A Markov Model for Blind Image Separation by a Mean-Field EM Algorithm

Anna Tonazzini, Luigi Bedini, and Emanuele Salerno

***Abstract*—This paper deals with blind separation of images from noisy linear mixtures with unknown coefficients, formulated as a Bayesian estimation problem. This is a flexible framework, where any kind of prior knowledge about the source images and the mixing matrix can be accounted for. In particular, we describe local correlation within the individual images through the use of Markov random field (MRF) image models. These are naturally suited to express the joint pdf of the sources in a factorized form, so that the statistical independence requirements of most independent component analysis approaches to blind source separation are retained. Our model also includes edge variables to preserve intensity discontinuities. MRF models have been proved to be very efficient in many visual reconstruction problems, such as blind image restoration, and allow separation and edge detection to be performed simultaneously. We propose an expectation–maximization algorithm with the mean field approximation to derive a procedure for estimating the mixing matrix, the sources, and their edge maps. We tested this procedure on both synthetic and real images, in the fully blind case (i.e., no prior information on mixing is exploited) and found that a source model accounting for local autocorrelation is able to increase robustness against noise, even space variant. Furthermore, when the model closely fits the source characteristics, independence is no longer a strict requirement, and cross-correlated sources can be separated, as well.**

***Index Terms*—Blind source separation (BSS), edge and feature detection, Markov random fields (MRFs), parameter learning, scene analysis.**

I. Introduction

BLIND SOURCE SEPARATION (BSS), which became an active research topic in signal processing in the last decade, consists of separating a set of unknown signals from a set of linear mixtures of them, when no knowledge is available about the mixing coefficients. The most well-known application example of BSS is the so-called "cocktail party" problem in audio processing. Other applications include the removal of underlying artifact components of brain activity from EEG records, and the search for hidden factors in parallel financial data series. It has only been more recently that BSS has received attention in image processing and computer vision, as well [1]. To cite just a few of the many emerging applications in these fields, we recall feature extraction or noise removal from natural images, separation of components in astrophysical maps [14], and document analysis and restoration [30].

In order to solve BSS, which is a severely ill-posed inverse problem [2], many techniques have been proposed so far. Among them, the independent component analysis (ICA) methods are based on the assumption of mutual independence of the sources. Most of these methods were developed in the case of noiseless data, and differ from one another in the way they enforce independence. The maximum likelihood (ML) method [6] directly assumes a factorized form for the joint source distribution; in the infomax method [7], entropy is used as a measure of independence; other methods ensure independence by minimizing contrast functions related to statistics of order greater than two [8]. The strict relationships among the various methods have been investigated as well [9], and some fast and efficient algorithms have been proposed, such as FastICA [10]. The independence requirement can be fulfilled in some applications, but in many cases, there is a clear evidence of correlation among the sources. Although some of the proposed algorithms have been experimentally shown to perform well even in the lack of independence, all of them perform poorly when noise affects the data. Recently, some work has been done to overcome this limitation. In particular, the noisy FastICA algorithm [11], and an independent factor analysis (IFA) method [12]–[14] have been developed, the latter being also capable of estimating the noise covariance matrix. Nevertheless, while providing satisfactory estimates of the mixing matrices, these methods still produce noisy source estimates. We believe that a way to obtain robust estimates for both the mixing matrix and the sources is to incorporate into the problem the available information about autocorrelation properties of the single sources. Indeed, correlation is an important feature of most real-world signals, especially of images and can be exploited to regularize many inverse problems.

In [15]–[17], Bayesian estimation has been proposed as a suitable, unifying framework for BSS, within which the other methods can be viewed as special cases. The Bayesian approach is also the most natural and flexible way to account for prior knowledge we may possess about a problem. In this paper, we apply Bayesian estimation to regularize the blind separation of noisy mixtures of images. We retain the independence constraint of the ICA approach, and reformulate the BSS problem as the joint maximum *a posteriori* (MAP) estimation of the mixing matrix and the sources. The flexibility of the Bayesian formulation is exploited to enforce constraints on the solution. In particular, Markov random field (MRF) models, under the form of suitable stabilizing Gibbs priors, are used for both enforcing independence and describing the local smoothness of the source

Manuscript received April 22, 2004; revised January 14, 2005. This work was supported in part by European Project Network of Excellence MUSCLE FP6-507752 (Multimedia Understanding through Semantics, Computation, and Learning). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Vicent Caselles.

The authors are with the Istituto di Scienza e Tecnologie dell'Informazione, I-56124 Pisa, Italy (e-mail: anna.tonazzini@isti.cnr.it; luigi.bedini@isti.cnr.it; emanuele.salerno@isti.cnr.it).

Digital Object Identifier 10.1109/TIP.2005.860323

images. Indeed, the autocorrelation property of images is usually only local, and is broken in correspondence of steep fronts, such as object boundaries. These are associated with the discontinuities of the intensity process that should be preserved and recovered for a reliable reconstruction of the original images. The interaction between the intensity field and the unknown discontinuity set is usually described by suitable cost functionals, in the form of a weak membrane energy. Blake and Zisserman [32] assumed a discrete image model and proposed a deterministic algorithm, namely the graduated nonconvexity (GNC) algorithm, for the optimization of the weak membrane energy. Mumford and Shah [33], in a continuous setting, proposed variational techniques to optimize a cost functional where a singular part represents the discontinuity set. Geman and Geman [3] derived an MRF image model that incorporates a binary line process associated to the discontinuities and proposed stochastic relaxation for the optimization of the overall cost functional. A review of the approaches above can be found in [34].

