MULTICHANNEL MORPHOLOGICAL COMPONENT ANALYSIS

J. Bobin[a], Y. Moudden[a], J-L. Starck[a,b] and M. Elad[c]

[a] DAPNIA/SEDI-SAP
Service d'Astrophysique
91191 Gif-sur-Yvette France

[b] Laboratoire APC
11 place Marcelin Berthelot
75231 Paris Cedex 05

[c] The Computer Science Department
The Technion , Israel Institute of
Technology, Haifa 32000, Israel

ABSTRACT

We present in this paper a new method for blind source separation which is adapted to the case where the sources have different morphologies. We show that the morphology diversity concept leads to a new and very efficient method, even in the presence of noise. The algorithm, named Multichannel Morphological Component Analysis (MMCA) is an extension of the Morphological Component Analysis (MCA) method which was proposed for separating a single image into texture and piecewise smooth parts or for inpainting applications. A range of example illustrates the results.

1. INTRODUCTION

A common assumption in signal or image processing is that measurements $\mathbf{X}$ made typically using an array of sensors, often consists of mixtures of contributions from various possibly independent underlying physical processes $\mathbf{S}$. The simplest mixture model is linear and instantaneous and takes the form :

$$\mathbf{X} = \mathbf{A}\,\mathbf{S} \; + \; \mathbf{N} \tag{1}$$

where $\mathbf{X}$ and $\mathbf{S}$ are random vectors of respective sizes $m \times 1$ and $n \times 1$ and $\mathbf{A}$ is an $m \times n$ matrix. Multiplying $\mathbf{S}$ by $\mathbf{A}$ linearly mixes the n sources into m observed processes. In some cases, an $m \times 1$ random vector $\mathbf{N}$ is included to account for instrumental noise.The problem is then to invert the mixing process so as to separate the data back into its constitutive elementary building blocks. In a blind approach assuming minimal prior knowledge of the mixing process, source separation is merely about devising quantitative measures of diversity or contrast. Classical Independent Component Analysis (ICA) methods assume the mixed sources are statistically independent; this techniques have proven successful in a wide range of applications JADE, FastICA, Infomax (see [1, 2, 3, 4], and references therein). Indeed, although statistical independence is a strong assumption, it is in many cases physically plausible.

An especially important case is when the mixed sources are highly sparse, meaning that each source is only rarely active and mostly nearly zero. The independence assumption then ensures that the probability for two sources to be significant simultaneously is extremely low so that the sources may be treated as having nearly disjoint supports. This is exploited for instance in Sparse Component Analysis [5]. And it is shown in [6] that first moving the data into a representation in which the sources are assumed to be sparse will greatly enhance the quality of the separation. Possible dictionaries include Fourier and related bases, wavelet bases, etc. Working with combinations of several bases or with very redundant dictionaries such as undecimated wavelet frames or the more recent ridgelets, curvelets [7], etc. could lead to even more efficient representations. However, selecting from a large dictionary, the smallest subset of elements, that will linearly combine to reproduce a given signal or image, is a hard combinatorial problem. Nevertheless, several algorithms have been proposed that can help build very sparse decompositions [8, 9] and in fact, a number of recent results prove that these algorithms will recover the unique optimal decomposition provided this solution is sparse enough and the dictionary is sufficiently incoherent [10, 11].
Morphological Component Analysis (MCA) is a method described in [12] that constructs a sparse representation of a signal or an image considering that it is a combination of features which are sparsely represented in different dictionaries. For instance, images commonly combine contours and textures: the former are well accounted for using *e.g.* curvelets while the latter may be well represented using local cosine functions [13, 14, 15].
In searching a sparse decomposition of a signal or image s, MCA makes the specific assumption that s is a sum of K components φ_k where a possibly overcomplete dictionary $\mathbf{\Phi}_k$ is given for each k, in which φ_k admits a sparse representation, $\varphi_k = \mathbf{\Phi}_k \alpha_k$ while its sparsest decomposition over the other $\mathbf{\Phi}_{k' \neq k}$ is essentially diffuse. The different $\mathbf{\Phi}_k$ can be seen as acting as discriminant between the different components of the initial signal s. Ideally, the α_k are the solutions of:

$$\min_{\{\alpha_1,\ldots,\,\alpha_K\}} \sum_{k=1}^{K} \|\alpha_k\|_0 \quad \text{subject to} \quad s = \sum_{k=1}^{K} \mathbf{\Phi}_k \alpha_k. \tag{2}$$

