Image decomposition and separation using sparse representations: an overview

M.J. Fadili , J.-L. Starck , J. Bobin and Y. Moudden

Abstract

This paper gives essential insights into the use of sparsity and morphological diversity in image decomposition and source separation by overviewing our recent work in this field. The idea to morphologically decompose a signal into its building blocks is an important problem in signal processing and has far-reaching applications in science and technology. Starck et al. [1], [2] proposed a novel decomposition method - Morphological Component Analysis (MCA) - based on sparse representation of signals. MCA assumes that each (monochannel) signal is the linear mixture of several layers, the so-called Morphological Components, that are morphologically distinct, e.g. sines and bumps. The success of this method relies on two tenets: sparsity and morphological diversity. That is, each morphological component is sparsely represented in a specific transform domain, and the latter is highly inefficient in representing the other content in the mixture. Once such transforms are identified, MCA is an iterative thresholding algorithm that is capable of decoupling the signal content. Sparsity and morphological diversity have also been used as a novel and effective source of diversity for blind source separation (BSS), hence extending the MCA to multichannel data. Building on these ingredients, we will overview the Generalized MCA (GMCA) introduced by the authors in [3], [4] as a fast and efficient BSS method. We will illustrate the application of these algorithms on several real examples. We conclude our tour by briefly describing our software toolboxes made available for download on the Internet for sparse signal and image decomposition and separation.

Index Terms

Sparse representations, image decomposition, blind source separation, morphological component analysis.

M.J. Fadili is with GREYC CNRS-ENSICAEN-University of Caen, Image Processing Group, 14050 Caen Cedex, France.

J.-L. Starck is with AIM, CEA/DSM-CNRS-University Paris Diderot, IRFU, SEDI-SAP, F-91191 Gif-Sur-Yvette Cedex, France.

J. Bobin is with the Department of Applied and Computational Mathematics, CalTech, CA 91125, USA.

Y. Moudden is with DSM /IRFU/SEDI, CEA/Saclay, F-91191 Gif-sur-Yvette, France.

I. INTRODUCTION

Although Mathematics has it million-dollar problems, in the form of Clay Math Prizes, there are several Billion Dollar problems in signal and image processing. Famous ones include the cocktail party problem (separate a speaker voice from a mixture of other recorded voices and background sounds at a cocktail party). These signal processing problems seem to be intractable according to orthodox arguments based on rigorous mathematics, and yet they keep cropping up in problem after problem.

One such fundamental problem involves decomposing a signal or image into superposed contributions from different sources; think of symphonic music which may involve superpositions of acoustic signals generated by many different instruments – and imagine the problem of separating these contributions. More abstractly, we can see many forms of media content that are superpositions of contributions from different 'content types', and we can imagine wanting to separate out the contributions from each. We can easily see a fundamental problem; for example an N-pixel image created by superposing K different types offers us N data (the pixel values) but there may be as many as $N \cdot K$ unknowns (the contribution of each content type to each pixel). Traditional mathematical reasoning, in fact the fundamental theorem of linear algebra, tells us not to attempt this: there are more unknowns than equations. On the other hand, if we have prior information about the underlying object, there are some rigorous results showing that such separation might be possible – using a sparsity prior.

The idea to morphologically decompose a signal into its building blocks is an important problem in signal and image processing. Successful methods for signal or image separation can be applied in a broad range of areas in science and technology including biomedical engineering, medical imaging, speech processing, astronomical imaging, remote sensing, communication systems, etc. An interesting and complicated image content separation problem is the one targeting decomposition of an image to texture and piece-wise-smooth (cartoon) parts. A functional-space characterization of oscillating textures was proposed in [5], and was used for variational cartoon+texture image decomposition [6]. Since then, we have witnessed a flurry of research activity in this application field.

In [1], [2], the authors proposed a novel decomposition method – Morphological Component Analysis (MCA) – based on sparse representation of signals. MCA assumes that each signal is the linear mixture of several layers, the so-called Morphological Components, that are morphologically distinct, e.g. sines and bumps. The success of this method relies on the assumption that for every component behavior to be separated, there exists a dictionary of atoms that enables its construction using a sparse representation. It is then assumed that each morphological component is sparsely represented in a specific transform domain. And when all transforms (each one attached to a morphological component) are amalgamated in one dictionary, each one must lead to sparse representation over the part of the signal it is serving, while being highly inefficient in

representing the other content in the mixture. If such dictionaries are identified, the use of a pursuit algorithm searching for the sparsest representation leads to the desired separation. MCA is capable of creating atomic sparse representations containing as a by-product a decoupling of the signal content.

Over the last few years, the development of multi-channel sensors motivated interest in methods for the coherent processing of multivariate data. Consider a situation where there is a collection of signals emitted by some physical sources. These could be, for example, different brain areas emitting electric signals; people speaking in the same room, thus emitting speech signals; or radiation sources emitting their electromagnetic waves. Assume further that there are several sensors or receivers. These sensors are in different positions, so that each records a mixture of the original source signals with different weights. The so-called blind source separation (BSS) problem is to find the original sources or signals from their observed mixtures, without prior knowledge of the mixing weights, and by knowing very little about the original sources. Some specific issues of BSS have already been addressed as testified by the wide literature in this field. In this context, as clearly emphasized by previous work, it is fundamental that the sources to be retrieved present some quantitatively measurable diversity or contrast (e.g. decorrelation, independence, morphological diversity, etc). The seminal work of [7], [8] paved the way for the use of sparsity in BSS. Recently, sparsity and morphological diversity have emerged as a novel and effective source of diversity for BSS both for underdetermined and overdetermined BSS; see the comprehensive review in [4]. Building on the sparsity and morphological diversity ingredients, the authors proposed the Generalized MCA (GMCA) as a fast and efficient multichannel sparse data decomposition and BSS method [3], [4], [9].

A. Organization of the paper

Our intent in this paper is to overview the recent work in monochannel image decomposition and multichannel source separation based on the concepts of sparsity and morphological diversity. The first part of the paper is devoted to monochannel sparse image decomposition, and the second part to blind sparse source separation. In this review, our goal is to highlight the essential concepts and issues, and to describe the main algorithms. Several applications to real data are given in each part to illustrate the capabilities of the proposed algorithms. We conclude our tour by providing pointers to our software toolboxes that implement our algorithms and reproduce the experiments on sparse signal and image decomposition and source separation.

B. Notations

The ℓ_p-norm of a (column or row) vector x is $\|x\|_p := \left(\sum_i |x[i]|^p \right)^{1/p}$ with the usual adaptation when $p = \infty$, and $\|x\|_0 := \lim_{p \to 0} \|x\|_p^p$ is the ℓ_0 pseudo-norm, i.e. the number of nonzero components. Bold symbols represent matrices and $\mathbf{X}^T$ is the transpose of $\mathbf{X}$. The Frobenius norm of $\mathbf{X}$ is $\|\mathbf{X}\|_F = \text{Trace} \left(\mathbf{X}^T \mathbf{X} \right)^{1/2}$. The k-th entry of y_i (resp. y^j) is $y_i[k]$ (resp. $y^j[k]$), where y_i is the i-th row and y^j is the j-th column of $\mathbf{Y}$. An

atom is an elementary signal-representing template. Examples might include sinusoids, monomials, wavelets, Gaussians. A dictionary $\mathbf{\Phi} = [\phi^1, \cdots, \phi^L]$ defines a $N \times L$ matrix whose columns are unit ℓ_2-norm atoms ϕ^i. When the dictionary has more columns than rows, it is called overcomplete or redundant. We are mainly interested here in overcomplete dictionaries.

