Smoothing and Adaptive Denoising for Blind Source Separation

Albert Bijaoui, Danielle Nuzillard

Abstract— **Pixel unmixing of multispectral astronomical images was examined as a blind source separation of an instantaneous mixture. The capabilities of separation algorithms were tested on different simulated images. The results showed that only in case of a high signal-to-noise ratio fine separations can be carried out. In this communication, the improvements resulting from a pre denoising are examined. Optimal linear smoothing and two adaptive wavelet denoisings were applied before separation. The increase of the separation quality of the obtained sources is discussed according to both the separation algorithms and the denoising methods.**

Keywords— **Multispectral analysis, Blind Source Separation, Independent Component Analysis, Wiener filter, Adaptive filter, Wavelet denoising.**

I. Introduction.

THE analysis of multispectral astronomical images can be examined as a blind source separation (BSS) of an instantaneous mixture [13]. The image model consists to set that each pixel value results from the contribution of different physical sources S_j giving a pixel mixing. In the hypothesis that this mixing is linear, each image X_i is written as:

$$X_i = \sum a_{ij} S_j + N_i \qquad (1)$$

where the matrix $A = [a_{ij}]$ is called the mixing matrix, N_i is the noise of image X_i. A and S, the vector formed by the sources S_j, are unknowns. The equation (1) is an ill-posed problem. Then a unique solution can only be obtained by introducing some constraint. It is a goal of BSS methods to provide the optimal one. The mixing generally increases the statistical dependence between the images; thus the search for the most physically independent sources is the main working direction.

A typical astronomical images set was analyzed with BSS algorithms [13]. The resulting sources obtained with different algorithms were pertinent from a physical insight. A set of experiments was also done on different simulated images the statistical properties of which being similar to astronomical ones [14].

The experimented BSS methods were based on different assumptions:

• A first kind of methods, called *Independent Component Analysis* (ICA) [5], takes into account the non-Gaussianity of the source probability density functions (PDF), which is measured by high order statistics (HOS);

A.Bijaoui is with the CERGA department, UMR CNRS 6527, Côte d'Azur Observatory, B.P.4229 F-06304, NICE CEDEX 4 (France), E-mail: Bijaoui@obs-nice.fr .

D.Nuzillard is with the Automatic and Microelectronic Laboratory, University of Reims Champagne-Ardennes, Moulin de la Housse, BP 1039, F-51687 Reims Cedex 2 (France), E-mail: Danielle.Nuzillard@univ-reims.fr

• The reduction of the spatial correlations between shifted sources is taken into account in the *Second Order Blind Identification* algorithm (SOBI,[1]) and related methods.

Our results on simulated images showed that fine separations can be provided from the two BSS types in case of a high signal-to-noise ratio (SNR). For low SNRs the separation quality felt down for both approaches.

The application of denoising methods is a natural idea. Linear smoothing was first applied in order to reduce the noise. That led to increase the separation quality. As smoothing decreases the peak intensities, we examined adaptive denoising methods in order to keep the largest values which play an important role in ICA. In this communication the effect of different denoising methods on different BSS algorithms was examined.

II. Tests of BSS algorithms.

A. The simulation.

We tested BSS algorithms on a set of simulated images with properties similar to astronomical ones. First, a program generated a set of images containing random Gaussian patterns (Figure 1). Each Gaussian was characterized by a central position, a major and a minor axes, an orientation and a central intensity. Each parameter was randomly determined with a probability density function (PDF) taking into account the properties of an astronomical image.

Then, the images were randomly mixed with positive weighting coefficients. This positivity constraint was set in order to take into account the specificity of astronomical images. The mixtures to be processed are displayed on Figure 2. We note that the PDFs and the autocorrelation functions of the sources are statistically identical.