However, the L_0 norm is non-convex and optimizing the above criterion is combinatorial by nature. Substituting an L_1 sparsity measure to the L_0 norm, as motivated by recent equivalence results *e.g.* in [10], and relaxing the equality constraint, the MCA algorithm seeks a solution to the following minimization problem:

$$\min_{\varphi_1,\ldots,\varphi_K} \sum_{k=1}^{K} \lambda_k \|\alpha_k\|_1 + \|s - \sum_{k=1}^{K} \varphi_k\|_2^2 \;\text{ with }\; \varphi_k = \mathbf{\Phi}_k \alpha_k \tag{3}$$

A detailed description of MCA is given in [12] along with results of experiments in contour/texture separation and image inpainting.

The purpose of this contribution is to extend MCA to the case of multi-channel data. This is described in section 2.

2. MULTICHANNEL MCA

We consider the mixing model (1) and make the additional assumption that each source s_k is well (*i.e.* sparsely) represented in a specific dictionary. Again, assigning a Laplacian prior with precision λ_k to the decomposition coefficients of the k^{th} source s_k in dictionary

Φ_k is a practical way to implement this property. Here, s_k denotes the $1 \times n$ array of the k^{th} source samples. Classically, we assume Gaussian white noise with known covariance Γ_n. This leads to the following joint estimator of the source processes $\mathbf{S} = \{s_1, \ldots, s_{n_s}\}$ and the mixing matrix $\mathbf{A}$:

$$\{\hat{\mathbf{S}}, \hat{\mathbf{A}}\} = \text{Arg} \min_{\mathbf{S},\mathbf{A}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|^2_{2,\mathbf{\Gamma}_n} + \sum_k \lambda_k \|s_k \mathbf{T}_k\|_1 \qquad (4)$$

where $\|\mathbf{M}\|^2_{2,\mathbf{\Gamma}_n} = \text{trace}\left(\mathbf{M}^T \mathbf{\Gamma}_n^{-1} \mathbf{M}\right)$. Unfortunately, this minimization problem suffers from the lack of scale invariance of the objective function. Indeed, combining a scaling of the mixing matrix, $A \leftarrow \rho A$, and an inverse scaling of the source matrix, $S \leftarrow \frac{1}{\rho} S$, leaves the quadratic measure of fit unchanged whereas the term measuring sparsity is deeply altered by the same inverse scale factor $\frac{1}{\rho}$. Consequently, the minimization will probably drive us to trivial solutions, $A \to \infty$ and $S \to 0$, since the sparsity term can be minimized *ad libitum* as ρ goes to $+\infty$. Nevertheless, scale-invariance can be artificially recovered by normalizing the columns a^k of the mixing matrix $\mathbf{A}$ at each iteration ($a^{k+} \leftarrow a^{k-}/\|a^{k-}\|_2$) and propagating the scale factor to the corresponding source, $s_k{}^+ \leftarrow \|a^{k-}\|_2 s_k{}^-$, and precision $\lambda_k{}^+ \leftarrow \|a^{k-}\|_2 \lambda_k{}^-$.