II. MONOCHANNEL SPARSE IMAGE DECOMPOSITION

A. Morphological Component Analysis

Suppose that the N-sample signal or image x is the linear superposition of K morphological components, possibly contaminated with noise

$$y = \sum_{k=1}^{K} x_k + \varepsilon, \qquad \sigma_\varepsilon^2 = \text{Var}[\varepsilon] < +\infty \,.$$

The MCA framework aims at solving the inverse problem that consists in recovering the components $(x_k)_{k=1,\cdots,K}$ from their observed linear mixture as illustrated in the top of Fig. 1. MCA assumes that each component x_k can be sparsely represented in an associated basis $\mathbf{\Phi}_k$, i.e.

$$x_k = \mathbf{\Phi}_k \alpha_k, \quad k = 1, \ldots, K \,,$$

where α_k is a sparse coefficient vector (sparse means that only a few coefficients are large and most are negligible). Thus, a dictionary can be built by amalgamating several transforms $(\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_K)$ such that, for each k, the representation of x_k in $\mathbf{\Phi}_k$ is sparse and not, or at least not as sparse, in other $\mathbf{\Phi}_l, l \neq k$. In other words, the sub-dictionaries $(\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_K)$ must be mutually incoherent. Thus, the dictionary $\mathbf{\Phi}_k$ plays a role of a discriminant between content types, preferring the component x_k over the other parts. This is a key observation for the success of the separation algorithm. Owing to recent advances in computational harmonic analysis, many novel representations, including the wavelet transform, curvelet, contourlet, steerable or complex wavelet pyramids, were shown to be very effective in sparsely representing certain kinds of signals and images. Thus, for decomposition purposes, the dictionary will be built by taking the union of one or several (sufficiently incoherent) transforms, generally each corresponding to an orthogonal basis or a tight frame.

However, the augmented dictionary $\mathbf{\Phi} = [\mathbf{\Phi}_1 \cdots \mathbf{\Phi}_K]$ will provide an overcomplete representation of x. Because there are more unknowns than equations, the system $x = \Phi\alpha$ is underdetermined. Sparsity can be used to find a unique solution, in some idealized cases; there is an extensive literature on the subject and the interested reader may refer to the comprehensive review paper [10].

In [1], [2], it was proposed to solve this underdetermined system of equations and recover the morphological components $(x_k)_{k=1,\cdots,K}$ by solving the following constrained optimization problem:

$$\min_{\alpha_1, \cdots, \alpha_K} \sum_{k=1}^{K} \|\alpha_k\|_p^p \ \text{ such that } \ \left\| y - \sum_{k=1}^{K} \mathbf{\Phi}_k \alpha_k \right\|_2 \leq \tau \,, \tag{1}$$
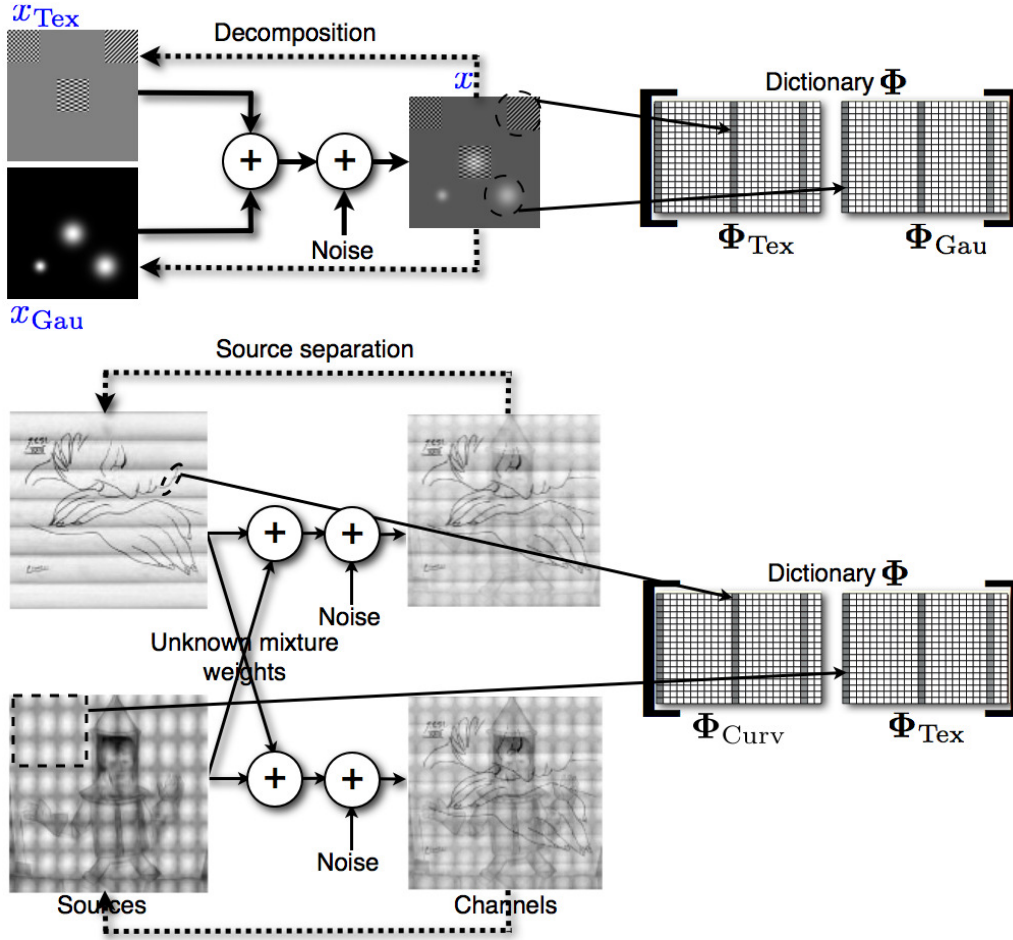
Fig. 1. Illustration of the image decomposition (top) and the blind source separation (BSS) problems (bottom) using sparsity and morphological diversity. For the bottom part, each source is itself a mixture of morphological components (see (4)) to be isolated.

where $\|\alpha\|_p^p$ is the penalty quantifying sparsity (the most interesting regime is for $0 \leq p \leq 1$), and τ is typically chosen as a constant times $\sqrt{N}\sigma_\epsilon$. The constraint in this optimization problem accounts for the presence of noise and model imperfection. If there is no noise and the linear superposition model is exact ($\tau = 0$), an equality constraint is substituted for the inequality constraint. This formulation is flexible enough to incorporate external forces that direct the morphological components to better suit their expected content, these forces will fine tune the separation process to achieve its task. As an example for such successful external force, [1] proposed to add a Total Variation (TV) penalty [11] to the cartoon part in order to direct this component to fit the piecewise-smooth model.

