Multiscale Contrasts for Blind Source Separation

Danielle Nuzillard, Albert Bijaoui

Abstract— **The analysis of multispectral astronomical images was examined as a blind source separation (BSS) of an instantaneous mixture. The experimented BSS methods were based on different assumptions: i/ the non-Gaussianity measured by high order statistics (HOS), ii/ the correlation between shifted sources. A HOS approach taking into account the spatial organization is proposed from a decomposition in the wavelet space. A multiscale contrast taking into account the probability density function (PDF) at each scale, was first introduced. The resulting contrast provides few improvement in the separation. Thus a definition of the multiscale contrast which took into account a mask associated to a thresholding was examined. We show in this communication that this new statistical quantity provides an available criterion for BSS.**

Keywords— **Multispectral analysis, Blind Source Separation, Mutual Information, Independent Component Analysis, Wavelet transform**

I. Introduction.

IN paper [16] the analysis of multispectral images was examined as a blind source separation (BSS) of an instantaneous mixture. Each image X_i was written as a linear mixture:

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

where the matrix $A = [a_{ij}]$ is the mixing matrix and N_i the noise of image X_i. The capabilities of different BSS algorithms were examined on simulated images [17]. The experimented methods were based on different assumptions:

- The non-Gaussianity, measured by high order statistics (HOS). This leads to what has been called *Independent Component Analysis* (ICA) [6];
- The spatial correlation by reducing the correlations between shifted sources as it is taken into account in *Second Order Blind Identification* (SOBI, [1]) and related methods.

Fine separations were carried out in case of a high signal-to-noise ratio (SNR), but for low SNRs the separation quality felt down. In paper [2] it was shown that an adaptive denoising by the wavelet transform improved this quality. Then the idea of building a HOS approach associated to the spatial organization from a decomposition in the wavelet space raised. A multiscale contrast taking into account the probability density function (PDF) at each scale was introduced. The improvement in the separation obtained by this contrast was faint. The gain obtained by denoising the images was due to a thresholding, i.e. a coefficient selection. Thus a definition of the multiscale contrast using such a mask was examined. In this paper, it is shown that this new statistical quantity provides an available criterion for BSS.

Some basic notions on BSS are given in section 2. The wavelet transform and the multiscale mask are introduced in the following section. Then, the related multiscale contrast is defined. Its capabilities to separate sources are shown on simulated mixtures in section four. The feasibility to build a new BSS algorithm based on this contrast is discussed in conclusion.

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

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 .

II. Decorrelation, Independence and Mutual Information.

A. Decorrelation and the Karhunen-Loève expansion.

BSS is a class of statistical methods that decompose a set of m signals into n independent sources and estimate the mixing matrix A. The simplest model corresponds to equation (1). The sources are constrained to have unit variance. The mixing matrix A can be estimated up to permutation and phase shifts. The problem remains ill-posed and additional considerations are needed to solve it.

The Karhunen-Loève expansion (KL) carries out decorrelated sources. Any rotation of the resulting sources keeps this property, since KL provides the sources which maximize the energy concentration. If v_i is pixel value of source i, the sources PDF is $p(v_1, \ldots, v_n)$. The statistical independence between the source intensities means that:

$$p(v_1, \ldots, v_n) = \prod_i p_i(v_i) \qquad (2)$$

where $p_i(v_i)$ is the marginal PDF of i-th source. This relation implies that the variables v_i are not correlated, but the converse proposition is not true. It may exist sources more independent than those given by KL.

B. The source mutual information.

The mutual information of a set of n sources (SMI) is defined as the Kullback-Leibler divergence between the joint probability and the probability obtained by the product of the marginal probabilities [6]:

$$I(S_1, \ldots, S_n) = \sum_{v_1,\ldots,v_n} p(v_1, \ldots, v_n) \log_2 \frac{p(v_1, \ldots, v_n)}{\prod_{i=1,n} p_i(v_i)}. \qquad (3)$$