Define the k^{th} multichannel residual $\mathbf{D}_k = \mathbf{X} - \sum_{k' \neq k} a^{k'} s_{k'}$ as corresponding to the part of the data unexplained by the other couples $\{a^{k'}, s_{k'}\}_{k' \neq k}$. Then, the minimization problem (4) is equivalent to jointly minimizing the following set of elementary criteria:

$$\forall k, \{\hat{s}_k, \hat{a}^k\} = \text{Arg} \min_{s_k, a^k} \|\mathbf{D}_k - a^k s_k\|^2_{2,\mathbf{\Gamma}_n} + \lambda_k \|s_k \mathbf{T}_k\|_1 \qquad (5)$$

Zeroing the gradient with respect to s_k and a^k of this criterion leads to the following coupled equations:

$$\left\{ \begin{array}{lcl} s_k & = & \frac{1}{a^{k^T} \mathbf{\Gamma}_n^{-1} a^k} \left(a^{k^T} \mathbf{\Gamma}_n^{-1} \mathbf{D}_k - \frac{\lambda_k}{2} \text{Sign}(s_k \mathbf{T}_k) \mathbf{R}_k \right) \\ a^k & = & \frac{1}{s_k s_k{}^T} \mathbf{D}_k s_k{}^T \end{array} \right. \qquad (6)$$

Although the above holds for unitary transforms for which $\mathbf{R}_k = \mathbf{T}_k^T$, we make the same approximation as in the previous section and consider that it is still valid for redundant transforms. Then, for a fixed a^k, the source process s_k is estimated by soft-thresholding the coefficients of the decomposition of a *coarse version* $\tilde{s_k} = (1/a^{k^T} \mathbf{\Gamma}_n^{-1} a^k) a^{k^T} \mathbf{\Gamma}_n^{-1} \mathbf{D}_k$ with threshold $\lambda_k / (2 a^{k^T} \mathbf{\Gamma}_n^{-1} a^k)$. Considering a fixed s_k, the update on a^k follows from a simple least squares linear regression. The MMCA algorithm is given below :

1. Set # of iterations $L_{\max}$ & thresholds $\forall k, \delta_k = L_{\max} \cdot \lambda_k / 2$
2. While $\delta_k > \lambda_k / 2$,
   For $k = 1, \ldots, n_s$:
   - Renormalize a^k, s_k and δ_k
   - Update of s_k assuming all $s_{k' \neq k}$ and $a^{k'}$ are fixed:
     – Compute the residual $\mathbf{D}_k = \mathbf{X} - \sum_{k' \neq k} a^{k'} s_{k'}$
     – Project $\mathbf{D}_k$: $\tilde{s}_k = \frac{1}{a^{k^T} \mathbf{\Gamma}_n^{-1} a^k} a^{k^T} \mathbf{\Gamma}_n^{-1} D_k$
     – Compute $\alpha_k = \tilde{s}_k \mathbf{T}_k$
     – Soft threshold α_k with threshold $= \delta_k$ gives $\hat{\alpha}_k$
     – Reconstruct s_k by $s_k = \hat{\alpha}_k \mathbf{R}_k$
   - Update of a^k assuming all $s_{k'}$ and $a^{k' \neq k}$ are fixed:
     $a^k = \frac{1}{s_k s_k{}^T} \mathbf{D}_k s_k{}^T$

   Lower the thresholds: $\delta_k = \delta_k - \lambda_k / 2$.

At each iteration, a *coarse* (and thus noise free) version of the sources are computed. The mixing matrix is then estimated from noise-free sources which only contains the most significant parts of the original sources and not their mixtures. The overall optimization leads to refine both the noise-free sources and the mixing matrix. The use of an iterative thresholding with a set of thresholds $\{\delta_k\}_{k=1,c_d ots,n}$ decreasing slowly guarantees robustness. Indeed, both alternate projections and iterative thresholding define a compelled path for the variables to estimate (sources and mixing matrix) during the optimization. This optimization scheme might lead to a not so bad estimation according to the MMCA hypothesis which stipulate that different sources are sparsified in different basis.

The next section will illustrate the efficiency of the MMCA algorithm when the sources to separate are morphologically different enough.