B. MCA algorithm

Problem (1) is not easy to solve in general, especially for $p < 1$ (for $p = 0$ it is even NP-hard). Nonetheless, if all components $x_l = \Phi_l \alpha_l$ but the k-th are fixed, then it can be proved that the solution α_k is given by

hard thresholding (for $p = 0$) or soft thresholding (for $p = 1$) the marginal residuals $r_k = y - \sum_{l \neq k} \Phi_l \alpha_l$. These marginal residuals r_k are relieved from the other components, and are likely to contain mainly the salient information of x_k. This intuition dictates a coordinate relaxation algorithm that cycles through the components at each iteration and applies a thresholding to the marginal residuals. This is what justifies the steps of the MCA algorithm summarized in Algorithm 1, where $\mathrm{TH}_\lambda(\alpha)$ denotes component-wise thresholding with threshold λ: hard thresholding $\mathrm{HT}_\lambda(u) = u$ if $|u| > \lambda$ and 0 otherwise, or soft-thresholding $\mathrm{ST}_\lambda(u) = u \max(1 - \lambda / |u|, 0)$.

Algorithm 1: MCA decomposition algorithm.

Task: Signal/image decomposition.

Parameters: The signal/image x, the dictionary $\Phi = [\Phi_1 \cdots \Phi_K]$, number of iterations N_{iter}, stopping threshold $\lambda_{\min}$, threshold update schedule.

Initialization:
- Initial solution $x_k^{(0)} = 0, \forall k$.
- Initial residual $r^{(0)} = y$.
- Initial threshold: let $k^\star = \max_k \left\| \Phi_k^T y \right\|_\infty$, set $\lambda^{(0)} = \max_{k \neq k^\star} \left\| \Phi_k^T y \right\|_\infty$.

Main iteration:

for $t = 1$ **to** N_{iter} **do**

    for $k = 1$ **to** K **do**

        Compute marginal residuals $r_k^{(t)} = r^{(t-1)} + x_k$.

        Update k−th component coefficients by thresholding $\alpha_k^{(t)} = \mathrm{TH}_{\lambda^{(t-1)}} \left(\Phi_k^T r_k^{(t)} \right)$.

        Update k−th component $x_k^{(t)} = \Phi_k \alpha_k^{(t)}$.

    Update the residuals $r^{(t+1)} = y - \sum_{k=1}^K x_k^{(t)}$.

    Update the threshold $\lambda^{(t)}$ according to the given schedule.

    if $\lambda^{(t)} \leq \lambda_{\min}$ **then** stop.

Output: Morphological components $\left(x_k^{(N_{\mathrm{iter}})} \right)_{k=1,\cdots,K}$.

Beside coordinate relaxation, another important ingredient of MCA is iterative thresholding with *varying threshold*. Thus, MCA can be viewed as a stagewise hybridization of matching pursuit (MP) [12] with block-coordinate relaxation (BCR) [13] to (approximately) solve (1). The adjective stagewise is because MCA exploits the fact that the dictionary is structured (union of transforms), and the atoms enter the solution by groups rather than individually unlike MP. As such, MCA is a *salient-to-fine* process where at each iteration, the most salient content of each morphological component is iteratively computed. These estimates are then progressively refined as the threshold λ decreases towards $\lambda_{\min}$.

In the noiseless case, a careful analysis of the recovery properties (uniqueness and support recovery) of MCA algorithm and its convergence behavior when all Φ_k are orthobases can be found in [14], [9].

C. Dictionary choice

From a practical point of view, given a signal x, we will need to compute its forward (or analysis) transform by multiplying it by $\mathbf{\Phi}^T$. We also need to reconstruct any signal from its coefficients α. In fact, the matrix $\mathbf{\Phi}$ and its adjoint $\mathbf{\Phi}^T$ corresponding to each transform are never explicitly constructed in memory. Rather they are implemented as fast implicit analysis and synthesis operators taking a signal vector x, and returning $\mathbf{\Phi}^T x = \mathbf{T} x$ (analysis side), or taking a coefficient vector α and returning $\mathbf{\Phi}\alpha$ (synthesis side). In the case of a simple orthogonal basis, the inverse of the analysis transform is trivially $\mathbf{T}^{-1} = \mathbf{\Phi}$; whereas assuming that $\mathbf{\Phi}$ is a tight frame implies that the frame operator satisfies $\mathbf{\Phi}\mathbf{\Phi}^T = c\mathbf{I}$, where $c > 0$ is the tight frame constant. Hence, $\mathbf{T}^T = \mathbf{\Phi}$ is the Moore-Penrose pseudo-inverse transform (corresponding to the minimal dual synthesis frame), up to the constant c. In other words, the pseudo-inverse reconstruction operator $\mathbf{T}^+$ corresponds to $c^{-1}\mathbf{\Phi}$. It turns out that $\mathbf{T}^+\alpha$ is the reconstruction operation implemented by most implicit synthesis algorithms.

Choosing an appropriate dictionary is a key step towards a good sparse decomposition. Thus, to represent efficiently isotropic structures, a qualifying choice is the wavelet transform [15], [16]. The curvelet system [17] is a very good candidate for representing piecewise smooth (C^2) images away from C^2 contours. The ridgelet transform [18] has been shown to be very effective for sparsely representing global lines in an image. For locally oscillating textures, one can think of the local DCT (discrete cosine transform) [15], waveatoms [19] or brushlets [20]. These transforms are also computationally tractable particularly in large-scale applications, and as stated above, never explicitly implement $\mathbf{\Phi}$ and $\mathbf{T}$. The associated implicit fast analysis and synthesis operators have typical complexities of $O(N)$, with N the number of samples or pixels (e.g. orthogonal or bi-orthogonal wavelet transform) or $O(N \log N)$ (e.g. ridgelets, curvelets and local DCT, waveatoms).

What happens if none of the known fixed transforms can efficiently sparsify a morphological component; e.g. a complex natural texture. In [21], the authors have extended the MCA algorithm to handle the case where the dictionary attached to each morphological component is not necessarily fixed *a priori* as above, but learned from a set of exemplars in order to capture complex textural patterns.

D. Thresholding strategy

In practice, hard thresholding leads to better results. Furthermore, in [14], we empirically showed that the use of hard thresholding is likely to provide the ℓ_0-sparsest solution. As far as the thresholding decreasing strategy is concerned, there are several alternatives. For example, in [1], [2], linear and exponential decrease were advocated. In [14], a more elaborated strategy coined MOM (for mean-of-max) was proposed.

E. Handling bounded noise

MCA handles in a natural way data perturbed by additive noise ε with bounded variance σ_ε^2. Indeed, as MCA is a coarse-to-fine iterative procedure, bounded noise can be handled just by stopping iterating when the residual is at the noise level. Assuming that the noise variance σ_ε^2 is known, the algorithm may be stopped at iteration t when the ℓ_2-norm of the residual satisfies $\left\| r^{(t)} \right\|_2 \leq \sqrt{N} \sigma_\varepsilon$. Alternatively, one may use a strategy reminiscent of denoising methods, by taking $\lambda_{\min} = \tau \sigma_\varepsilon$ where τ is a constant, typically between 3 and 4.

F. Applications

Fig. 2 shows examples of application of the MCA sparse decomposition algorithm to three real images: Barbara (a)-(c), X-ray riser image (d)-(f) and an astronomical image of the galaxy SBS 0335-052 (g)-(j). The riser in Fig. 2(d) is made of a composite material layer, a layer of steel-made fibers having opposite lay angles, and lead-made markers used as a reference to calibrate the X-ray camera. The observed image is corrupted by noise. The structures of interest are the curvilinear fibers. The astronomical image of Fig. 2(g) is contaminated by noise and a stripping artifact, the galaxy of interest is hardly visible in the original data.