B. Evaluation of the restoration quality.

As the true mixing coefficients $\overline{a}_{ij}$ are known, the quality of the restoration can be evaluated. The restored sources S are computed by the relation:

$$S = BX \qquad (2)$$

where $B \equiv [b_{ij}] = A^{-1}$ is the demixing matrix and X the mixtures. Without taking into account the noise, the following relation is verified:

$$X = \overline{A}\ \overline{S} \qquad (3)$$

where $\overline{A}$ is the true mixing matrix and $\overline{S}$ the true source vector. Hence it comes:

$$S = B\overline{A}\ \overline{S}. \qquad (4)$$

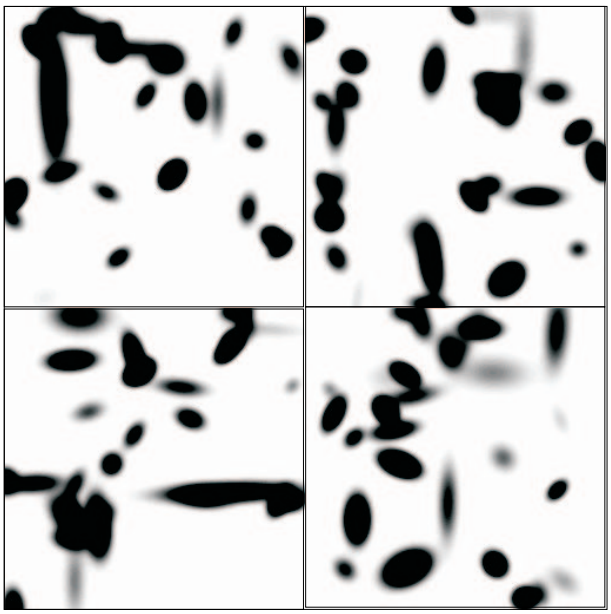

Fig. 1. Simulated Gaussian patterns. Each block corresponds to an initial source.

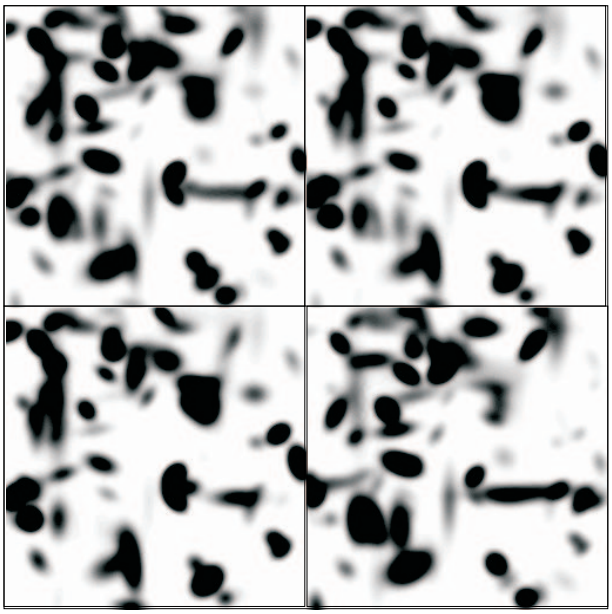

Fig. 2. Simulated mixtures without noise.

The matrix $D = B\overline{A}$ is the product of a permutation matrix by a diagonal one in case of a perfect separation. If not, its coefficients are spread since the restored sources are a combination of the original ones. A criterion is then provided by this dispersion. For a given restored source S_j we have:

$$S_j = \sum_{j'} d_{jj'} \overline{S}_{j'}. \tag{5}$$

The energy terms $e_{j'} = d_{jj'}^2$ are computed and sorted by increasing values. For a perfect restoration, only the last term is not equal to 0 for any permutation between the restored and the original sources. Let us call $r_{j'}$ the $e_{j'}$ rank; the concentration of the energy is measured through the Gini index [9]:

$$G_j = \frac{1}{n-1}[2\frac{\sum_{j'} e_{j'} r_{j'}}{\sum_{j'} e_{j'}} - (n+1)]. \tag{6}$$

$G_j = 0$ if all the $e_{j'}$ are equal, $G_j = 1$ if all the $e_{j'}$ but one are null. Then G_j provides a perfect information on the quality of the source restoration. The separation is evaluated by:

$$G = \sum_j G_j. \tag{7}$$

Experimentally G defines a criterion which is consistent with a visual inspection of the resulting sources.