If the sources are independent this divergence is equal to 0, then SMI carries out an available independence criterion between the sources. We have also:

$$I(S_1, \ldots, S_n) = \sum_i H(S_i) - H(S_1, \ldots, S_n) \qquad (4)$$

where $H(S_i)$ is the entropy of source i defined by:

$$H(S_i) = -\sum_v p_i(v) \log_2 p_i(v) \qquad (5)$$

and $H(S_1, \ldots, S_n)$ is the mutual entropy:

$$H(S_1, \ldots, S_n) = -\sum_{v_1, \ldots, v_n} p(v_1, \ldots, v_n) \log_2 p(v_1, \ldots, v_n). \quad (6)$$

In the case of Gaussian PDFs, SMI is equal to the expression [12]:

$$IM_2 = -\frac{1}{2} \log_2 \det R \quad (7)$$

where R is the matrix of the coefficients of correlation. If the sources are not correlated, $IM_2 = 0$. Any set of sources obtained by rotation from the KL expansion carries out a null IM_2, i.e. a null SMI. Then SMI in this case does not add any information.

If the source PDFs are non-Gaussian, SMI depends also on HOS. The application of Edgeworth's expansion of the PDF [13] introduces cumulants from which SMI can be estimated. This approximation may be not sufficient and a set of methods has been developed in order to get the best SMI evaluation.

C. SMI and its related contrast function.

The observed SMI can be derived from equation (3). It is generally impossible to obtain an available estimation of the probability $p(v_1, v_2, \ldots, v_n)$ from experimental data. Then $H(S_1, S_2, \ldots, S_n)$ is badly estimated. Comon [6] noticed that the exact SMI value was not needed and it was only necessary to know for which mixing matrix A, or its inverse $B = A^{-1}$, the mutual information is minimum. The sources are obtained from the images from the demixing relation:

$$S = BX \quad (8)$$

since we obtain:

$$H(S_1, \ldots, S_n) = H(X_1, \ldots, X_n) + log_2 |\det B|. \quad (9)$$

Then we get:

$$I(S_1, \ldots, S_n) = \sum_{i=1}^{n} H(S_i) - H(X_1, \ldots, X_n) - log_2 |\det B|. \quad (10)$$

Next it is sufficient to maximize the function [6]:

$$\begin{aligned} C &= -\sum_{i=1}^{n} H(S_i) + log_2 |\det B| \\ &= -\sum_{i=1}^{n} H(S_i) - log_2 |\det A|. \end{aligned} \quad (11)$$

C is a contrast function which depends on the entropy evaluation for each source separately.

III. Contrast and Multiresolution.

A. Contrast of a pixel pair.

Let us consider two image pixels with values $y_i \in Y_i$ and $z_i \in Z_i$ for image $X_i = Y_i \bigoplus Z_i$. Applying relation (8) to this pair we get:

$$T = BY \qquad U = BZ \quad (12)$$

where T_j and U_j are the subsets of the source S_j pixels homologous to y_i and z_i ($S_j = T_j \bigoplus U_j$).

The mutual information related to the pair is: $I(T_1, \ldots, T_n, U_1, \ldots, U_n)$ such that:

$$\begin{aligned} I(T_1, \ldots, T_n, U_1, \ldots, U_n) &= \sum_i (H(T_i) + H(U_i)) \\ &- H(T_1, \ldots, T_n, U_1, \ldots, U_n). \end{aligned} \quad (13)$$

We have the relation:

$$\begin{pmatrix} T \\ U \end{pmatrix} = \begin{pmatrix} B & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} Y \\ Z \end{pmatrix}. \quad (14)$$

If we set:

$$V = \begin{pmatrix} T \\ U \end{pmatrix}, W = \begin{pmatrix} X \\ Y \end{pmatrix}, C = \begin{pmatrix} B & 0 \\ 0 & B \end{pmatrix}. \quad (15)$$

The relation (14) is written as:

$$V = CW. \quad (16)$$

Then we get:

$$H(V_1, \ldots, V_n) = H(W_1, \ldots, W_n) + \log_2 |\det C|. \quad (17)$$

Now, we have the relation:

$$\det C = \det B^2 \quad (18)$$

Thus we obtain:

$$\begin{aligned} I(V_1, \ldots, V_n) &= \sum_i H(V_i) \\ &- H(W_1, \ldots, W_n) - 2\log_2 |\det B|. \end{aligned} \quad (19)$$

Consequently the following contrast for a BSS insight is:

$$C_2 = -\sum_i H(V_i) + 2\log_2 |\det B|. \quad (20)$$

Obviously $H(V_i) = H(T_i) + H(U_i)$ et $H(T_i) = H(U_i)$. Thus we can write:

$$C_2 = 2C_1. \quad (21)$$

Then the pixel dependencies do not play a role for BSS. But we implicitly admit that the image is stationary, the y_i and z_i PDFs being the same. Now if we apply a transformation on the image the resulting coefficients can have different PDFs so that new contrasts, eventually more efficient than the direct one, may be computed from these PDFs.