Following the idea in [3], we adopt an MRF source model with a binary line process for breaking the smoothness constraint where the original sources are likely to present steep fronts. For the maximization of the posterior probability with respect to all the unknowns, we propose to use an expectation–maximization (EM) algorithm. We derive the method considering, at the moment, a known number of independent sources, which is equal to the number of mixtures, and then propose specific approximations to develop a computationally feasible algorithm. Following the EM strategy, the source variables plus edges are viewed as hidden data, and averaged out in the so-called E-step of the procedure, and the mixing matrix elements are viewed as the parameters to be estimated. To reduce the computational complexity of the original formulation, i.e., to compute the expectations with respect to the hidden data in the E-step, we propose using the mean field approximation. At convergence, alternating the E-step and the maximization of the resulting objective function (the M-step) yields an estimate of the mixing parameters, the source images and the related edge maps. We experimentally verified that using source models accounting for local autocorrelation enables us to increase robustness against noise, both space invariant and space variant, and also against cross correlation of the sources. We worked on images of documents containing overlapped texts, with the aim at separating them. A possible application of our method could be bleed-through or show-through removal from digital scans of degraded text documents. The estimated edge maps can also be useful in applications such as image segmentation and classification. In the particular application of digital document analysis, this could also be a preliminary processing to optical character recognition.

II. BSS Through Bayesian Estimation

According to the BSS formalism, the data model we consider is

$$\mathbf{x}(t) = A\mathbf{s}(t) + \mathbf{n}(t) \qquad t = 1, 2, \ldots, T \tag{1}$$

where $\mathbf{x}(t)$ is the vector of the measurements, $\mathbf{s}(t)$ is the column vector of the unknown sources, and $\mathbf{n}(t)$ is the noise or measurement error vector, at location t, and A is the unknown mixing matrix, assumed location independent. We assume the same number N of measured and source signals, so that A is an $N \times N$ matrix. For later use, we also define the vectors

$$\mathbf{s}_i = (s_i(1), s_i(2), \ldots, s_i(T)) \qquad i = 1, 2, \ldots, N.$$

Considering a white and Gaussian noise with zero mean, the likelihood is given by

$$P(\mathbf{x}|\mathbf{s}, A) = \frac{1}{Z_R} \times \exp\left\{-\frac{1}{2}\sum_t (\mathbf{x}(t) - A\mathbf{s}(t))' R^{-1}(t)(\mathbf{x}(t) - A\mathbf{s}(t))\right\} \tag{2}$$

where $\mathbf{x} = (\mathbf{x}(1), \ldots, \mathbf{x}(T))$ and $\mathbf{s} = (\mathbf{s}(1), \ldots, \mathbf{s}(T))$ are the matrices of data and source images, respectively, R is the noise covariance matrix, assumed, in general, to be location dependent, and Z_R is the normalizing constant. The prime means transposition. Note that the ML principle applied to (2) would clearly give an underdetermined problem, unless more information is exploited. In a fully Bayesian approach, both A and $\mathbf{s}$ are assumed as independent unknowns, and are assigned with prior distributions $P(A)$ and $P(\mathbf{s})$, respectively. Then, A and $\mathbf{s}$ can be simultaneously estimated by maximizing some joint distribution, e.g., $P(\mathbf{s}, A, \mathbf{x})$. Our problem, thus, becomes

$$(\hat{\mathbf{s}}, \hat{A}) = \arg\max_{\mathbf{s},A} P(\mathbf{s}, A, \mathbf{x}). \tag{3}$$

Note that $P(\mathbf{s}, A, \mathbf{x})$ can be decomposed as

$$P(\mathbf{s}, A, \mathbf{x}) = P(\mathbf{x}|\mathbf{s}, A)P(\mathbf{s})P(A) = P(\mathbf{s}, A|\mathbf{x})P(\mathbf{x}). \tag{4}$$

Since $P(\mathbf{x})$ depends on neither A nor on $\mathbf{s}$, $P(\mathbf{s}, A, \mathbf{x})$ is, thus, proportional to the posterior distribution $P(\mathbf{s}, A|\mathbf{x})$. Hence, problem of (3) is equivalent to the following MAP estimation problem

$$\begin{aligned}(\hat{\mathbf{s}}, \hat{A}) &= \arg\max_{\mathbf{s},A} P(\mathbf{s}, A, |\mathbf{x}) \\ &= \arg\max_{\mathbf{s},A} P(\mathbf{x}|\mathbf{s}, A)P(\mathbf{s})P(A).\end{aligned} \tag{5}$$

The posterior distribution, or equivalently distribution $P(\mathbf{s}, A, \mathbf{x})$, accounts for all information we have about the problem, and can restrict the set of solutions associated to the likelihood part of (5) by means of the priors $P(\mathbf{s})$ and $P(A)$. For instance, the prior $P(\mathbf{s})$ can be used to introduce any known dependence between different sources or time/space correlation within individual sources. In the ICA approach, A is assumed to have a uniform prior and the sources are only assumed to be mutually independent, that is

$$P(\mathbf{s}(t)) = \prod_i P_i(s_i(t)) \qquad \forall t \tag{6}$$

whatever the form of the marginals P_i.