3. EXPERIMENTS

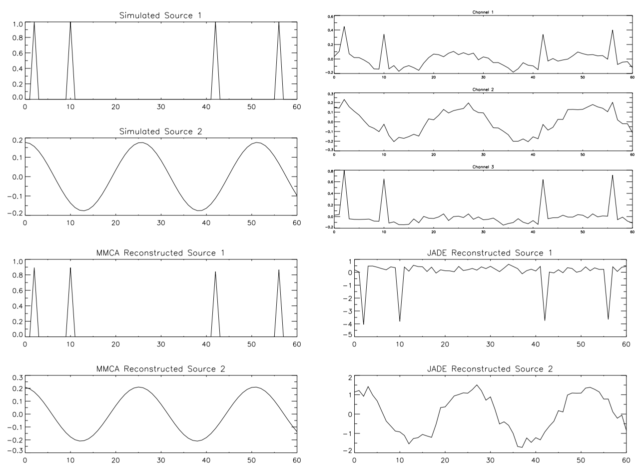

Fig. 1. **top left :** the two initial source signals. **top right :** three noisy *observed* mixtures. **bottom left :** the two source signals reconstructed using MMCA. **bottom right :** the two source signals reconstructed with Jade.

MMCA is clearly able to efficiently separate the initial source signals. Note that denoising is an intrinsic part of the algorithm.

Fig. 2 shows a similar experiment with 2D data. From top to bottom, we see respectively the two initial source images, two noisy mixtures, the two images reconstructed with Jade and the two images reconstructed using MMCA. In this case, two transforms using in MMCA were the isotropic wavelet transform [16] and the curvelet transform [7]. The first one is well suited for representing Gaussians and the second one represents well anisotropic features.

In figure 3, the two left pictures are the sources. The first one is composed of a set several pointwise singularities which are well represented - according to the sparsity sense - in a wavelet dictionary. The second one is a kind of "cloud" spatially diffused over the whole picture. The latter is well sparsified using a global Discrete Cosine Transform. The mixtures are pictured in the second couple of images (i.e. second column) . An anisotropic (not the same noise variance in each channel) Gaussian noise has been added.

For the sake of comparison with standard source separation tech-

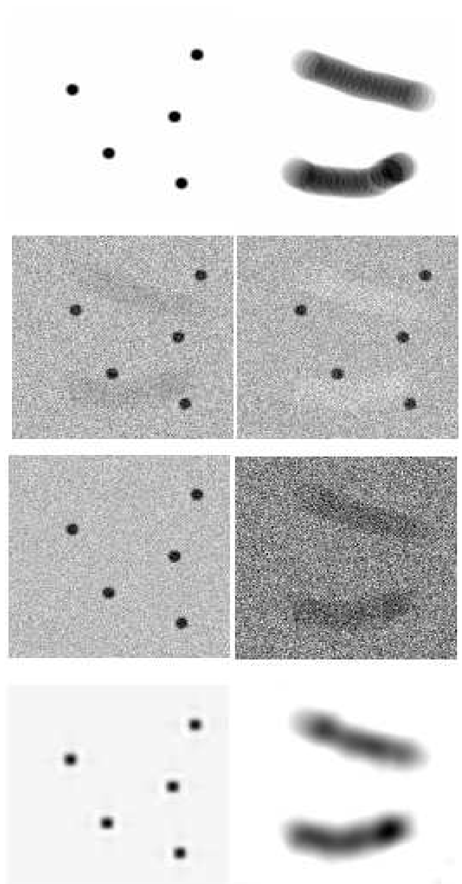

Fig. 2. From top to bottom. the two initial source images, two noisy mixtures, the two images reconstructed with Jade and the two images reconstructed using MMCA.

niques, the third couple of pictures shows the sources estimated by the well-known JADE. As the genuine JADE algorithm has not been devised to take into account additive noise, we decided to compare our results with a denoised version of the JADE estimates; thus the estimated sources has been denoised using a standard undecimated wavelet denoising technique assuming that the noise variances are known. Note that we could have denoised the data before separating; the non-linear wavelet denoising erasing the coherence between the channels, an ICA-based method fails to separate sources from denoised data. The fourth couple of pictures illustrates this latter point. Finally the two last images shows the MMCA source estimates. Visually the MMCA results are better than those of JADE and JADE denoised estimates.