III. Smoothing and Adaptive denoising.

A. BSS and linear smoothing.

Since the noise decreases the algorithm performances, it is natural to try to reduce it. A linear smoothing of the data is a common way to do it. Many filters can be implemented. The Wiener filter is theoretically the optimal one for stationary Gaussian signal and noise. If we may often admit that the noise is effectively stationary Gaussian, there is generally no evidence that the signal can be associated to a stationary Gaussian process.

ICA is based on the non-Gaussianity and the main information for these related algorithms is therefore supported by the largest pixel values. The smoothing tends to decrease the peak intensities reducing the non-Gaussianity of the distribution.

The cross correlations between shifted sources are used in SOBI [1]. The information at the highest frequencies is removed by a smoothing and this slightly modifies the correlations between the sources.

Hence a linear smoothing would not be *a priori* optimal for improving the separation of the sources for the two types of BSS algorithms.

In order to avoid the reduction of the peak intensities related to the signal an adaptive denoising on the mixtures can be applied. Within this approach, the signal properties are not considered as stationary and the smoothing depends on the local environment. In the past decade, many algorithms based on the wavelet transform were proposed for this task. This transformation concentrates the significant information into few coefficients, while the noise is spread on the whole set. By restoring the signal from the relevant coefficients a large part of the noise is removed. When compared to a linear smoothing the results are completely different. The smoothing spreads the peaks whereas the adaptive denoising selects them and restores a signal quite free of noise while keeping the peaks with their true intensity. The information in the highest frequency band is also selectively kept, leading to larger shifted cross correlations than in case of a uniform smoothing.

B. The optimal smoothing.

The Wiener filter is computed by [15]:

$$W(u,v) = \frac{S(u,v)}{S(u,v) + N(u,v)} \tag{8}$$

where $S(u,v)$ and $N(u,v)$ are respectively the spectral energy density of the signal and of the noise. (u,v) are the spatial frequencies.

$S(u,v)$ was estimated by taking into account the four mixtures without noise. The different Wiener filters were computed for each experimented Gaussian white noise. The filters looked like the Fourier transform of a Gaussian. Then the optimal smoothing was achieved by a simple fitted Gaussian smoothing. The mixture images corresponding to the same noise are smoothed with the same Gaussian mask.

C. The wavelet transform denoising.

Many different approaches of denoising with the wavelet transform were developed. The main principles are :

• *Application of a discrete wavelet transform (DWT)*: The standard DWT based on the multiresolution theory [10] is unredundant. The shift invariance is not kept and artifacts are consequently generated. Starck & Bijaoui [17] developed another approach based on the so-called *à trous* algorithm while Coifman & Donoho introduced a special case of this algorithm, called the shift-invariant discrete wavelet transform (SIDWT) [6].

• Thresholding: Donoho applied first a hard thresholding [7], zeroing the coefficients under the threshold, then a soft thresholding which reduces artifacts implied by the resulting discontinuities. Starck & Bijaoui developed a thresholding from a statistical decision rule [17].

• Restoration: In the case of a unredundant transform the restoration is generally done by a simple inverse transform. The application of a redundant transform enlarges the restoration possibilities. A regularization criterion can be introduced for selecting the best restoration [2].

The *à trous* algorithm with a cubic B-spline wavelet was chosen for denoising the mixtures. This transform is well suited for processing astronomical images which display diffuse spots.