B. Transformation and Information.

At the second order the best transform of a single image is the Karhunen-Loève expansion related to the spatial auto-correlation. For an increasing window size this transform tends to the Fourier one. The coefficient PDFs, then their entropy, depend on the spatial frequency, the variance at low frequencies being generally higher than the one at high frequencies. Thus we could compute a contrast from the Fourier coefficients and then use it as a criterion for BSS. The entropy per coefficient becomes the problem to solve. This implies a statistical model on the Fourier transform. These remarks lead us to examine a linear transform close to the Fourier transform but allowing one to estimate the PDF or at least the first four cumulants. Taking into account its compression capabilities, the wavelet transform seemed to be more suitable than the shift windowed Fourier transform [7].

C. The wavelet transforms and multiscale contrasts.

The wavelet transform was popularized by the introduction of the multiresolution theory and the discrete wavelet transform (DWT) [14].

The goal is to define a multiscale contrast based on the information contained in the wavelet coefficients at each scale. Let us consider a DWT. After transformation, a set of coefficients $w(i,k,l,d)$ is obtained. At a scale level i, the scale factor is 2^i, (k,l) is the current pixel location and d is the direction (horizontal, vertical and diagonal). For an image of $2^I \times 2^I$ pixels, three arrays of $2^{I-i} \times 2^{I-i}$ pixels are computed at each scale i.

At the first scale, the number of pixels per array is sufficient for computing the empirical entropy $H(1,d)$. But the number of pixels exponentially decreasing with i, the entropy can not be any more directly estimated. A way to solve this drawback lies in the application of a redundant transform for which at each scale the number of coefficients is not decimated. It is the case of the *à trous* algorithm (AT) [18] or of the Coifman & Donoho shift-invariant discrete wavelet transform (SIDWT) [5]. Clearly, it exists a statistical dependence between the coefficients; the information amount is quite not increased by the redundancy, but the number of values allows to compute an empirical PDF and then its related entropy. The Haar transform [8], in its SIDWT version, was first applied in order to compute a contrast on the data. An AT transform based on a B-spline scaling function was also applied. In this case the wavelet function was regular up to order two, contrary to the Haar wavelet which is irregular from order 0. The AT transform carried out better results.

The redundancy is only applied for the PDF evaluation. The empirical entropy per coefficient at given scale and direction is computed from this estimation. But the image can be restored by the decimated coefficients, then the number of useful coefficients decreases by a factor 4 from a scale to the following one. Consequently, the multiscale contrast (MSC) is defined as:

$$MSC = -\sum_{i=1,I}\sum_{d}[\frac{1}{4^i}H(i,d) - \log_2|\det B|]. \quad (22)$$

The redundancy allowed us to keep the full pixel sample for the PDF estimation, but the wavelet coefficients, at given scale and direction, are correlated. Then $H(i,d)$ is still badly estimated for increasing i, but its value has a decreasing influence.

The first results obtained with different transforms did not show a significant gain for this criterion. It is useful to remind that the idea to apply the wavelet transform for BSS came from the denoising. In this operation a coefficient selection was performed in order to further restore a clean image. For this selection a mask was computed scale per scale from different rules. Then a new ingredient was introduced : the multiscale mask.

D. Multiscale mask of a set of images.

Let us consider a set of n images. At the i-th scale and the (k,l) location the wavelet coefficients $w(m,i,k,l,d)$ are obtained, where m is the image index. The local energy is defined as:

$$E(i,k,l,d) = \sum_{m=1,n} w^2(m,i,k,l,d). \quad (23)$$

If the images are globally constant in the neighborhood defined by the compact support of the wavelet at scale i, the PDF of the energy E results from the noise distribution. If the noise is Gaussian, the wavelet coefficients are also Gaussianly distributed. Its mean is equal to 0 whatever the local mean because the wavelet function has a null mean. Its standard deviation is $k_i\sigma$, where k_i is a coefficient factor depending on the chosen transform and on the current scale i. k_i is equal to 1 for DWT and its SIDWT variant. For AT it is easy to evaluate numerically the k_i coefficients.