The dictionaries used for the three images are respectively: local DCT+curvelets for Barbara to decompose it into cartoon+texture parts; translation invariant wavelets+curvelets for the riser image; ridgelets+curvelets+translation invariant wavelets for the astronomical image. The details of the experimental set-up including the parameters of the dictionaries for each image are detailed in [22]. From Fig. 2(e)-(f), one can clearly see how MCA managed to get rid of the lead-made markers while reconstructing the curvilinear fibers structure. In Fig. 2(j), the galaxy has been well detected in the wavelet space while the stripping artifact was remarkably captured and removed owing to ridgelets and curvelets.

III. MULTICHANNEL SPARSE SOURCE SEPARATION

A. The blind source separation problem

In the blind source separation (BSS) setting, the instantaneous linear mixture model assumes that we are given m observations (channels) $\{y_1, \cdots, y_m\}$ where each y_j is a *row-vector* of size N; each channel is the linear mixture of n sources

$$\forall j \in \{1, \cdots, m\}, \qquad y_j = \sum_{i=1}^{n} a_j[i] s_i + \varepsilon_j , \tag{2}$$

or equivalently in matrix form

$$\mathbf{Y} = \mathbf{AS} + \mathbf{E} , \tag{3}$$

where $\mathbf{Y} = [y_1^T, \cdots, y_m^T]^T$ is the $m \times N$ measurement matrix whose rows are the channels y_j's, $\mathbf{S} = [s_1^T, \cdots, s_n^T]^T$ is the $n \times N$ source matrix whose rows are the sources s_i, and $\mathbf{A}$ is the $m \times n$ mixing matrix.
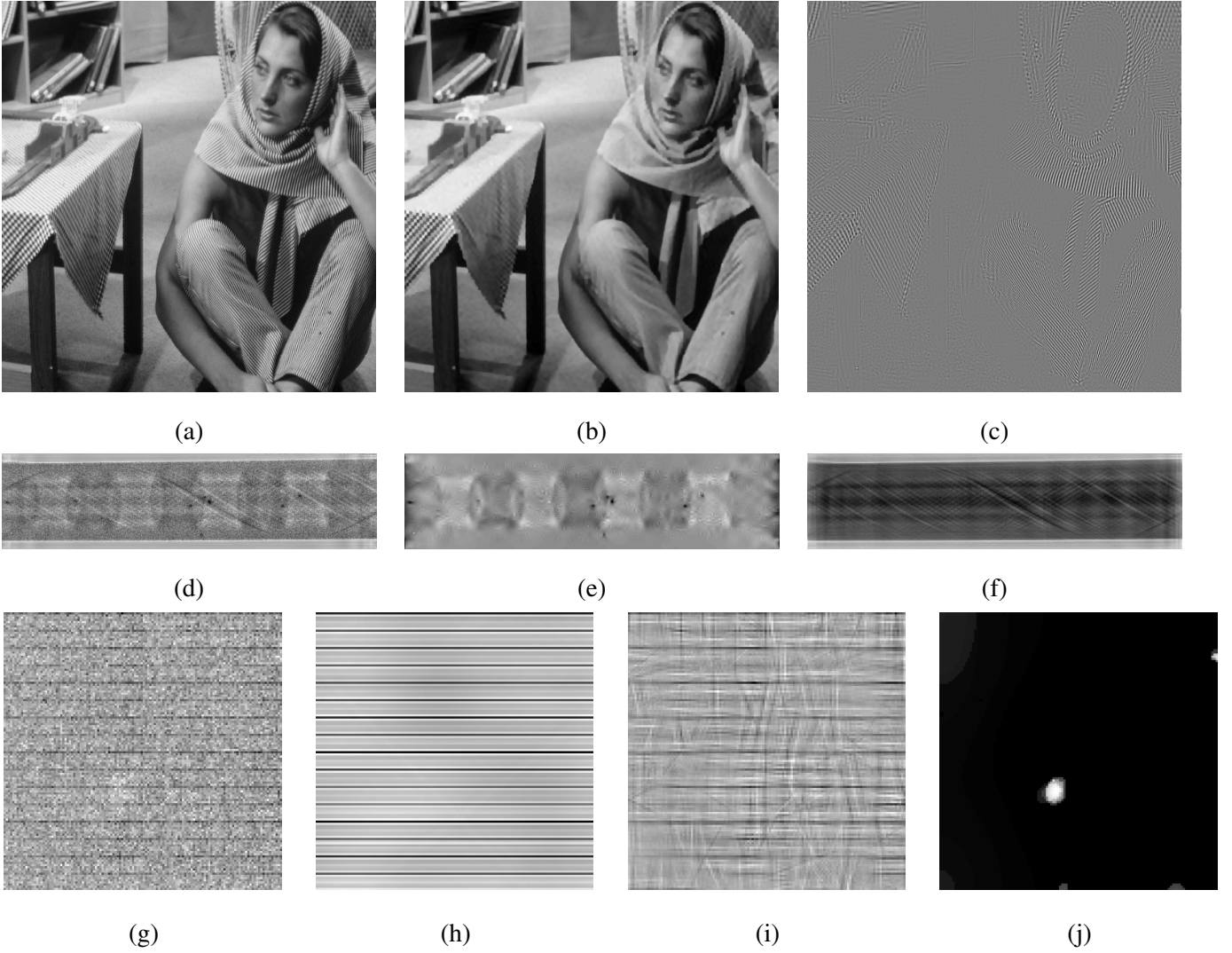
Fig. 2. MCA of three real 2D images. Barbara: (a) Original, (b) cartoon component (curvelets), (c) texture (local DCT). X-ray riser image: (d) Observed, (e) isotropic structures and background (wavelets), (f) curvilinear fibers (curvelets). Galaxy SBS 0335-052: (g) Observed, (h) ridgelet component, (i) curvelet component, (j) detected galaxy (wavelets).

$\mathbf{A}$ defines the contribution of each source to each channel. An $m \times N$ matrix $\mathbf{E}$ is added to account for instrumental noise or model imperfections. See Fig. 1.

Source separation techniques aim at recovering the original sources $\mathbf{S}$ by taking advantage of some information contained in the way the signals are mixed. In the blind approach, both the mixing matrix $\mathbf{A}$ and the sources $\mathbf{S}$ are unknown. Source separation is overwhelmingly a question of contrast and diversity. Indeed, source separation boils down to devising quantitative measures of diversity or contrast to extricate the sources. Typical measures of contrast are statistical criteria such as independence (i.e. ICA [23]), or sparsity and morphological diversity, see [7], [24], [25], [3], [4] and references therein.

B. Generalized Morphological Component Analysis

The GMCA framework assumes that the observed data $\mathbf{Y}$ is a linear instantaneous mixture of unknown sources $\mathbf{S}$ with an unknown mixing matrix $\mathbf{A}$ as in (3). For notational convenience, the dictionaries in the multichannel case will be transposed versions of those considered in the single channel case in Section II; each dictionary $\mathbf{\Phi}_k$ is now a matrix whose *rows* are unit-norm atoms. Thus, we let $\mathbf{\Phi}$ be the concatenation of K transforms: $\mathbf{\Phi} = \left[\mathbf{\Phi}_1^T, \cdots, \mathbf{\Phi}_K^T\right]^T$.