D. The à trous algorithm.

Let us consider a discrete image $v(k,l)$. We pose :

$$v(k,l) = c(0,k,l) = < f(x,y), \phi(x-k, y-l) > \qquad (9)$$

where $\phi(x,y)$ is the scaling function which satisfies the dilation equation:

$$\frac{1}{4}\phi(\frac{x}{2},\frac{y}{2}) = \sum_{n,m} h(n,m)\phi(x-n, y-m). \qquad (10)$$

We define the approximation coefficients as:

$$c(i,k,l) = \frac{1}{4^i} < f(x,l), \phi(\frac{x-k}{2^i}, \frac{y-l}{2^i}) > \qquad (11)$$

where i is the current scale. The recursive relation is obtained:

$$c(i+1,k,l) = \sum_{n,m} h(n,m)c(i, k+n2^i, l+m2^i). \qquad (12)$$

The centered cubic B-spline is chosen. It corresponds to:

$$h(n,m) = h_3(n)h_3(m) \qquad (13)$$

were $h_3(n)$ is equal to 0 outside $[-2,2]$ and otherwise:

$$h_3(n) = \frac{1}{16}C_4^{2+n}. \qquad (14)$$

The wavelet function $\psi(x,l)$ is defined as:

$$\frac{1}{4}\psi(\frac{x}{2},\frac{y}{2}) = \phi(x) - \frac{1}{4}\psi(\frac{x}{2},\frac{y}{2}) \qquad (15)$$

the wavelet coefficients are computed by the relation:

$$w(i+1,k,l) = c(i,k,l) - c(i+1,k,l) \qquad (16)$$

These coefficients are the details lost from one approximation to the following one. Due to the redundancy, the inversion rule is not unique, but a trivial one exists:

$$c(i,k,l) = c(i+1,k,l) + w(i+1,k,l) \qquad (17)$$

This relation is true only if the wavelet coefficients correspond to a wavelet transform. After thresholding this is not the case and a correct inversion requires a restoration algorithm.

E. The multiresolution mask.

After transformation we get a set of noisy coefficient $w(i,k,l)$. A process on this set must be done in order to denoise them. Many rules were proposed :

• *Donoho's hard thresholding.* The threshold t is given by the rule :

$$t = \sigma_b\sqrt{2\log(N)} \qquad (18)$$

where σ_b is the standard deviation of the wavelet coefficient and N the number of independent coefficients. Even with a redundant transform, this number N decreases with the scale by a factor 4.

• *Donoho's soft thresholding.* Artifacts are generated with this previous rule. The thresholding is softened with the rule:

$$\begin{aligned} \hat{w} &= w + t \qquad w < -t \\ \hat{w} &= 0 \qquad -t \le w \le t \\ \hat{w} &= w - t \qquad w > t \end{aligned} \qquad (19)$$

• *Significant coefficients.* For object detection a selection rule based on the coefficient PDF in the case of local constant signal was introduced [16]. Here $t = k\sigma_b$ where k is a coefficient which depends on an uncertainty threshold. A soft thresholding can be done by taking into account two levels of detection k_1 and k_2.

We tested different rules and we selected two soft thresholdings with two levels t_1 and t_2. Below t_1 (in absolute value) the coefficients are set to 0. Above t_2 they are kept. Between the two thresholds the relation $\hat{w}$ versus w is linear.

Two processes were done `wave1` and `wave2`. For `wave1` denoising $t1$ and $t2$ are:

$$t_1 = 3.5\sigma_i \qquad t_2 = 4.5\sigma_i \tag{20}$$

while for `wave2` denoising t_1 and t_2 were computed with a rule similar to Donoho's one:

$$t_1 = 0.7\sigma_i(I-i) \qquad t_2 = \sigma_i(I-i) \tag{21}$$

where I is the largest scale. This rule gave better results than the classical Donoho's one for our images.

F. Reconstruction and iterations

The image can be directly restored from the set $\hat{w}(i,k,l)$, but many artifacts are generally observed. This is due to the fact that the applied transform is redundant and then the wavelet transform of the restored image is not exactly the set $\hat{w}(i,k,l)$. Some iterative processes were proposed in order to fit the restored image with the threshold coefficients.