To get convinced by the efficiency of MMCA in a noisy context, we choose to compare our method (with the latter example) with well-known source separation methods such as JADE and FASTICA [1] and denoised versions of their estimates. Figures 4 shows the correlation between the genuine sources and their estimates as the data noise variance increases. One can note that both JADE and FASTICA have similar performances. In the low noise case, MMCA performs as well as standard methods. As the data noise variance increases, MMCA clearly achieves better source estimation. In fact, when the variance of the data noise increases, the ICA-based estimated mixing matrix are biased and thus the standard methods fails to correctly estimate the sources. Quantitative results are shown in figure 5. It represents a mixing matrix criterion defined by $\rho_A = ||I - \Lambda\tilde{A}^{-1}A||_1$ to assess the mixing matrix estimation quality as the data noise variance increases. A is the true mixing matrix, $\tilde{A}$ is the estimated one and Λ is a matrix which zeros the effect of scaling and permutation on the estimated matrix. If $\tilde{A} = \Lambda A$ (*i.e* $\tilde{A}$ is equal to A up to scaling and permutation) then $\rho_A = 0$; thus ρ_A measures a deviation from the true estimate.

In opposition to standard ICA methods, MMCA iteratively estimates the mixing matrix A from coarse (noise-free) version of the sources and thus is not penalized by the presence of noise.

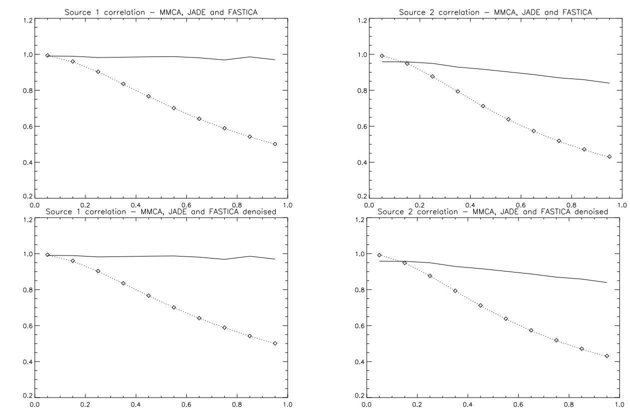

Fig. 4. **Two first graphs:** Correlation between the true source 1 and 2 (respectively) and the sources estimated by JADE (*dotted line*), FASTICA ($\diamond$) and MMCA.**Two last graphs:** Correlation between the true source 1 and 2 (respectively) and the sources estimated by JADE denoised (*dotted line*), FASTICA denoised ($\diamond$) and MMCA. **Abscissa :** value of σ such that the data noise variance is equal to 0.8σ in the first channel et $1,2\sigma$ in the second one. **Ordinate :** correlation coefficient.

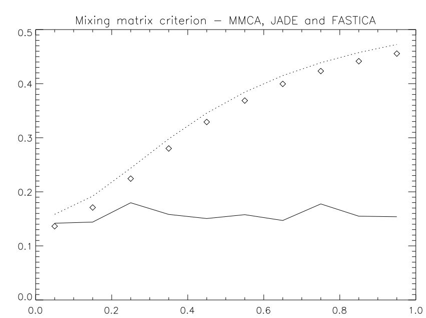

Fig. 5. Mixing matrix criterion assessing the efficiency of the mixing matrix estimation with JADE (*dotted line*), FASTICA ($\diamond$) and MMCA. Abscissa: value of σ such that the data noise variance is equal to 0.8σ in the first channel et $1,2\sigma$ in the second one. Ordinate : ρ_A (see text)