If the images have a white Gaussian noise with identical deviations σ, $E(i,k,l,d)$ follows a χ^2-square law with n degrees of freedom, $P_n(E)$. Its mean is equal to $nk_i^2\sigma^2$ and its standard deviation is $\sqrt{2n}k_i^2\sigma^2$. Let us consider the case of constant but noisy images, the energy coefficients $E(i,k,l,d)$ are distributed according to the distribution $P_n(E)$. At the i-th scale, the (k,l) location and the d direction, $E_0(i,k,l,d)$ is measured, the hypothesis H_0 that the images are locally constant can be tested. A statistical level ϵ is introduced. If:

$$P_n[E(i,k,l,d) > E_0(i,k,l,d] < \epsilon \quad (24)$$

is satisfied the probability for the local energy to be greater that $E_0(i,k,l,d)$ under the H_0 hypothesis is smaller than ϵ. Then the H_0 hypothesis is rejected and a detection is obtained. For a decreasing ϵ value, the number of false detections decreases but the misses of real local variations increase. ϵ is chosen as a compromise between false detection and misses. This allows one to define the selection thresholds $T(i,d)$ such that:

$$P_n[E(i,k,l,d) > T(i,d)] < \epsilon. \quad (25)$$

The multiscale mask is obtained by taking account this selection. In our experiments a fuzzy mask was computed:
- For $E(i,k,l,d) < m+3s$, where m is the expectation and s the deviation in case of uniform image, the mask value $M(i,k,l,d) = 0$;
- For $E(i,k,l,d) > m+8s$, $M(i,k,l,d) = 1$;
- For an intermediate energy, the mask value is computed by a linear relation.

E. Multiscale contrasts and the multiscale mask.

A contrast can be computed from the multiscale mask by applying equation (22). For each wavelet set the number of selected coefficients is too small for computing an empirical entropy, even with a redundant transform. Then we derived it from the third and the fourth order cumulants, C_3 and C_4, using the Edgeworth expansion [10]:

$$H(i,d) = \log \sigma - \frac{1}{12}C_3^2 - \frac{1}{48}C_4^2. \quad (26)$$

For the cumulant estimations the wavelet coefficients are weighted by the mask values. Then the criterion called the *Multiscale Masked Contrast* (MSMC) is computed by the following algorithm:

- Compute the wavelet transform of the images;
- Deduce the full energy at each scale, each location and, if necessary, each direction;
- From a statistical detection rule, determine the fuzzy mask;
- For each scale and, if necessary, each direction, compute the entropy from the cumulants (equation 26);
- Compute the contrast from relation (22).

IV. Experimental tests.

A. The simulation.

We tested BSS algorithms on a set of simulated images with properties similar to astronomical ones. A program generated a set of images each containing random Gaussian patterns (Figure 1). Then the images were randomly mixed with positive weighting coefficients. The mixtures to be processed are drawn on Figure 2.

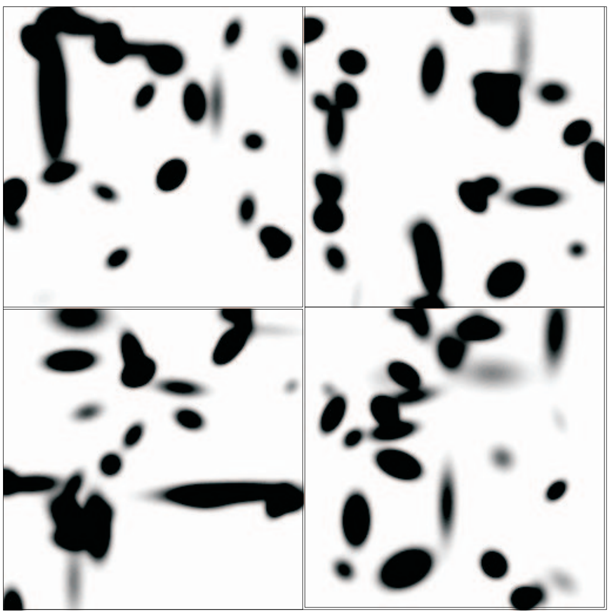

Fig. 1. Simulated Gaussian patterns. Each block corresponds to an initial source.