For `wave1` the iterative process is based on the significant residual [11]. At each step the difference between the original image and the previously restored one is done, the coefficients of its wavelet transform with absolute value greater than the thresholds are kept, the image restored from these values is called the significant residue. This residue is added to the previous restored image for a next step. Few iterations are needed.

For restoring compressed images we introduced a method based on a regularization [2]. This method was applied for `wave2` in order to improve the restoration. The negative image values are thresholded to zero.

IV. BSS on denoised mixtures.

A. Noisy and denoised images.

A stationary Gaussian noise was added for each mixture. Three standard deviations 0.007 (b1), 0.07 (b2), 0.7 (b3) were examined. The introduced Gaussian noise led to signal-to-noise ratios (SNR) equal to 14.73, -5.27 and -25.27 dB respectively.

The optimal smoothing was performed with a Gaussian smoothing with a parameter of respectively 1.67 (b1), 2.94 (b2) and 13.05 (b3) pixel.

On Figure 3 the denoised images are drawn for the smoothing and the wavelet denoisings for b1 simulation. The differences are faint. On Figure 4 the denoised images are drawn for b2. With the wavelet denoisings the apparent structures are not smoothed, but the algorithm selects the information which seems significant. These differences are still more evident for b3 (Figure 5). The quality of the denoising can be estimated from the correlation between the images without noise and the denoised ones (Table I). `wave1` appears as the worst method. The signal-to-noise ratios (SNR) are also given.

B. BSS on denoised images.

Many experiments were done. We selected four methods:

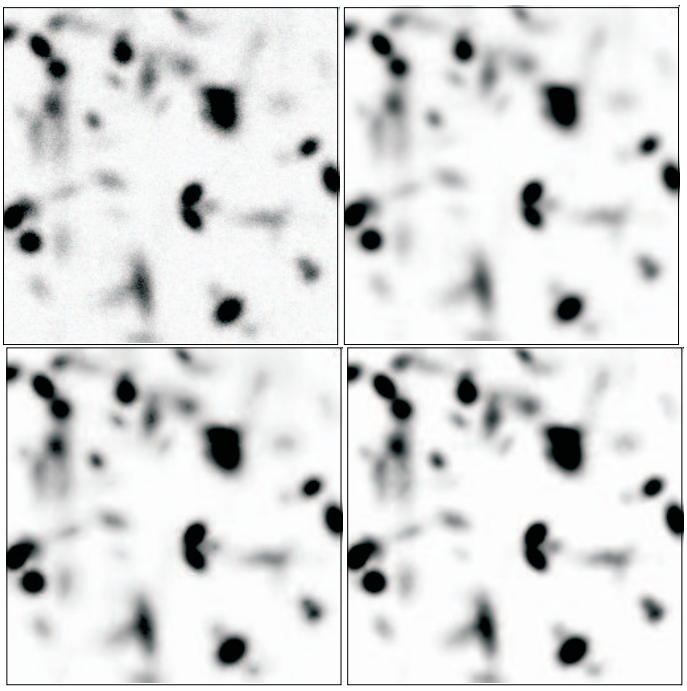

Fig. 3. The noisy simulated mixtures ($SNR = 14.73$ dB), its optimal smoothing and the wavelet denoisings wave1 and wave2.

TABLE I
Correlation between the original first mixture and its denoised one and signal to noise ratios.

Image	b1		b2		b3	
Method	C	SNR	C	SNR	C	SNR
Raw	0.979	14.73	0.435	-5.27	0.054	-25.27
Wiener	0.998	25.13	0.965	12.78	0.447	1.22
Wave1	0.999	26.40	0.960	12.15	0.218	-0.53
Wave2	0.999	29.32	0.978	14.60	0.606	0.80

• JADE. In JADE (Joint Approximate Diagonalization of Eigen-matrices) [3] the statistical independence of the sources is obtained through the joint maximization of the fourth order cumulants since these fourth order terms behave as contrast functions [5].