Note that, assuming a given order for the P_i, for any permutation matrix C, we have, in general, $P(\mathbf{s}) \neq P(C\mathbf{s})$. Conversely, if $P(\mathbf{s})$ is invariant to permutation of the s_i, then for any maximizer $(\mathbf{s}^*, A^*)$ to (5), the couple $(C\mathbf{s}^*, A^*C^T)$ is still a maximizer to (5). Any solution to the problem can, thus, be found up to an arbitrary permutation of the components of $\mathbf{s}$ and the corresponding columns of A. In particular, when the ICA position (6) holds true, the property of invariance to permutation implies that all the P_is are equal. Note also that most ICA models do not assume any noise in the data, and,

thus, the problem reduces to obtain a vector $\mathbf{s}^*$ and a matrix A^*, such that the divergence between $P(\mathbf{s}^*)$ and the factorized density in (6) is minimized, subject to $A^*\mathbf{s}^* = \mathbf{x}$. This implies that another basic indetermination in the ICA approach is in scale, that is, for each $(\mathbf{s}^*, A^*)$, if D is a diagonal scaling matrix $P(D\mathbf{s}^*, A^*D^{-1}|\mathbf{x}) = P(\mathbf{s}^*, A^*|\mathbf{x})$. These indeterminations, however, are not so important in BSS, provided that $P(\mathbf{s})$ is informative enough to properly constrain the solution. A more substantial indetermination is obtained with ICA models when two or more marginals are Gaussian. For instance, if the marginals are all identical Gaussian densities, for any solution $(\mathbf{s}^*, A^*)$, we obtain another solution $(K\mathbf{s}^*, A^*K^{-1})$ through any rotation matrix K. Together with the permutation and scale ambiguities, this makes ICA unable to separate Gaussian sources. The presence of the prior $P(A)$ in (5) may restrict the set of solutions to our problem. In particular, any prior which is sensitive to column interchanges eliminates the permutation ambiguity. If $P(A)$ is able to constrain the values of the elements of A, the scale ambiguity is also eliminated.

In this paper, we assume independent sources, thus letting (6) hold true, but adopt MRF models to describe regularity properties, under the form of local spatial correlation (local smoothness), for the individual sources. As already mentioned, this will be shown to make the Bayesian approach advantageous over ICA in terms of robustness against noise, and also against cross correlation of the sources.

As usual in blind estimation, joint maximization of (5) is very hard. An equivalent way to write it is the following:

$$\hat{A} = \arg\max_{A} P\left(\mathbf{x}|\hat{\mathbf{s}}(A), A\right) P\left(\hat{\mathbf{s}}(A)\right) P(A) \tag{7}$$

where

$$\hat{\mathbf{s}}(A) = \arg\max_{\mathbf{s}} P(\mathbf{x}|\mathbf{s}, A) P(\mathbf{s}). \tag{8}$$

If $\hat{\mathbf{s}}(A)$ were available for each A, for instance, through an analytical formula, the problem would reduce to a maximization with respect to the only variable A. However, this is not usually the case, so that the problem, either in the form of (5) or in the form of (7) and (8), needs to be reduced somehow.

In the following, we propose to solve (5) via an EM algorithm, whose derivation in this specific case takes advantage of the use of the mean field approximation.

III. Choice of the Gibbs Source Priors

MRF models have become very popular since the middle 1980s [3], [4], especially in connection to inverse problems of image processing and computer vision, such as restoration, denoising, segmentation, optical flow estimation, and so on. Through MRF models, it is, indeed, possible to describe local image features, such as edges, in order to enforce space-variant smoothness constraints on the solutions. Furthermore, the local nature of these models allows us to devise distributed and parallel algorithms. In this paper, we propose to use MRF in a BSS and ICA context, for modeling both the independence among the sources and the local properties of spatial correlation within each source. Let us consider, then, the distribution of the

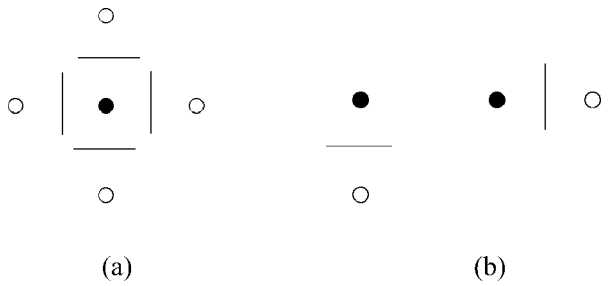


Fig. 1. Local interactions in our MRF source image model. (a) Neighborhood system. (b) Clique system.

ith source $\mathbf{s}_i$ in our problem. According to the MRF formalism, it must have the following Gibbs form:

$$P_i(\mathbf{s}_i) = \frac{1}{Z_i} \exp\left\{-U_i(\mathbf{s}_i)\right\} \tag{9}$$

where Z_i is the normalizing constant and $U_i(\mathbf{s}_i)$ is the prior energy in the form of a sum of potential functions over the set of interacting locations (cliques). These potential functions should describe some regularity in the ith source, by penalizing high gradient values. This regularity is physically plausible in many real-world applications, and, as already said, it is an essential constraint to prevent the reconstructions from being unstable when the data are noisy. Nevertheless, the source signals can present steep fronts, which must be preserved as well. We, thus, refer to edge-preserving priors (or stabilizers) [19], [20]. Among the many proposed, some priors have the property of being convex [21], [23], so that gradient descent can be used to optimize with respect to $\mathbf{s}$. More generally, specific priors were also proposed that allow us to "correct" their nonconvexity by providing a sequence of approximations, to be optimized in turn according to the GNC strategy [22]. In a previous paper [28], we experimented both convex and nonconvex edge-preserving stabilizers for the separation of noisy mixtures of images, by adopting an alternating maximization strategy for the joint MAP estimation of the mixing matrix and the sources.