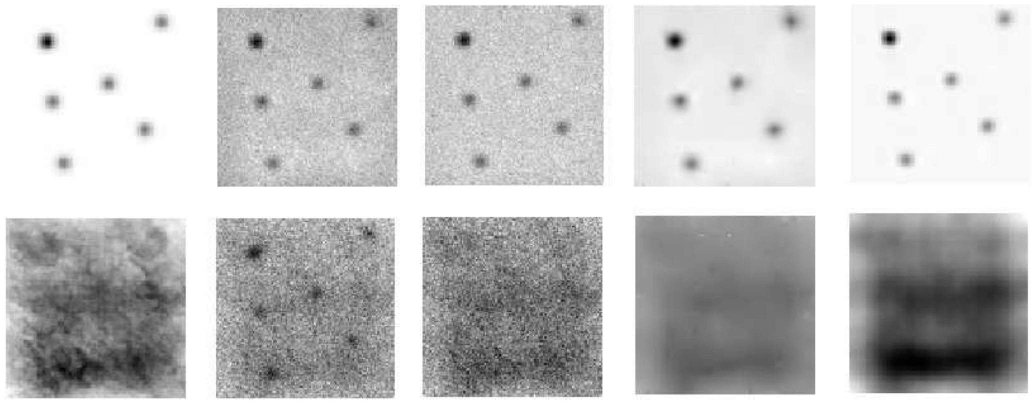

Fig. 3. **From left to right :** the original sources, their mixtures (a Gaussian noise is added : $\sigma = 0.4$ and 0.6 for channel 1 and 2 respectively - the mixtures are such that $x_1 = 0,5s_1 - 0,5s_2$ and $x_1 = 0,3s_1 + 0,7s_2$), the sources estimated by JADE, their denoised version and the sources estimated using MMCA.

4. CONCLUSION

The MMCA algorithm described in this paper extends MCA to the multichannel case. For blind source separation, this extension is shown to perform well provided the original sources are morphologically different meaning that the sources are sparsely represented in different bases. We also demonstrated that MMCA performs better than standard ICA-based source separation in noisy context. The next step might be an extension of the MCA philosophy to separate sources which might be morphologically similar.

5. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley, New York, 2001, 481+xxii pages.

[2] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[3] Dinh-Tuan Pham and Jean-François Cardoso, "Blind separation of instantaneous mixtures of non stationary sources," *IEEE Trans. on Sig. Proc.*, vol. 49, no. 9, pp. 1837–1848, Sept. 2001.

[4] Jacques Delabrouille, Jean-François Cardoso, and Guillaume Patanchon, "Multi–detector multi–component spectral matching and applications for CMB data analysis," *Monthly Notices of the Royal Astronomical Society*, vol. 346, no. 4, pp. 1089–1102, Dec. 2003, to appear, also available as http://arXiv.org/abs/astro-ph/0211504.

[5] P. G. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 992–996, 2005.

[6] M. Zibulevsky and B.A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural-Computation*, vol. 13, no. 4, pp. 863–882, April 2001.

[7] J.-L. Starck, E. Candès, and D.L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 131–141, 2002.

[8] S.G. Mallat and Z. Zhang, "Atomic decomposition by basis pursuit," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, 1993.

[9] S.S. Chen, D.L. Donoho, and M.A. Saunder, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[10] D. L. Donoho and M. Elad, "Maximal sparsity representation via l_1 minimization," *the Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, 2003.

[11] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[12] J.-L. Starck, M. Elad, and D.L. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," *Advances in Imaging and Electron Physics*, vol. 132, 2004.

[13] J.-L. Starck, M. Elad, and D.L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[14] M. Elad, J.-L Starck, D. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *ACHA*, 2005, to appear.

[15] F. Fadili and J.-L. Starck, "EM algorithm for sparse representation-based image inpainting," in *IEEE International Conference on Image Processing ICIP'05*. 2005, IEEE.

[16] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press, 1998.