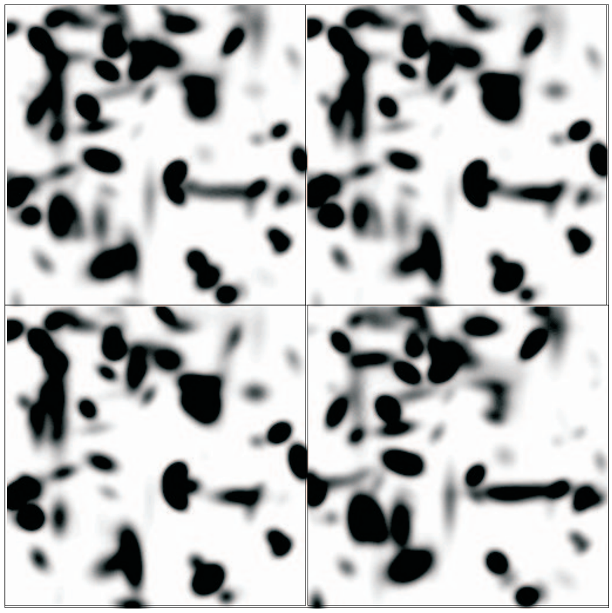

Fig. 2. Simulated mixtures without noise.

B. Evaluation of the restoration quality.

As we know the true mixing coefficients $\overline{A} = [\overline{a}_{ij}]$, it is possible to quantify the quality of the restoration. A criterion based on the estimation of the energy concentration of the restored sources versus the original ones can be applied [17]. The restored sources are computed by the relation (8). If $\overline{S}$ is the true source vector we get:

$$S = B\overline{A}\,\overline{S} \quad (27)$$

In the case of a perfect separation the matrix $D = [d_{jj'}] = B\overline{A}$ is the product of a diagonal matrix with a permutation one. If not, its coefficients are spread, the restored sources being a combination of the original ones. For a given j, The energy terms $e_{j'} = d_{jj'}^2$ are computed and sorted by increasing values; $r_{j'}$ is the rank of $e_{j'}$. For perfect restoration and for any permutation between the restored and the original sources, only the last term is not equal to 0. The Gini concentration index of the energy is [11]:

$$G_j = \frac{1}{n-1}[2\frac{\sum_{j'} e_{j'} r_{j'}}{\sum_{j'} e_{j'}} - (n+1)]. \quad (28)$$

$G_j = 0$ if all the $e_{j'}$ are equal, $G_j = 1$ if all the $e_{j'}$ are null except one value. The separation is then evaluated by:

$$G = \sum_j G_j. \quad (29)$$

G defines a criterion which is experimentally consistent with visual inspection.

C. ICA applied tools.

We applied two different tools. In JADE [3] the statistical independence between the sources is obtained through the joint maximization of the fourth order cumulants since these fourth order terms behave as contrast functions [6].

In FastICA [9] non-Gaussianity is measured by a fixed-point algorithm using an approximation of negentropy through a neural network. FastICA was extended for contrast functions such as:

$$J_G(y) = |Ey\{G(y)\} - E\nu\{G(\nu)\}|^p \quad (30)$$

where ν corresponds to the Gaussian variable which has the same mean and the same variance as y. $G(y)$ is used in its derivative form, $g(y)$, in the algorithm. The function $g(y) = \tanh(ay)$ was selected with $p = 2$. The algorithm works in two different ways: a symmetric solution is used for which the sources are computed simultaneously (FastICA-st), or a deflation one for which the sources are extracted successively (FastICA-dt).

D. SOBI and its derived versions.

SOBI (Second Order Blind Identification, [1]) is an efficient second order algorithm. It depends on the number of spatial shifts p of sources with themselves and on their values. The variance-covariance matrix at any shift is computed from the cross correlations between sources and the shifted ones. Cross-correlation terms are minimized, thus diagonal terms are maximized. A joint diagonalization criterion of several p covariance matrices is applied.