In this paper, we adopt a different prior, which contains an extra binary process $\mathbf{l}$, marking "explicitly" the image edges, ideally located between each pair of adjacent pixel locations [3]. The ith source is represented by the couple $(\mathbf{s}_i, \mathbf{l}_i)$, where $\mathbf{s}_i$ is also called "intensity process," and $\mathbf{l}_i$ is called "line process" (see also [19]). In the two-dimensional grid of the image pixels, for each location t, we consider the first order neighborhood shown in Fig. 1(a), and the clique system shown in Fig. 1(b). We measure the local regularity of the image as the sum of the squared differences between the values of the two elements in all the image cliques, weighted by the proper edge elements, in order to break the smoothness constraint where it is not plausible. However, the line process is also unknown, and has to be treated as an extra variable of the problem. Let us call $N(t)$ the set of the two neighbor locations adjacent to t horizontally from right and vertically from bottom. We then define $U_i(\mathbf{s}_i, \mathbf{l}_i)$ as

$$U_i(\mathbf{s}_i, \mathbf{l}_i) = \sum_{t} \sum_{r \in N(t)} \lambda_i \left(s_i(t) - s_i(r)\right)^2 \left(1 - l_i(t,r)\right) + \alpha_i l_i(t,r) \tag{10}$$

where λ_i and α_i are two positive weights related, respectively, to the degree of smoothness and to the number of edges we foresee

for the source $\mathbf{s}_i$. As is easily seen, indeed, a line element set to 1 (an active edge) breaks the smoothness constraint between locations t and r, and its presence is paid by a constant α_i, which penalizes the presence of too many active edges. Conversely, a line element set to 0 lets the difference of intensity between locations t and r be strongly penalized. The most probable images present steep fronts where the gradient is higher than a threshold of value $\sqrt{\alpha/\lambda}$, and are smooth elsewhere.

The prior distribution P_i of (9) is now augmented with the line process, and becomes

$$P_i(\mathbf{s}_i, \mathbf{l}_i) = \frac{1}{Z_i} \exp\{-U_i(\mathbf{s}_i, \mathbf{l}_i)\} \tag{11}$$

and the MAP estimation problem of (5) becomes

$$(\hat{\mathbf{s}}, \hat{\mathbf{l}}, \hat{A}) = arg \max_{\mathbf{s},\mathbf{l},A} P(\mathbf{x}|\mathbf{s}, A) P(\mathbf{s}, \mathbf{l}) P(A). \tag{12}$$

Note that the likelihood $P(\mathbf{x}|\mathbf{s}, A)$ does not depend on $\mathbf{l}$ because the data are conditionally independent of the line process, i.e., the line status does not affect the data.

IV. EM Method and the Mean Field Approximation

According to the EM formalism [5], BSS can be described as an incomplete data problem, where part of the data [the sources $(\mathbf{s}, \mathbf{l})$] is hidden, and has to be estimated along with the parameters, i.e., the elements of matrix A. In this respect, the true problem to be solved, in the incomplete data set of the observations $\mathbf{x}$, becomes

$$\hat{A} = \arg\max_{A} P(\mathbf{x}|A) P(A) \tag{13}$$

while the original problem in (12) is seen as a transposition of the problem of (13) in the complete data set. The many-to-one relationship among complete and incomplete data in this case is the projection from $(\mathbf{s}, \mathbf{l}, \mathbf{x})$ to $\mathbf{x}$, and the two probability distributions are related as follows:

$$P(\mathbf{x}|A) P(A) = \int_{\mathbf{s}} \sum_{\mathbf{l}} P(\mathbf{x}|\mathbf{s}, A) P(\mathbf{s}, \mathbf{l}) P(A) d\mathbf{s}. \tag{14}$$

The EM algorithm is an iterative procedure for solving (13) by making use of the associated distribution $P(\mathbf{x}|\mathbf{s}, A) P(\mathbf{s}, \mathbf{l}) P(A)$. In this specific case, each iteration consists of the two following steps.

E-step: Compute the expectation of function $\log[P(\mathbf{x}|\mathbf{s}, A) P(\mathbf{s}, \mathbf{l}) P(A)]$, with respect to $(\mathbf{s}, \mathbf{l})$ and conditioned on the observed data and the current estimate $A^{(k)}$ of the parameters. The whole expectation becomes

$$\begin{aligned} &Q\left(A|A^{(k)}\right) \\ &= E_{\mathbf{s},\mathbf{l}}\left[\log P(\mathbf{x}|\mathbf{s}, A) + \log P(\mathbf{s}, \mathbf{l}) + \log P(A)|\mathbf{x}, A^{(k)}\right] \\ &= E_{\mathbf{s},\mathbf{l}}\left[\log P(\mathbf{x}|\mathbf{s}, A)|\mathbf{x}, A^{(k)}\right] \\ &\quad + E_{\mathbf{s},\mathbf{l}}\left[\log P(\mathbf{s}, \mathbf{l})|\mathbf{x}, A^{(k)}\right] + \log P(A) \\ &= E_{\mathbf{s},\mathbf{l}}\left[\log P(\mathbf{x}|\mathbf{s}, A)|\mathbf{x}, A^{(k)}\right] + \log P(A) + \text{const.} \end{aligned} \tag{15}$$

M-step: Compute

$$\begin{aligned} A^{(k+1)} &= \arg\max_{A} Q\left(A|A^{(k)}\right) \\ &= \arg\max_{A} E_{\mathbf{s},\mathbf{l}}\left[\log P(\mathbf{x}|\mathbf{s}, A)|\mathbf{x}, A^{(k)}\right] \\ &\quad + \log P(A). \end{aligned} \tag{16}$$

It is to be noted that the expectation must be computed with respect to the posterior distribution $P(\mathbf{s}, \mathbf{l}|\mathbf{x}, A^{(k)})$ of $(\mathbf{s}, \mathbf{l})$ given the data and the previous estimate of A, that is

$$\begin{aligned} &E_{\mathbf{s},\mathbf{l}}\left[\log P(\mathbf{x}|\mathbf{s}, A)|\mathbf{x}, A^{(k)}\right] \\ &\qquad = \int_{\mathbf{s}} \sum_{\mathbf{l}} \log P(\mathbf{x}|\mathbf{s}, A) P\left(\mathbf{s}, \mathbf{l}|\mathbf{x}, A^{(k)}\right) d\mathbf{s}. \end{aligned} \tag{17}$$

Although the EM formulation allows us to reduce the complexity of the original joint maximization problem, the integral in (17) is still very complicated, and a closed form is not generally available. In the following, we show how the mean field theory [25] can be exploited to efficiently compute the expectation [26], [27].