• FastICA. In FastICA [8] non-Gaussianity is measured by a fixed-point algorithm using an approximation of negentropy through a neural network. FastICA was extended for general contrast functions such as:

$$J_G(y) = |Ey\{G(y)\} - E\nu\{G(\nu)\}|^p \tag{22}$$

where ν corresponds to the Gaussian variable which has the same mean and the same variance as y. The algorithm was applied with $p = 2$, $G(y) = \log(\cosh(ay))$ and a deflation solution for which the sources are extracted successively.

• SOBI. SOBI (Second Order Blind Identification, [1]) is an efficient second order algorithm. It depends on the number p of spatial shifts of sources with themselves and their values. The variance-covariance matrix at any shift is computed from the cross correlations between the sources

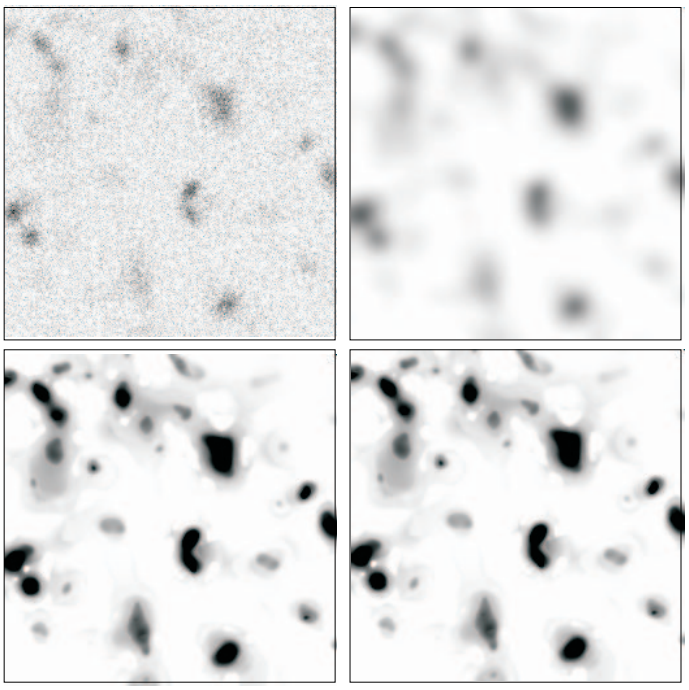

Fig. 4. The noisy simulated mixtures ($SNR = -5.27$ dB), its optimal smoothing and the wavelet denoisings wave1 and wave2.

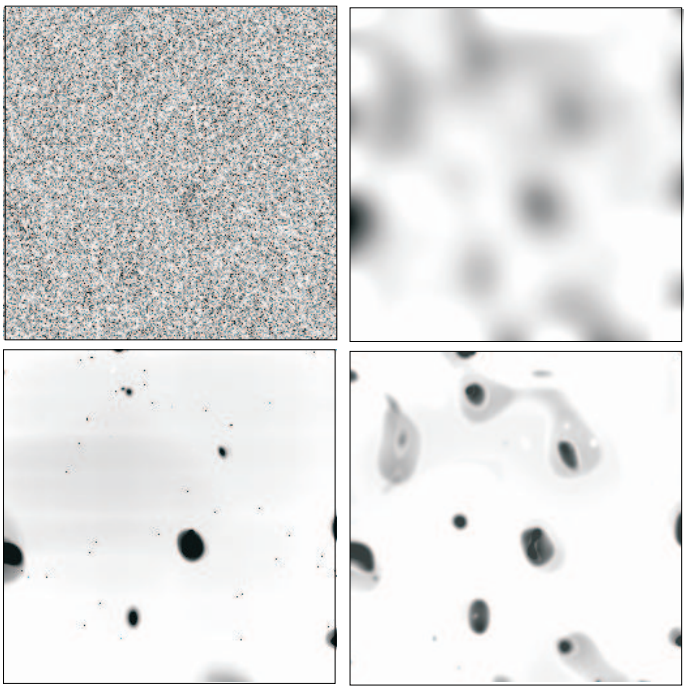

Fig. 5. The noisy simulated mixtures ($SNR = -25.27$ dB), its optimal smoothing and the wavelet denoisings wave1 and wave2.

and the shifted ones. After the data whitening, the set of p variance-covariance matrices is computed. Cross-correlation terms are minimized, thus diagonal terms are maximized. A joint diagonalization criterion of several p covariance matrices is applied [4]. The algorithm was adapted to $2D$ images.