The GMCA framework assumes *a priori* that the sources $(s_i)_{i=1,\cdots,n}$ are sparse in the dictionary $\mathbf{\Phi}$; $\forall i \ s_i = \alpha_i \mathbf{\Phi}$ where α_i is sparse (or compressible). More precisely, in the GMCA setting, each source is modeled as the linear combination of K morphological components where each component is sparse in a specific basis

$$\forall i \in \{1, \cdots, n\}; \qquad s_i = \sum_{k=1}^{K} x_{ik} = \sum_{k=1}^{K} \alpha_{ik} \mathbf{\Phi}_k \ . \tag{4}$$

GMCA seeks an unmixing scheme, through the estimation of $\mathbf{A}$, which leads to the sparsest sources $\mathbf{S}$ in the dictionary $\mathbf{\Phi}$. This is expressed by the following optimization problem

$$\min_{\mathbf{A}, \alpha} \sum_{i=1}^{n} \sum_{k=1}^{K} \|\alpha_{ik}\|_p^p \ \text{ such that } \ \|\mathbf{Y} - \mathbf{A}\alpha\mathbf{\Phi}\|_F \leq \sigma \text{ and } \|a^i\|_2 = 1 \ \forall i = 1, \cdots, n \ , \tag{5}$$

where typically $p = 0$ or a relaxed version with $p = 1$. But other sparsity regularization terms can be used in (5), e.g. mixed-norms [26]. The unit ℓ_2-norm constraint on the columns of $\mathbf{A}$ avoids the classical scale indeterminacy of the product $\mathbf{AS}$ in (3). The reader may have noticed that the MCA problem (1) is a special case of the GMCA problem (5) when there is only one source $n = 1$ and one channel $m = 1$ (no mixing). Thus GMCA is indeed a multichannel generalization of MCA.

The program (5) is a difficult non-convex optimization problem even for convex penalties $p \geq 1$. More conveniently, the product $\mathbf{AS}$ can be split into $n \cdot K$ multichannel morphological components: $\mathbf{AS} = \sum_{i,k} a^i x_{ik} = \mathbf{A}\alpha\mathbf{\Phi} = \sum_{i,k} a^i \alpha_{ik} \mathbf{\Phi}_k$. Based on this decomposition, GMCA yields an alternating minimization algorithm to estimate iteratively one term at a time. It has been shown in [3] that estimating the morphological component $x_{ik} = \alpha_{ik} \mathbf{\Phi}_k$ assuming $\mathbf{A}$ and $x_{\{pq\} \neq \{ik\}}$ can be obtained through iterative thresholding for $p = 0$ and $p = 1$.

Now, considering fixed $\{a^p\}_{p \neq i}$ and $\mathbf{S}$, updating the column a^i is then just a least-squares estimate

$$a^i = \frac{1}{\|s_i\|_2^2} \left(\mathbf{Y} - \sum_{p \neq i} a^p s_p\right) s_i^T \ . \tag{6}$$

This estimate is then projected onto the unit sphere to meet the unit ℓ_2-norm constraint in (5).

C. GMCA algorithm

The GMCA algorithm is summarized in Algorithm 2. In the same vein as MCA, GMCA also relies on a salient-to-fine strategy. More precisely, GMCA is an iterative thresholding algorithm such that at each iteration it first computes coarse versions of the morphological components for a fixed source s_i. These raw sources

are estimated from their most significant coefficients in $\boldsymbol{\Phi}$. Then, the corresponding column a^i is estimated from the most significant features of s_i. Each source and its corresponding column of $\mathbf{A}$ are then alternately and progressively refined as the threshold λ decreases towards $\lambda_{\min}$. This particular iterative thresholding scheme provides robustness to noise, model imperfections and initialization by working first on the most significant features in the data and then progressively incorporating smaller details to finely tune the model parameters. As a multichannel extension of MCA, GMCA is also robust to noise and can be used with either linear or exponential decrease of the threshold [3], [4]. Moreover, hard thresholding leads to its best practical performance.

Algorithm 2: GMCA BSS algorithm.

Task: Blind source separation.
Parameters: Observed channels $\mathbf{Y}$, the dictionary $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1^T \cdots \boldsymbol{\Phi}_K^T]^T$, number of sources n, number of iterations N_{iter}, initial and stopping thresholds $\lambda^{(0)}$ and $\lambda_{\min}$, threshold update schedule.
Main iteration:
for $t = 1$ **to** N_{iter} **do**

    for $i = 1$ **to** n **do**

        for $k = 1$ **to** K **do**

            Compute marginal residuals $\mathbf{R}_{ik}^{(t)} = \mathbf{Y} - \sum_{\{p,q\} \neq \{i,k\}} a^{p^{(t-1)}} x_{pq}^{(t-1)}$.

            Update $(i,k)-$th coefficients by thresholding $\alpha_{ik}^{(t)} = \mathrm{TH}_{\lambda^{(t-1)}} \left(a^{i^{(t-1)^T}} \mathbf{R}_{ik}^{(t)} \boldsymbol{\Phi}_k^T \right)$.

            Update $(i,k)-$th component by reconstructing from the selected coefficients $x_{ik}^{(t)} = \alpha_{ik}^{(t)} \boldsymbol{\Phi}_k$.

        Update i-th source $s_i^{(t)} = \sum_{k=1}^{K} x_{ik}^{(t)}$.

        Update i-th mixing matrix column $a^{i^{(t)}} = \frac{1}{\|s_i^{(t)}\|_2^2} \left(\mathbf{Y} - \sum_{p \neq i}^{n} a^{p^{(t-1)}} s_p^{(t)} \right) s_i^{(t)^T}$ and normalize to a unit norm.

    Update the threshold $\lambda^{(t)}$ according to the given schedule.

    if $\lambda^{(t)} \leq \lambda_{\min}$ **then** stop.

Output: Sources $\left(s_i^{(N_{\text{iter}})} \right)_{i=1,\cdots,n}$ and mixing matrix $\mathbf{A}^{(N_{\text{iter}})}$.

D. Unknown number of sources

In BSS, the number of sources n is assumed to be a fixed known parameter of the problem. This is rather an exception than a rule, and estimating n from the data is a crucial and strenuous problem. Only a few works have attacked this issue. One can think of using model selection criteria such as the minimum description length used in [27]. In [4], a sparsity-based method to estimate n within the GMCA framework was proposed. Roughly speaking, this selection procedure uses GMCA to solve a sequence of problems (5) for each constraint radius $\sigma(q)$ with increasing q, $1 \leq q \leq m$. In [4], it was argued to set $\sigma(q)$ to the Frobenius-norm of the error when approximating the data matrix $\mathbf{Y}$ with its largest q singular vectors; see [4] for further details.

E. Hyperspectral data

In standard BSS, $\mathbf{A}$ is often seen as a mixing matrix of small size $m \times n$. On the other hand, there are applications in which one deals with data from instruments with a very large number of channels m which are well organized according to some physically meaningful index. A typical example is *hyperspectral* data where images are collected in a large number (i.e. hundreds) of contiguous regions of the electromagnetic spectrum. Regardless of other definitions appearing in other scientific communities, the term hyperspectral is used here for multichannel data model (3) where the number of channels m is large, and these channels achieve a uniform sampling of some meaningful physical index (e.g. wavelength, space, time) which we refer to as the *spectral* dimension. For such data, it then makes sense to consider the regularity of the spectral signatures $(a^i)_{i=1,\cdots,n}$. For instance these spectral signatures may be known *a priori* to have a sparse representation in some specified possibly redundant dictionary $\mathbf{\Psi}$ of *spectral* waveforms.

