually the sum of the average distances of the points from one another within each cluster on all the clusters.

The Hartigan statistics is then:

$H(k) = [(W(k)/W(k+1))-1](n-k-1)$ 4.6

where k is the current number of clusters.

H(k) for k = 1...MAX is calculated and choose the estimated number of clusters as the minimal k so that:
$H(k) \leq TR$ 4.7

TR is some threshold value that Hartigan proposed to take as 10, TR is a function of the number of points n.

5. Results

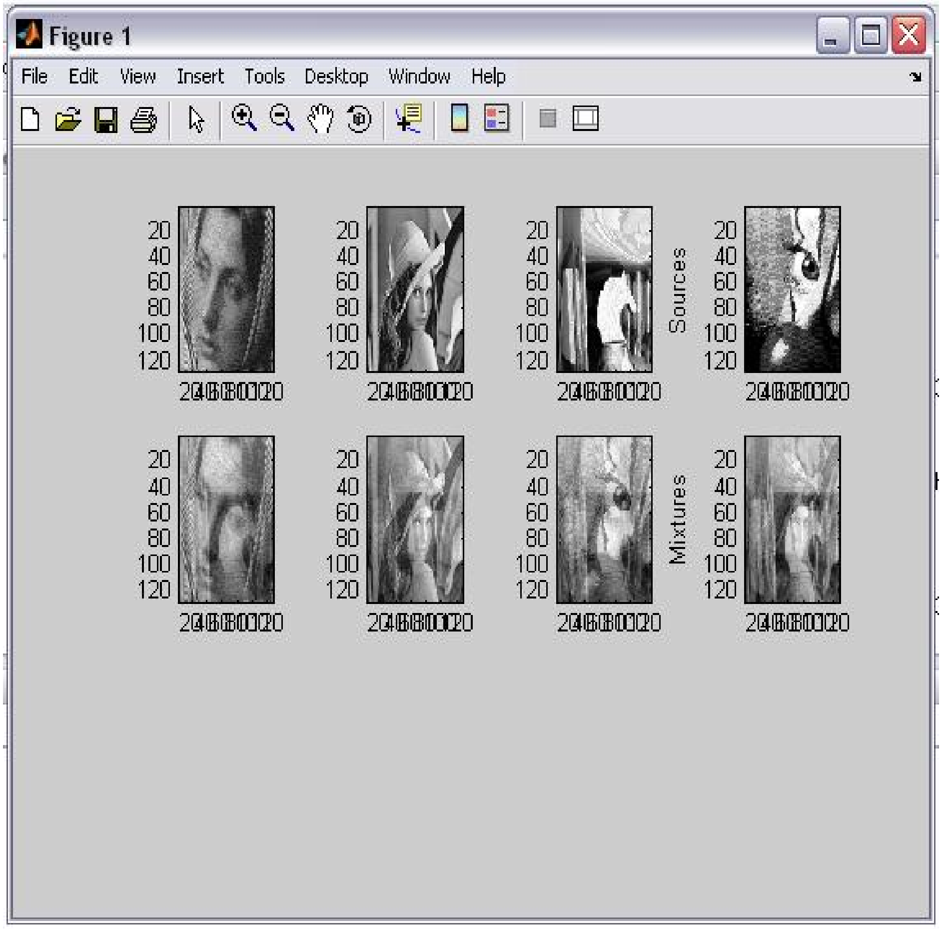

Fig 5.1: The Trained Output as a result of ICA Algorithm. – MIXING

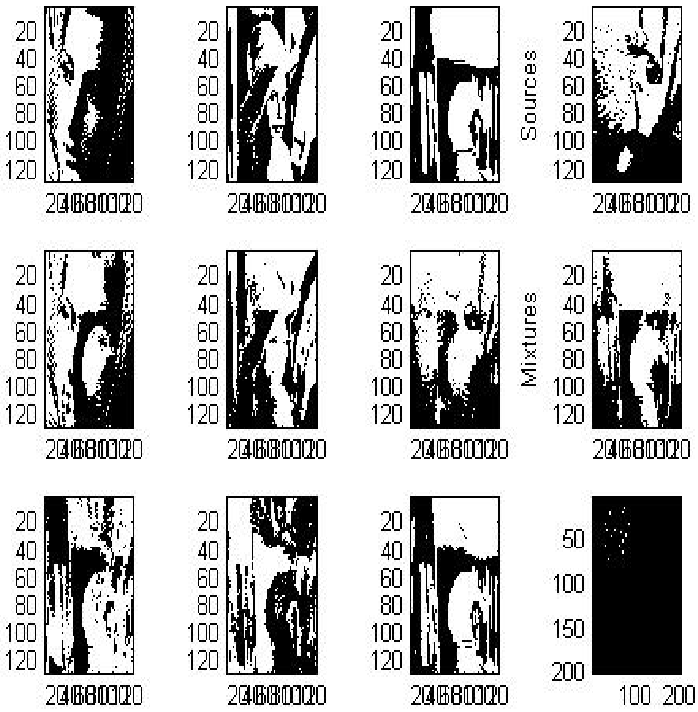

Fig 5.2 : The extract sources from the Unknown Sources as a Hartigan Product.