• f-SOBI. Nuzillard [12] modified SOBI by computing the cross-correlations of the Fourier transforms, which are easily estimated in the direct space. This can be viewed as an alternative correlation choice. SOBI takes into account the correlation at short distances, while the correlations at short frequency distances play the main role in f-SOBI. The algorithm was also adapted to the two-dimensional field.

C. Results.

In Table II the resulting separation quality indices G are indicated for the different BSSs. We remark:

• A decrease of the BSS quality with an increasing noise;

• With SOBI, working in the direct space, quite no improvement is obtained for all methods. So we do not take further into account this algorithm.

• A significant gain is obtained with the three other algorithms. This improvement depends on SNR:

– At 15 dB, smoothing and wavelet denoisings lead to similar results. The three BSS algorithms can be considered as quite equivalent.

– At -5 dB, a significant gain is obtained, especially for f-SOBI. For BSS the denoising with the wavelet transform appears better than the optimal smoothing.

– At -25 dB, the gain is faint but real. The best BSS is obtained with JADE on the optimal smoothing. Nevertheless the wavelet denoisings and f-SOBI lead both to comparable results.

In the case of adaptive denoising the multiscale mask depends on the images. So, the sources computed from the denoised mixtures are not optimal. It could be better to compute them from the noisy mixtures and to denoise the resulting images. It is possible to iterate on the new denoised sources, and so to improve the separation. Then the procedure is the following one:

• Denoise the mixture and apply a BSS algorithm. A demixing matrix $B1$ is got;

• Compute the sources by applying $B1$ on the noisy mixtures;

• Process again the resulting sources with a BSS algorithm, which leads to a new demixing matrix $B2$;

• Compute the new sources by applying $B2$ on the preceding noisy sources.

The process may be iterated up to convergence. In Table (II) the results were obtained from one iteration (`wave1i`). At first step `wave1` was applied and the BSS was done with f-SOBI. Then `wave1` was used again, but all the BSSs methods were tested. We note the significant gain in the separation quality.

As the optimal smoothing was the same for all the mixtures, no gain can be obtained by this procedure.

D. Discussion.

The couple f-SOBI and the wavelet denoisings appears to be the most robust scheme for BSS of noisy images, even if it does not bring always the best separation. The denoising based on the detection on the significant wavelet

TABLE II

INDICES OF THE SEPARATION QUALITY OBTAINED ON THE DENOISED IMAGES. FASTICA IS APPLIED WITH $G(y) = \log(\cosh(ay))$ AND A DEFLATION ALGORITHM. WN MEANS THE MIXTURES WITHOUT NOISE, RAW THE NOISY MIXTURES WITHOUT SMOOTHING OR DENOISING.

Image	FastICA	JADE	SOBI	f-SOBI
wn	3.99	3.10	1.49	3.96
b1 raw	2.27	3.10	1.27	3.22
b1 wiener	3.93	3.96	2.01	3.95
b1 wave1	3.94	3.97	2.18	3.96
b1 wave2	3.93	3.96	2.15	3.96
b2 raw	1.88	1.90	2.01	2.36
b2 wiener	2.40	2.62	1.58	3.26
b2 wave1	2.73	2.79	1.97	3.43
b2 wave2	2.35	2.78	1.70	3.39
b3 raw	2.23	2.03	1.99	2.06
b3 wiener	2.33	2.88	2.46	2.38
b3 wave1	2.67	2.82	2.16	2.80
b3 wave2	2.62	2.77	2.20	2.66
b3 wave1i	2.70	2.92	2.52	2.90

coefficients, `wave1`, carries out better results than the one based on decreasing thresholds with the scale, `wave2`, although its SNR was smaller than the SNR of the second algorithm.