In [4], [28], the authors propose a modified GMCA algorithm capable of handling hyperspectral data. This is achieved by assuming that each rank-one matrix $\mathbf{X}_i = a^i s_i$ has a sparse representation in the multichannel dictionary $\mathbf{\Phi} \otimes \mathbf{\Psi}$ [9], [29] ($\otimes$ is the Kronecker product); i.e. $\forall i \; s_i = \alpha_i \mathbf{\Phi}$ and $a^i = \mathbf{\Psi}\beta^i$ where α_i and β_i are both sparse. The separation optimization problem for hyperspectral data is then

$$\min_{\alpha,\beta} \frac{1}{2} \left\| \mathbf{Y} - \sum_{i=1}^{n} \mathbf{\Psi}\beta^i \alpha_i \mathbf{\Phi} \right\|_F^2 + \sum_{i=1}^{n} \lambda_i \left\| \beta^i \alpha_i \right\|_1 , \tag{7}$$

where a Laplacian prior is imposed on each α_i conditionally on β_i and vice versa. Remarkably, this joint prior preserves the scale invariance of (3). Problem (7) is again a non-convex optimization problem for which no closed-form solution exists. In the line of the GMCA algorithm, thanks to the form of the ℓ_1-penalty in (7), a block-relaxation iterative thresholding algorithm was proposed in [4], [28] that alternately minimizes (7) with respect to α and β. It was shown by these authors that the update equations on the coefficient matrices are

$$\begin{aligned}
\alpha^{(t+1)} &= \mathrm{ST}_{\eta^{(t)}} \left(\left(\beta^{(t)T}\beta^{(t)} \right)^{-1} \beta^{(t)T} \mathbf{\Psi}^T \mathbf{Y} \mathbf{\Phi}^T \right), \\
\beta^{(t+1)} &= \mathrm{ST}_{\zeta^{(t)}} \left(\mathbf{\Psi}^T \mathbf{Y} \mathbf{\Phi}^T \alpha^{(t)T} \left(\alpha^{(t)}\alpha^{(t)T} \right)^{-1} \right).
\end{aligned} \tag{8}$$

$\eta^{(t)}$ is a vector of length n and entries $\eta^{(t)}[i] = \lambda^{(t)} \|\beta^{i(t)}\|_1 / \|\beta^{i(t)}\|_2^2$, $\zeta^{(t)}$ has length m and entries $\zeta^{(t)}[j] = \lambda^{(t)} \|\alpha_j^{(t)}\|_1 / \|\alpha_j^{(t)}\|_2^2$, and $\lambda^{(t)}$ is a decreasing threshold. The multichannel soft-thresholding operator ST_η acts on each row i with threshold $\eta[i]$ and ST_ζ acts on each column j with threshold $\zeta[j]$.

F. Applications

BSS: We first report a simple BSS application. Fig. 3 shows two original sources (left) and the two noisy mixtures SNR=10dB (middle). GMCA was applied with a dictionary containing curvelets and local DCT. As a quantitative performance measure of BSS methods, we use the mixing matrix criterion $\Delta_{\mathbf{A}} = \|\mathbf{I}_n - \mathbf{P}\hat{\mathbf{A}}^+ \mathbf{A}\|_1$, where $\hat{\mathbf{A}}^+$ is the pseudo-inverse of the estimate of the mixing matrix $\mathbf{A}$, and $\mathbf{P}$ is a matrix that reduces the
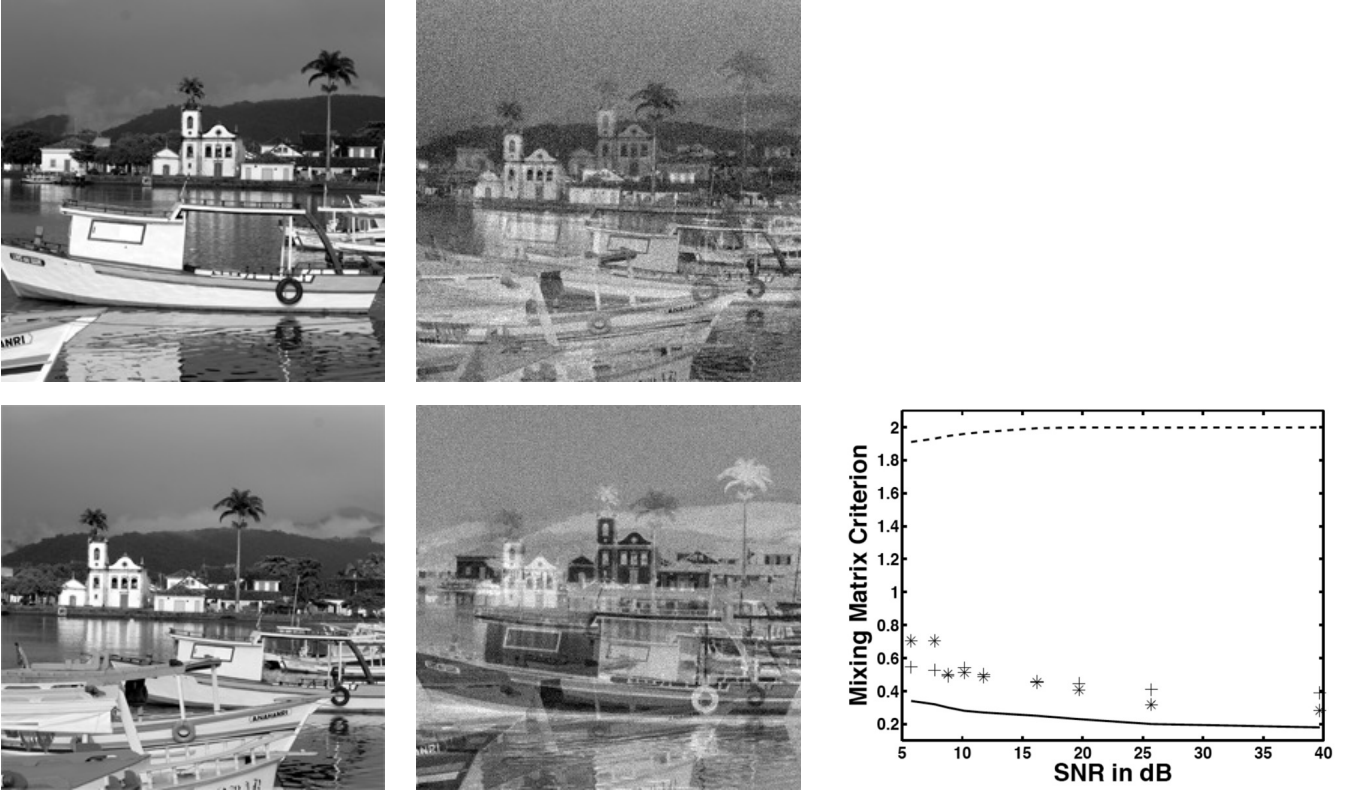
Fig. 3. Example of BSS with two sources (left) and two noisy mixtures (middle). The right panel depicts the evolution of the mixing matrix criterion $\Delta_{\mathbf{A}}$ with input SNR (*solid* line : GMCA, *dashed* line : JADE, *(+) :* RNA, *(⋆) :* EFICA).

scale/permutation indeterminacy of the mixing model. The right panel of Fig. 3 compares GMCA to popular BSS techniques in terms of $\Delta_{\mathbf{A}}$ as the SNR increases. We compare GMCA to ICA JADE [30], Relative Newton Algorithm (RNA) [31] that accounts for sparsity and EFICA [32]. Both RNA and EFICA were applied after "sparsifying" the data via an orthonormal wavelet transform. It can be seen that JADE performs rather badly, while RNA and EFICA behave quite similarly. GMCA seems to provide much better results especially at high noise level.