In (17), the form of $P(\mathbf{x}|\mathbf{s}, A)$ is given by (2), and $P(\mathbf{s}, \mathbf{l}|\mathbf{x}, A^{(k)})$ is a complicated expression containing interactions between different pixels through the prior $P(\mathbf{s}, \mathbf{l})$ given by (6), (10), and (11). The mean field theory suggests that, in the computation of the statistics of a single element in a random field, the effect of all the other elements can be approximated by the effect of their means. Applying this theory to each vector-valued field element $\mathbf{s}(t)$, $P(\mathbf{s}, \mathbf{l}|\mathbf{x}, A^{(k)})$ in (17) is replaced by the product of the mean field marginals $P_t(\mathbf{s}(t)|\langle\mathbf{s}(t)\rangle, \langle\mathbf{l}\rangle, \mathbf{x}, A^{(k)})$, which are obtained from $P(\mathbf{s}, \mathbf{l}|\mathbf{x}, A^{(k)})$ when each $\mathbf{s}(r)$, $r \neq t$, and each $\mathbf{l}(t, r)$ are replaced by their mean values $\langle\mathbf{s}(r)\rangle$ and $\langle\mathbf{l}(t, r)\rangle$, respectively

$$P\left(\mathbf{s}, \mathbf{l}|\mathbf{x}, A^{(k)}\right) \approx \prod_{t} P_t\left(\mathbf{s}(t)|\langle\mathbf{s}(t)\rangle, \langle\mathbf{l}\rangle, \mathbf{x}, A^{(k)}\right) \tag{18}$$

where $\langle\mathbf{s}(t)\rangle$ denotes the set of the mean vector values in all the locations different from t, and $\langle\mathbf{l}\rangle$ denotes the vector of all the mean line values. In this way, the summation in $\mathbf{l}$ disappears in (17), and we get

$$\begin{aligned} &E_{\mathbf{s},\mathbf{l}}\left[\log P(\mathbf{x}|\mathbf{s}, A)|\mathbf{x}, A^{(k)}\right] \approx \\ &\quad = \sum_{t} \int_{\mathbf{s}} -\frac{1}{2} \left[\mathbf{x}(t) - A\mathbf{s}(t)\right]' R^{-1}(t) \left[\mathbf{x}(t) - A\mathbf{s}(t)\right] \\ &\qquad \times P_t\left(\mathbf{s}(t)|\langle\mathbf{s}(t)\rangle, \langle\mathbf{l}\rangle, \mathbf{x}, A^{(k)}\right) d\mathbf{s}(t) \\ &\quad = -\frac{1}{2} \sum_{t} \int_{\mathbf{s}} Tr\left\{R^{-1}(t) \left(\mathbf{x}(t) - A\mathbf{s}(t)\right) \left(\mathbf{x}(t) - A\mathbf{s}(t)\right)'\right\} \\ &\qquad \times P_t\left(\mathbf{s}(t)|\langle\mathbf{s}(t)\rangle, \langle\mathbf{l}\rangle, \mathbf{x}, A^{(k)}\right) d\mathbf{s}(t) \end{aligned} \tag{19}$$

where use was made of (2) and (18). By neglecting the terms not depending on A, we obtain

$$\begin{aligned} &E_{\mathbf{s},\mathbf{l}}\left[\log P(\mathbf{x}|\mathbf{s}, A)|\mathbf{x}, A^{(k)}\right] \approx \\ &\quad = -\frac{1}{2} \sum_{t} Tr\left\{R^{-1}(t) \left[-A\langle\mathbf{s}(t)\rangle \mathbf{x}'(t) - \mathbf{x}(t)\langle\mathbf{s}'(t)\rangle A'\right.\right. \\ &\qquad \left.\left. + A\langle\mathbf{s}(t)\mathbf{s}'(t)\rangle A'\right]\right\}. \end{aligned} \tag{20}$$

The expectation is, thus, calculated by a sum of expectations on the individual pixels. The related densities have the following form:

$$P_t\left(\mathbf{s}(t)|\left\langle\mathbf{s}(t)\right\rangle,\langle\mathbf{l}\rangle,\mathbf{x},A^{(k)}\right) = \frac{1}{Z}\times\exp\left\{-\frac{1}{2}\left(\mathbf{x}(t)-A^{(k)}\mathbf{s}(t)\right)' R^{-1}(t) \times\left(\mathbf{x}(t)-A^{(k)}\mathbf{s}(t)\right)\right\} \times\exp\left\{-\sum_{r\in N(t)}\left(\mathbf{s}(t)-\langle\mathbf{s}(r)\rangle\right)'\Lambda_t(r)\left(\mathbf{s}(t)-\langle\mathbf{s}(r)\rangle\right) -\sum_i\sum_{r\in N(t)}\alpha_i\left\langle l_i(t,r)\right\rangle\right\} \tag{21}$$

where $\Lambda_t(r)$ is a diagonal matrix whose ith element is given by $\lambda_i(1-\langle l_i(t,r)\rangle)$ and $N(t)$ has been redefined as the set of all the four pixels in the neighborhood of Fig. 1(a). Apart from terms independent of $\mathbf{s}(t)$, it is easy to verify that the density in (21) can be expressed as a Gaussian

$$P_t\left(\mathbf{s}(t)|\left\langle\mathbf{s}(t)\right\rangle,\langle\mathbf{l}\rangle,\mathbf{x},A^{(k)}\right) \propto exp\left\{-\frac{1}{2}\left(\mathbf{s}(t)-\mu(t)\right)' B^{-1}(t)\left(\mathbf{s}(t)-\mu(t)\right)\right\} \tag{22}$$

with

$$B(t)=\left[A^{(k)\prime}R^{-1}(t)A^{(k)}+2\sum_{r\in N(t)}\Lambda_t(r)\right]^{-1} \tag{23}$$

and

$$\mu(t)'=\left[\mathbf{x}(t)'R^{-1}(t)A^{(k)}+2\sum_{r\in N(t)}\left\langle\mathbf{s}(r)'\right\rangle\Lambda_t(r)\right]B(t). \tag{24}$$