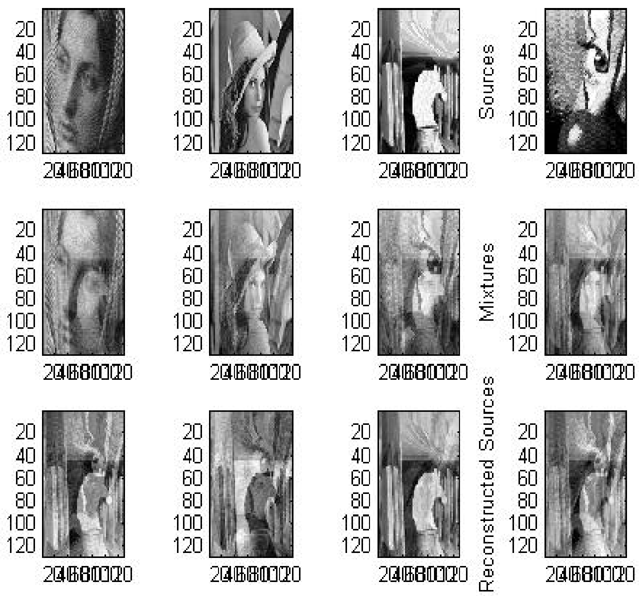

Fig 5.3 : Resultant images after removing Noise from the Images as a Restructured Result.

6. Conclusions

For the case of mixtures without white noise the NSE and CTE measures are very much close and therefore they are practICAlly the same measure of the quantity of separation, for the noised case there are larger differences between the two error, but still they are measure, but still they are close and the CTE is considered (cross talk error) as the error that shows us the separation quality when the nse is very much influenced by the noise energy.

- For the case of two non noised images none of the algorithms checked had problem to achieve visually good results in the sources separation.

- For the cases of three non noised images checked the best algorithm changed from clustering algorithm for the first case of lena & woman & chess (error of 0.14-0.15%) to the 3rd ICA algorithm for the 2nd case of chess & woman & clown (error of 0.63-0.64%). in general the 1st and 3rd ICA algorithms have relatively good performance here while the 2nd ICA algorithm reaches pretty high errors (12% and 21-22%) and also the clustering algorithm for the second case reaches the error of about 10%.
- for the case of two noised images the algorithm that achieves the lowest errors is undoubtedly the 3rd ICA algorithm with CTE of 0.25% and 0.001%. the other algorithms have much worse performance with relatively very high errors for the clustering algorithm and the 1st ICA algorithms for the case of lena & woman and better performance from them in the 2nd case of clown & chess (still highest error for the clustering algorithm).
- when three images were involved with noise the best algorithm still remains the 3rd ICA algorithm with higher errors than in the 2 images case of 0.64% and 3.1% CTE. the clustering algorithm fails to compete with the others and reaches errors of 1400 – 1500%. also the 1st and 2nd ICA algorithms return much worse errors than the 3rd ICA algorithm.
- the results the algorithms presented can operate well also in the case of more than three sources and mixtures and are not limited to a certain size of the problem. the examples in the results are for four and five sources and mixtures as returned by the 2nd and 3rd ICA algorithms and the quality of separationis exellent.

7. References

[1] http://www.cis.hut.fi/projects/ICA/cocktail/cocktail_en.cgi
Http://www.cis.hut.fi/projects/ICA/fastICA/code/dlcode.shtml ICA software in matlab.

[2] Http://www.cs.helsinki.fi/u/ahyvarin/papers/nn00new.pdf

[3] S. Amari, A. Cichocki, and H. H. Yang, “A new learning algorithm for blind signal separation,” *Advances in Neural Information Processing Systems*, vol. 8, pp. 752–763, 1996.

[4] [A.J. Bell and T.J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[5] T-W. Lee, M. Girolami, A.J. Bell, and T.J. Sejnowski, “A unifyinginformation-theoretic framework for independent componentanalysis,” *International Journal on Mathematical and Computer Modeling*, in press, 1998.

[6] J.-F. Cardoso and B. Laheld, “Equivariant adaptive source separation,” *IEEE Trans. on Sig. Proc.*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.