Color image denoising: GMCA can be applied to color image denoising. This is illustrated in Fig. 4 where the original RGB image is shown on the left. The middle image shows the RGB image obtained using a classical undecimated wavelet-domain hard thresholding on each color plane independently. GMCA is applied to the RGB color channels using the curvelet dictionary. In the notation of section III-A, we have $m = 3$ channels (color planes), each color channel is $y_j, j = 1, 2, 3$, $n = 3$ sources and $\mathbf{A}$ plays the role of the color space conversion matrix. Unlike classical color spaces (e.g. YUV, YCC), where the conversion matrix from RGB is fixed, the color space conversion matrix is here estimated by GMCA from the data. As such, GMCA is able to find adaptively the appropriate color space corresponding to the color image at hand. Once $\mathbf{A}$ is estimated by GMCA, we applied the same undecimated wavelet-based denoising to the estimated sources. The
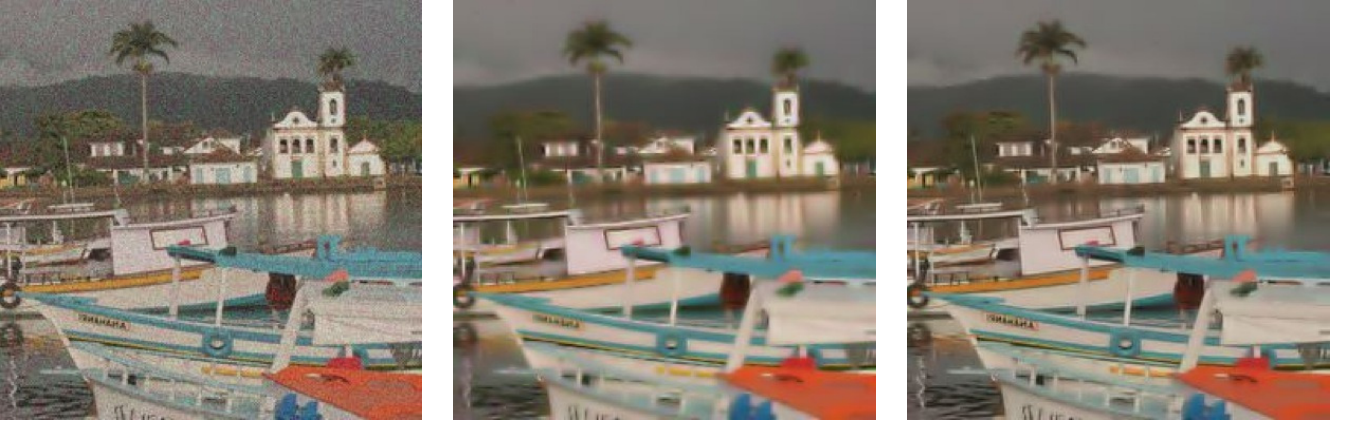
Fig. 4. Left: Original RGB image image with additive Gaussian noise SNR=15 dB. Middle: Wavelet-based denoising in the RGB space. Right: Wavelet-based denoising in the "GMCA color space".

denoised data are obtained by coming back to the RGB space via the estimated mixing matrix. Fig.(4) on the right shows the GMCA-based denoised image. Clearly, denoising in the "GMCA color space" is substantially better than in the RGB space (or other color spaces such as YUV or YCC; see [3]).

Hyperspectral data processing: In this experiment, we consider $m = 128$ mixtures of $n = 5$ source images. The sources are drawn at random from a set of structured images shown in Fig. 5(a). For the spectra (i.e. columns of $\mathbf{A}$), we randomly generated sparse coefficient vectors β^i ($i = 1, \ldots, n$) with independent Gaussian-distributed non-zero entries, and then applied the inverse orthogonal wavelet transform to these sparse vectors to get the spectra. $\mathbf{\Phi}$ was chosen as the curvelet dictionary. Fig. 5(b) gives four typical noisy observed channels with SNR=20dB. The sources recovered using GMCA (Algorithm 2) and its hyperspectral extension (iteration of section III-E) are shown respectively in Fig. 5(c) and (d). Visual inspection shows that GMCA is outperformed by hyperspectral GMCA which better accounts for both spatial and spectral sparsity.

IV. REPRODUCIBLE RESEARCH SOFTWARE

Following the philosophy of reproducible research, two toolboxes MCALab and GMCALab [22] are made available freely for download at the address