The expectations in (20) are, thus

$$\left\langle\mathbf{s}(t)\right\rangle=\mu(t)\qquad\forall t \tag{25}$$

$$\left\langle\mathbf{s}(t)\mathbf{s}'(t)\right\rangle=B(t)+\left\langle\mathbf{s}(t)\right\rangle\left\langle\mathbf{s}'(t)\right\rangle\qquad\forall t. \tag{26}$$

The only quantity we still need to calculate is matrix $\Lambda_t(r)$. This is defined by the mean values of the line elements, that is

$$\left\langle l_i(t,r)\right\rangle=\int_{\mathbf{s}}\sum_{\mathbf{l}}l_i(t,r)P\left(\mathbf{s},\mathbf{l}|\mathbf{x},A^{(k)}\right)d\mathbf{s}. \tag{27}$$

By applying again the mean field approximation, it is easy to verify that (27) results in

$$\left\langle l_i(t,r)\right\rangle=\frac{\exp\{-\alpha_i\}}{\exp\left\{-\lambda_i\left(\left\langle s_i(t)\right\rangle-\left\langle s_i(r)\right\rangle\right)^2\right\}+\exp\{-\alpha_i\}}. \tag{28}$$

The mean field values needed in (20) should be calculated iteratively, since each of them depends on the status of its neighborhood. Thus, for each $A^{(k)}$, (23), (24), and (28) must be iterated from the previous mean field status.

The explicit, final form for function Q is

$$Q\left(A|A^{(k)}\right)=logP(A)-\frac{1}{2}\sum_t Tr\left\{R^{-1}(t)(-A\left\langle\mathbf{s}(t)\right\rangle\mathbf{x}'(t) -\mathbf{x}(t)\left\langle\mathbf{s}'(t)\right\rangle A'+A\left(\left\langle\mathbf{s}(t)\right\rangle\left\langle\mathbf{s}'(t)\right\rangle+B(t)\right)A')\right\}. \tag{29}$$

Alternately evaluating $Q(A|A^{(k)})$ for the current $A^{(k)}$ and maximizing it for A yields an approximated MAP estimate $\hat{A}$ for the mixing matrix. As soon as this procedure has reached convergence, the mean field evaluated for $A=\hat{A}$, $(\hat{\mathbf{s}},\hat{\mathbf{l}})=(\langle\mathbf{s}(\hat{A})\rangle,\langle\mathbf{l}(\hat{A})\rangle)$, is an approximated maximum of the posterior marginals estimate for the source images [29]. Note that, although each binary edge element can only assume values 0 or 1 the mean edge field elements $\langle l_i(t,r)\rangle$ are continuous valued. A final edge map can be extracted from the mean field by simple thresholding.

In the cases where a uniform prior is assumed for A, the maximizer at each M-step can be obtained by setting to zero the derivative of (29) with respect to A, thus yielding

$$A^{(k+1)}=\left[\sum_t R^{-1}(t)\mathbf{x}(t)\left\langle\mathbf{s}'(t)\right\rangle\right] \times\left[\sum_t R^{-1}(t)\left(\left\langle\mathbf{s}(t)\right\rangle\left\langle\mathbf{s}'(t)\right\rangle+B(t)\right)\right]^{-1} \tag{30}$$

where, in the particular case of space-invariant noise, the inverse covariance matrix R^{-1} disappears.

When a nonuniform prior is assumed for A, the function Q can be nonconcave. In this case, a simulated annealing optimization strategy can be a viable option, since the small number of variables makes the computational cost relatively low.

V. Experimental Results

The efficiency of the EM algorithm described in Section IV was tested on both synthetic and real piecewise smooth images. For generating the synthetic images we did not refer to any specific probability law, and, in this first application, we did not assume any a priori knowledge about the mixing matrix elements. As far as the prior is concerned, in the absence of specific knowledge about the single sources, we assumed the same Gibbs prior for each of them, with the same hyperparameters λ_i and α_i. These were selected empirically, based on a rough evaluation of the noise amount and the intensity jumps in the mixtures. However, it is to be noted that our computational scheme could be augmented with a step for the automatic, dynamic selection of the MRF hyperparameters, based on ML or MAP criteria. We refer to [35] for a recent survey on this subject. Furthermore, when appropriate, assigning different priors/hyperparameters to the different sources permits to avoid the typical permutation ambiguity of BSS; we verified that the order in which we assign the priors to the sources is reflected in the order of the reconstructions [18], [24]. To avoid premature "freezing" of the source estimates, we employed a few steps of the iterative scheme for computing the mean values of the sources and the