Nuzillard [15] 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 algorithms were adapted to $2D$ images. SOBI-1D and f-SOBI-1D correspond to these algorithms for which the image data are considered as one line, while the $2D$ structure is taken in account for SOBI-2D and f-SOBI-2D.

E. MSMC and separation quality.

In table I the different contrast values are given. In the first column the BSS algorithms are indicated. A positive constraint [15] was also applied (the symbol '+' is added at the end of the BSS algorithm name in the table). This algorithm variant essentially improves SOBI and f-SOBI.

It exists a general correlation between the contrast and the index of the quality separation. The coefficients of correlation are given in table II.

F. Discussion.

Clearly the criterion based on the multiscale masked contrast carries out the best correlation with the quality separation criterion. Getting the highest MSMC value allows one to obtain a separation which could be as close as possible of the original sources. Nevertheless this assumption supposes that the conditions of validity of equation (1) are true, providing that the mixtures are separable and that the sources check the MSMC criterion.

Let us consider a set of sources $\overline{S}$, their observed mixtures are X on which a BSS algorithm T is applied bringing the set of sources S. If we apply T on $\overline{S}$ we obtain S also when:

TABLE I

Comparison of the quality index G with different separation criteria. C is the contrast deduced from the empirical entropy, C-BS, the multiscale contrast from a redundant wavelet (AT) with a cubic B-spline as scaling function, C-H, the multiscale contrast obtained with the SI-DWT with the Haar transform, MSMC, the multiscale masked contrast with the AT transform and a cubic B-spline scaling function.

BSS	G	C	C-BS	C-H	MSMC
Sources	4.	30.09	62.24	45.49	185.99
Mixtures	2.53	17.88	47.46	34.63	30.04
KL	2.60	17.88	49.15	35.85	31.92
FastICA-dt	3.99	28.84	61.48	45.05	184.95
FastICA-st	3.49	29.00	61.50	45.17	185.34
JADE	3.10	26.99	60.90	44.54	180.98
SOBI-1D	1.44	21.47	54.93	40.42	39.75
SOBI-2D	1.49	22.03	54.37	40.75	26.25
f-SOBI-1D	3.97	26.96	60.89	44.52	181.21
f-SOBI-2D	3.96	26.71	60.73	44.42	179.64
FastICA-dt+	2.58	23.57	56.79	41.51	112.21
FastICA-st+	3.17	23.94	57.31	41.81	118.90
JADE+	2.80	28.98	61.88	45.30	185.36
SOBI-1D+	3.24	20.58	53.72	39.36	150.19
SOBI-2D+	2.85	23.12	55.78	40.83	64.88
f-SOBI-1D+	4.00	30.03	62.21	45.49	186.00
f-SOBI-2D+	3.99	28.93	61.86	45.27	185.19

TABLE II

Correlation between the contrasts and the quality index.

Criterion	Correlation
C	0.693
C-BS	0.657
C-H	0.620
MSMC	0.838

- The algorithm takes into account the same constraint on the mixing matrix;
- The convergence towards the absolute maximum of the criterion is obtained;
- Only one absolute maximum exists.

Now, if T is applied on S, S is also restored. T is a projection operator. All images set which carries out S can be a solution of equation (1), even the original mixtures X. But the set which maximizes a contrast is preferable because a mixing with random coefficients generally decreases it.

SOBI and f-SOBI take into account the spatial organization. In the case of the tested mixtures f-SOBI is efficient. JADE and FastICA based on HOS are also efficient. That means that the true original sources $\overline{S}$ have a mutual information near to 0 and that its cross correlations between their shifted Fourier transform have very few energy.

The use of the wavelet transform permits to introduce indirectly the spatial organization. If two pixels are permuted the mutual information is not changed, but the distribution of the wavelet coefficients is slightly modified. Then MSC

TABLE III

Comparison of the quality index G with different separation criteria for the noisy mixtures. C is the contrast deduced from the empirical entropy, and MSMC the multiscale masked contrast with the AT transform and a cubic B-spline scaling function.