http://www.morphologicaldiversity.org

MCALab and GMCALab have been developed to demonstrate key concepts of MCA and GMCA and make them available to interested researchers and technologists. These toolboxes are libraries of Matlab routines that implement the decomposition and source separation algorithms overviewed in this paper. They contain a variety of scripts to reproduce the figures in our own articles, as well as other exploratory examples not included in the papers.
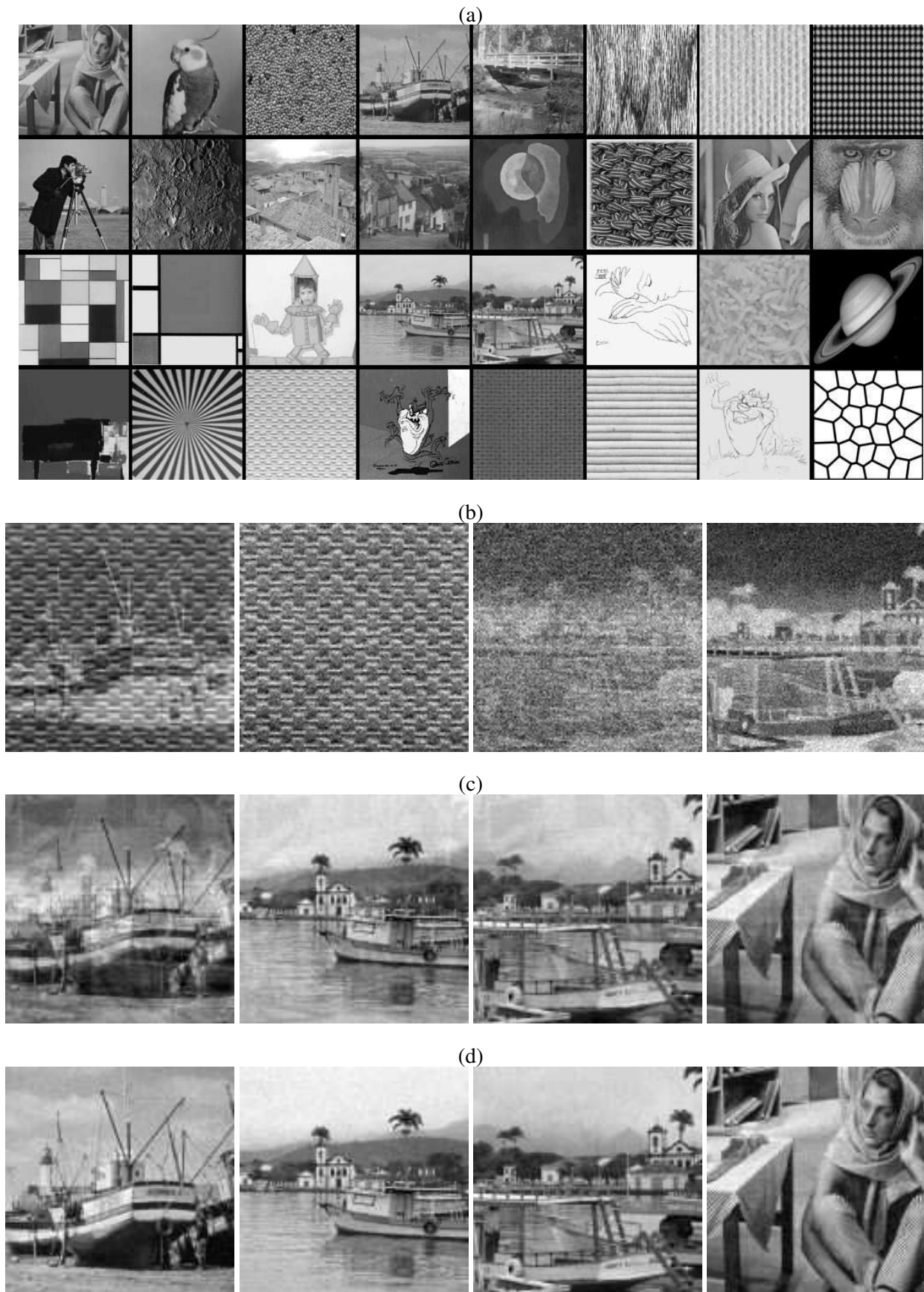
(a)

(b)

(c)

(d)

Fig. 5. (a) Images used as sources. (b) Four noisy mixtures out of $m = 128$ (SNR=20dB). (c) Recovered sources using GMCA. (d) Recovered sources using hyperspectral GMCA.

V. Conclusion

In this paper, we gave an overview of how sparsity and morphological diversity can be used advantageously to regularize image decomposition and blind source separation problems. We also reported several numerical experiments to illustrate the wide applicability of the algorithms described. We believe that this is an exciting field where many interesting problems are still open. Among them, we may cite for instance the theoretical guarantees of the sparsity-regularized BSS problem, and sharper theoretical guarantees for the decomposition problem by exploiting geometry.

References

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

[2] ——, "Redundant multiscale transforms and their application for morphological component analysis," in *Advances in Imaging and Electron Physics*, P. Hawkes, Ed. Academic Press, Elsevier, 2004, vol. 132.

[3] J. Bobin, J.-L. Starck, M. J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Trans. Image Processing*, vol. 16, no. 11, pp. 2662–2674, November 2007.

[4] J. Bobin, J.-L. Starck, Y. Moudden, and M. J. Fadili, "Blind source separation : The sparsity revolution," in *Advances in Imaging and Electron Physics*, P. Hawkes, Ed. Academic Press, Elsevier, 2008, pp. 221–298.

[5] Y. Meyer, "Oscillating patterns in image processing and in some nonlinear evolution equations." *The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures*, 2001.

[6] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. of Scientific Computing*, vol. 19, no. 1-3, pp. 553–572, 2003.

[7] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition," *Neural Computations*, vol. 13/4, 2001.

[8] P.Bofill and M.Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[9] J. Bobin, Y. Moudden, M. J. Fadili, and J.-L. Starck, "Morphological diversity and sparsity for multichannel data restoration," *Journal of Mathematical Imaging and Vision*, vol. 10, no. 2, pp. 149–168, 2009.

[10] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[11] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation noise removal algorithm," *Physica D*, vol. 60, pp. 259–268, 1992.

[12] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[13] S. Sardy, A. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 361–379, 2000.

[14] J. Bobin, J.-L. Starck, M. J. Fadili, Y. Moudden, and D. L. Donoho, "Morphological component analysis: An adaptive thresholding strategy," *IEEE Trans. Image Processing*, vol. 16, no. 11, pp. 2675 – 2681, November 2007.

[15] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, 1998.

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

[17] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Model. Simul.*, vol. 5, pp. 861–899, 2005.

[18] E. Candès and D. Donoho, "Ridgelets: the key to high dimensional intermittency?" *Philosophical Transactions of the Royal Society of London A*, vol. 357, pp. 2495–2509, 1999.

[19] L. Demanet and L. Ying, "Wave atoms and sparsity of oscillatory patterns," *Appl. Comput. Harmon. Anal.*, vol. 23, no. 3, pp. 368–387, 2007.

[20] F. G. Meyer and R. R. Coifman, "Brushlets: A tool for directional image analysis and image compression," *Appl. and Comput. Harmon. Anal.*, vol. 5, pp. 147–187, 1997.

[21] G. Peyré, M. J. Fadili, and J.-L. Starck, "Learning adapted dictionaries for geometry and texture separation," in *Proceedings of the SPIE conference Wavelets XII*, San Diego, August 2007.

[22] M. J. Fadili, J.-L. Starck, M. Elad, and D. L. Donoho, "MCALab: Reproducible research in signal and image decomposition and inpainting," *IEEE Computing in Science and Engineering*, 2009, to appear.

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

[24] 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.

[25] J. Bobin, Y. Moudden, J.-L. Starck, and M. Elad, "Morphological diversity and source separation," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 409–412, 2006.

[26] M. Kowalski, E. Vincent, and R. Gribonval, "Under-determined source separation via mixed-norm regularized minimization," in *EUSIPCO'08*, Lausanne, August 2008.

[27] R. Balan, "Estimator for number of sources using minimum description length criterion for blind sparse source mixtures," in *Independent Component Analysis and Signal Separation*, ser. LNCS, M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, Eds., vol. 4666. Springer, 2007, pp. 333–340.

[28] Y. Moudden, J. Bobin, J.-L. Starck, and M. J. Fadili, "Dictionary learning with spatio-spectral sparsity constraints," in *Proceedings of SPARS'09*, St Malo, France, April 2009.

[29] R.Gribonval and M.Nielsen, "Beyond sparsity : recovering structured representations by l_1-minimization and greedy algorithms: Application to the analysis of sparse underdetermined ica," *Advances in Computational Mathematics*, vol. 28, no. 1, pp. 23–41, 2008.

[30] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE. Special issue on blind identification and estimation*, vol. 9, no. 10, pp. 2009–2025, Oct. 1998.

[31] M. Zibulevski, "Blind source separation with relative newton method," *Proccedings ICA2003*, pp. 897–902, 2003.

[32] Z. Koldovsky, P. Tichavsky, and E. Oja, "Efficient variant of algorithm fastica for independent component analysis attaining the cramer-rao lower bound," *IEEE Transactions on neural networks*, vol. 17, pp. 1265–1277, 2006.