The constant thresholding rule keeps only significant coefficients and then removes signal. At the first scale the percentage of lost information is the same for `wave1` and `wave2`, but at larger scales `wave2` keeps all the signal. As the noise plays a negligible role at large scales, the correlation with the original image is high. But BSS is mainly based on extreme values, which determines the high order statistics computed for the separation. With `wave1` only the significant values are kept and the separation acts with quite sure data, while the statistic is polluted by noise residues with `wave2` resulting in a separation rather worse than for `wave1`.

The main goal for denoising signals for BSS is not to get the best SNR but to keep the extreme values.

V. CONCLUSION.

BSS is a signal processing sensitive to the noise. In this paper it was shown that standard procedures of noise reduction can really improve the results. The experiments do not show that the best denoising carries out the best BSS. Only one simulation was processed; it could then be interesting to build other ones with different textures in order to get a general overview on the link between denoising and BSS.

An algorithm based on the spatial organization, like f-SOBI, carried out more robust results than the ones based on the independence between the pixel values. This point urges us to examine a method combining independence and spatial criteria.

REFERENCES

[1] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, E. Moulines. A blind source separation technique using second-order statistics. *IEEE Trans. SP,* 45, pp. 434-444, 1997.

[2] Y. Bobichon, A. Bijaoui. Regularized multiresolution methods for astronomical image enhancement. *Wavelet applications in signal and image processing V* ed. A.Aldroubi, A.F.Laine, M.A.Unser pp.36-45 SPIE 1997.

[3] J.F. Cardoso, A. Souloumiac. Blind Beamforming for non-Gaussian signals. *IEE Proceedings-F*, 40(6), pp. 362-370, 1993.

[4] J.F. Cardoso, A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, 17, pp. 161-164, 1996.

[5] P. Comon. Independent Component Analysis, a new concept. *Signal Processing*, 36, pp.287-314, 1994.

[6] R.R. Coifman, D. Donoho. Translation invariant denoising. *Technical report* 475, Dept. of Statistics, Standord, University, May 1995.

[7] D.L. Donoho, I.M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81, pp.425-455, 1994.

[8] A. Hyvärinen A., E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9, pp.1483-1492,1997.

[9] M.G. Kendall, A. Stuart. *The Advanced Theory of Statistics*, vol. 1, p. 48, Charles Griffin, London, 1969.

[10] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11, pp.674-693, 1989.

[11] F. Murtagh, J.L. Starck, A. Bijaoui. Image restoration with noise suppression using the Wavelet Transform II. *Astron. Astrophys. sup. ser.*, 112, pp. 179-189, 1995.

[12] D. Nuzillard. Adaptation de SOBI à des données fréquentielles. *GRETSI'99 Vannes*, pp. 745-748, 1999.

[13] D. Nuzillard, A. Bijaoui. Blind Source Separation and Analysis of Multispectral Astronomical Images. *Astron. Astrophys. Sup. Ser.*, v.147, pp.129-138, 2000.

[14] D. Nuzillard, A. Bijaoui. Multispectral analysis, Blind Source Separation and Mutual Information. *IEEE. Image Proc.*, submitted 2001.

[15] A. Papoulis. *Probability, Random variables and Stochastic Processes*, p.457, Mac Graw Hill, New York, 1965.

[16] E. Slezak, A. Bijaoui, G. Mars. Structures identification from galaxy counts. *Astron. Astrophys.*, 227, pp.301-316, 1990.

[17] J.L. Starck, A.Bijaoui. Filtering and deconvolution by the wavelet transform. *Signal Processing*, 35, pp. 195-211, 1994.