BSS	G	C	MSMC
Sources	4.	13.50	9.70
Mixtures	2.53	14.67	7.64
KL	2.60	15.23	7.90
FastICA-dt	2.27	16.81	6.57
FastICA-st	3.49	16.56	10.35
JADE	3.10	16.01	10.72
SOBI-1D	1.44	16.30	4.92
SOBI-2D	1.49	16.25	6.63
f-SOBI-1D	3.35	16.37	10.78
f-SOBI-2D	3.22	15.80	13.68
FastICA-dt+	2.58	12.95	5.52
FastICA-st+	3.17	14.37	5.63
JADE+	2.80	13.87	5.96
SOBI-1D+	2.28	15.59	7.27
SOBI-2D+	2.33	15.40	7.31
f-SOBI-1D+	3.38	16.48	13.95
f-SOBI-2D+	3.30	15.89	13.22

TABLE IV

Correlation between the contrasts and the quality index for the noisy mixtures.

Criterion	Correlation
C	-0.196
MSMC	0.664

contains an information both on the HOS of the sources and their spatial organization. But we saw that this criterion was not determinant. The main progress came from the masking. By retaining only coefficients carrying out a relevant information a more selective criterion than both f-SOBI and JADE was derived.

G. The noise influence.

The noise reduces drastically the capabilities of the BSS algorithms. Then the performance of the empiral contrast with MSMC was tested on the simulated mixtures where a Gaussian noise was added so that the signal to noise ratio became around 15 dB. The results are given in Table III.

We remark that the original mixtures have a contrast greater than the one obtained after computing the sources from the original demixing matrix. The added noises were partially interpreted as independent sources, then their mixing decreases the contrast. With MSMC the situation is inverted, but the maximum of this contrast is not obtained for the original mixing matrix. In table IV the coefficients of correlation are given confirming the improvement of the BSS selection from MSMC.

V. Conclusion.

A new criterion for quantifying BSS was presented. It is based on the idea that the main information in the images was included in selected wavelet coefficients and that a multiscale masked contrast could be derived. The sparsity of the wavelet coefficients was elsewhere used for BSS [19], but in a very different way.

MSMC is a quantity which seems to be closely correlated to the separation quality index. It is evaluated on the sources without any knowledge on the real sources. Instead of using it for quantifying a BSS result it is possible also to exploit it as a criterion for computing the sources.

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] A. Bijaoui, D. Nuzillard. Blind source separation,smoothing and adaptive denoising. *IEEE Trans. on Signal Processing*, submited, 2001.

[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] R. R.Coifman, D. Donoho. Translation invariant de-noising. Technical report 475, Dpt of Statistics,Stanford Unversity, May 1995.

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

[7] D. Gabor. Theory of communication. *Journal of the IEE*, vol. 93, pp. 429-441, 1946.

[8] A. Haar. Zur theorie der orthogonalen funktionensysteme. *Math. Annal*, vol. 69, pp. 331-371, 1910.

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

[10] A. Hyvärinen A., E. Oja. Independent component analysis: a tutorial. *Neural Networks*, 13, pp. 411-430, 2000.

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

[12] J.-L. Lacoume, P.-O. Amblard, P.Comon. *Statistiques d'ordre supérieur pour le traitement du signal*, p.25, Masson, Paris, 1997.

[13] J.-L. Lacoume, P.-O. Amblard, P.Comon. *Statistiques d'ordre supérieur pour le traitement du signal*, p.22, Masson, Paris, 1997.

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

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

[16] D. Nuzillard, A. Bijaoui. Blind source separation and analysis of multispectral astronomical images. *Astron. Astrophys. Sup. Ser.*, v.147, pp.129-138, 2000.

[17] D. Nuzillard, A. Bijaoui. Multispectral analysis, blind source separation and mutual information. *IEEE Trans. on Image Proc*, submited, 2001.

[18] J.L. Starck, F. Murtagh, A. Bijaoui. *Image Processing and Data Analysis. The Multiscale Approach*, p.21, Cambridgde University Press, Cambridge, 1998.

[19] M. Zibulevsky, B. A.Pearlmutter. Blind separation of sources with sparse representations in a given signal dictionary. *Independent Component Analysis and Blind Source Separation*, pp. 181-186, ed. P. Pajunen and J. Karhunen, Helsinki University of Technology, 2000.