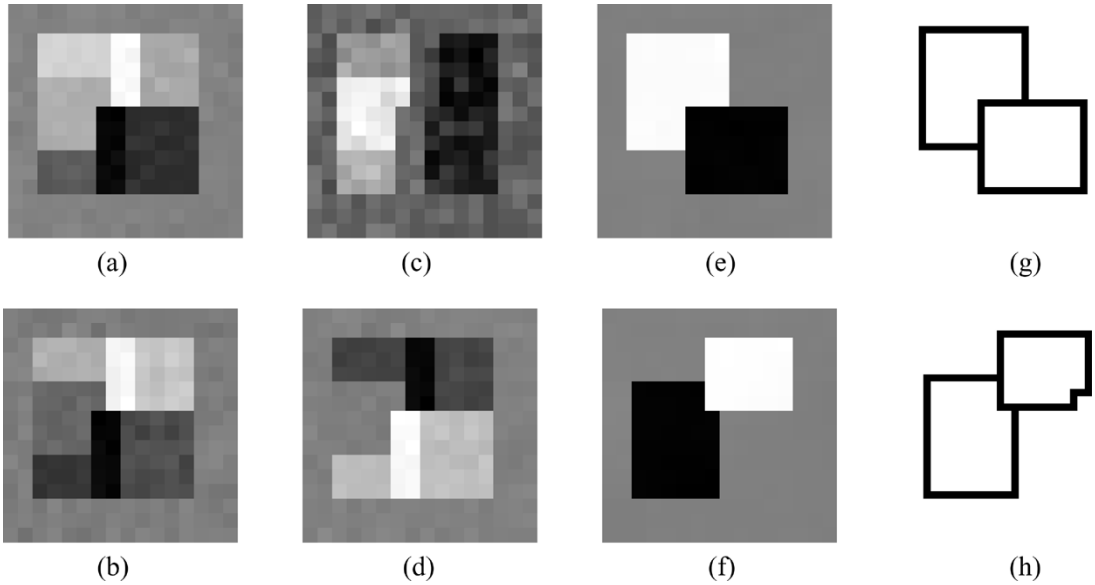


Fig. 2. Synthetic images: (a) first mixture (26-dB SNR); (b) second mixture (26-dB SNR); (c) first output of FastICA; (d) second output of FastICA; (e) first demixed image; (f) second demixed image; (g) edge map of the first demixed image; (h) edge map of the second demixed image.

related edge maps. Since no prior was assumed for A, we employed the updating rule (30) for estimating the mixing matrix at each iteration.

In all the experiments, the mixtures were generated numerically, by letting the ideal matrix coefficients to be selected randomly. In most cases, the sources were cross correlated, and we considered both space-invariant and space-variant noise. However, the results reported below mainly refer to the space-invariant noise situation. The starting point for the mixing matrix was always chosen randomly, while the starting point for the sources was always set to the mixtures, and the starting point for the edge maps was always zero (no edges). Although the EM strategy, as any alternate optimization scheme, is known to be suboptimal, i.e., dependent on the initialization of the parameters to be estimated, in our experiments this choice of the starting points always led to stabilization of the solutions and to estimates close to the ideal mixing matrices and sources. On the other hand, for the kind of cost function we considered, the uniqueness of a global optimum, as well as the number and the locations of local optima, are very hard to study.

For a quantitative analysis of the method, we focused both on the matrix estimate and the source estimates. With respect to the source estimates, we considered the following normalized least-squares error (NLSE) for each couple of original and estimated sources

$$\mathrm{NLSE}(i) = \sqrt{\frac{\sum_t \left(s_i(t) - \hat{s}_i(t)\right)^2}{\sum_t s_i(t)^2}} \tag{31}$$

where s_i and $\hat{s}_i$ are the original and the estimated ith source, respectively. With respect to the matrix estimate, we defined two quality indexes: 1) the root mean squared error (rmse) between the original and estimated matrices and 2) the element-by-element ratio between the two matrices. This latter is someway more suitable, since we know that, due to the scale ambiguity, each column of the mixing matrix can be estimated up to a scale factor. It is to be noted, however, that these indices can be computed only when separation is achieved, in such a way that the estimated sources can be correctly associated with the original ones. In particular, since in some cases a permutation in the reconstructions and then in the matrix columns can occur, this must be corrected before computing the quality indices. Of course, when separation is not achieved, the estimated sources cannot be recognized, and the estimated mixing matrix is very far from the ideal one, so that the quality indices defined cannot be computed.

In a first set of experiments, we considered two simple piecewise constant synthetic images, generated in such a way to be very little correlated. Fig. 2 shows typical results from mixtures with a 26-dB signal-to-noise ratio (SNR). By trying with several randomly selected mixing matrices and several noise realizations, we always obtained similar results in terms of rmse.

It is to be noted that for these synthetic images the FastICA algorithm is not able to separate the mixture in any of the cases we considered, even for very little amounts of noise. This fact is probably due to the small number of levels present in the original images. Fig. 2(a) and (b) shows the mixtures, Fig. 2(c) and (d) shows the results of the FastICA algorithm, and Fig. 2(e) and (f) show the images demixed with our method, for the following original mixing matrix

$$A = \begin{bmatrix} 0.6998 & 0.3005 \\ 0.2977 & 0.4026 \end{bmatrix}.$$

The matrix recovered with our method was in this case

$$\hat{A} = \begin{bmatrix} 0.6961 & 0.3020 \\ 0.2901 & 0.4002 \end{bmatrix}$$

obtained with $\lambda = 1$ and $\alpha = 0.009$. The matrix rmse was 0.0045, while the element-by-element ratio was

$$\frac{\hat{A}}{A} = \begin{bmatrix} 0.9947 & 1.0050 \\ 0.9745 & 0.9940 \end{bmatrix}.$$

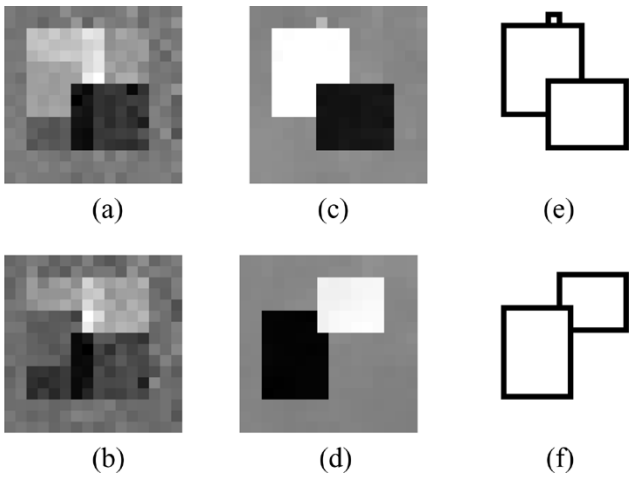


Fig. 3. Synthetic images with space-variant noise (14-dB SNR at pixels with $i < j$ and 20-dB SNR at pixels with $i \geq j$): (a) first mixture; (b) second mixture; (c) first demixed image; (d) second demixed image; (e) edge map of the first demixed image; (f) edge map of the second demixed image.

The mixing matrix estimated by FastICA was instead

$$\hat{A} = \begin{bmatrix} 0.3476 & -0.6804 \\ -0.0334 & -0.4962 \end{bmatrix}.$$

For the same images and matrix of the experiment above, we also attempted to separate from mixtures with higher noise levels, and we still obtained satisfactory results, although we obviously observed a degradation as the noise level increases. As far as the hyperparameters are concerned, we left λ unchanged and adopted higher values of α for higher noise, in order to avoid detection of spurious edges in correspondence of noise peaks. This strategy was also followed for the other experiments.

For 20-, 16.5-, and 14-dB SNR, the mixing matrices were estimated with rmse $= 0.0091$, rmse $= 0.0173$, and rmse $= 0.0251$, respectively. As far as the images are concerned, from a visual inspection we always observed a more than satisfactory quality of the reconstructions. Moreover, in all the cases considered above, the NLSE of (31) between the original and reconstructed images, scaled to zero mean and unit variance, was no greater than 0.14. In particular, for the case of 26-dB noise, we obtained $\mathrm{NLSE}(1) = 0.015$ and $\mathrm{NLSE}(2) = 0.020$, while for the case of 14-dB noise, we obtained $\mathrm{NLSE}(1) = 0.082$ and $\mathrm{NLSE}(2) = 0.135$.

As an example, Fig. 3 shows the mixtures, the reconstructed images, and related edge maps for the case of space-variant noise (14-dB SNR at pixels with $i < j$ and 20-dB SNR at pixels with $i \geq j$). In this case, the rmse between the original and estimated matrices was 0.0179.

The performance of the method on real images artificially mixed was similar to the one described for synthetic images. We considered two images of real texts, mixed with the following matrix:

$$A = \begin{bmatrix} 0.7451 & 0.7884 \\ 0.5879 & 0.2596 \end{bmatrix}.$$

In this case, due to the larger size of these images and the higher number of gray levels, for little amounts of noise (e.g., with 26-dB SNR) FastICA is able to separate quite well too, apart from a sensible noise component in the reconstructed images. However, for increasing amounts of noise, the reconstructions obtained with our algorithm remain satisfactorily good, whereas the ones from FastICA degrade quickly. For a 14-dB SNR, we obtained the results shown in Fig. 4. The matrix estimated with our method was

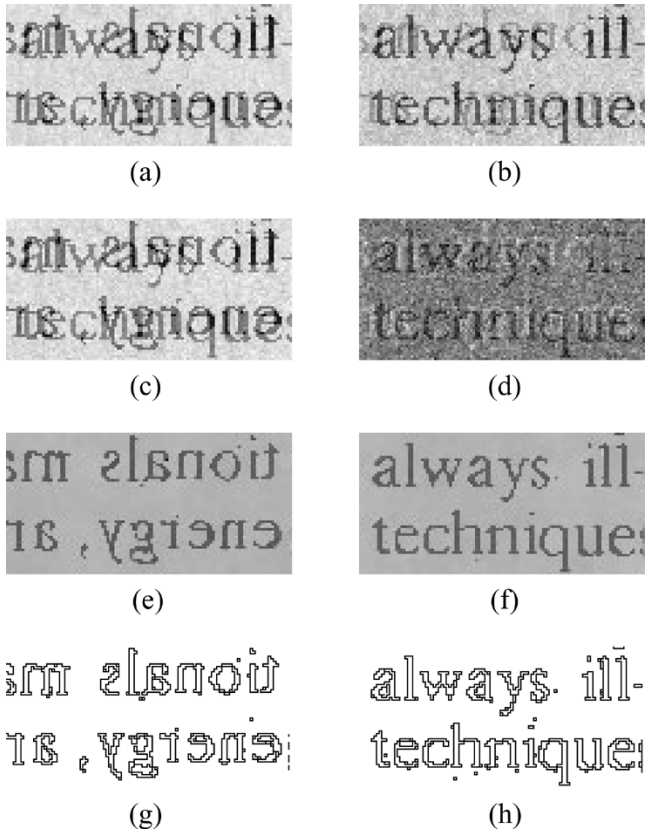


Fig. 4. Real images: (a) first mixture (14-dB SNR); (b) second mixture (14-dB SNR); (c) first FastICA output; (d) second FastICA output; (e) first demixed image; (f) second demixed image; (g) edge map of the first demixed image; (h) edge map of the second demixed image.

$$\hat{A} = \begin{bmatrix} 0.7999 & 0.7552 \\ 0.2643 & 0.5947 \end{bmatrix}$$

while the element-by-element ratio with the original (after column permutation) was

$$\frac{\hat{A}}{A} = \begin{bmatrix} 1.0135 & 1.0146 \\ 1.0115 & 1.0181 \end{bmatrix}.$$

The matrix estimated by the FastICA algorithm was instead

$$\hat{A} = \begin{bmatrix} 1.1594 & 0.0873 \\ 0.5926 & 0.3691 \end{bmatrix}.$$

The NLSEs on the images reconstructed with our method were, respectively, $\mathrm{NLSE}(1) = 0.205$ and $\mathrm{NLSE}(2) = 0.173$. It is to be noted that the two source images used here, two portions of a scanned document, present a nonnegligible correlation of 0.129

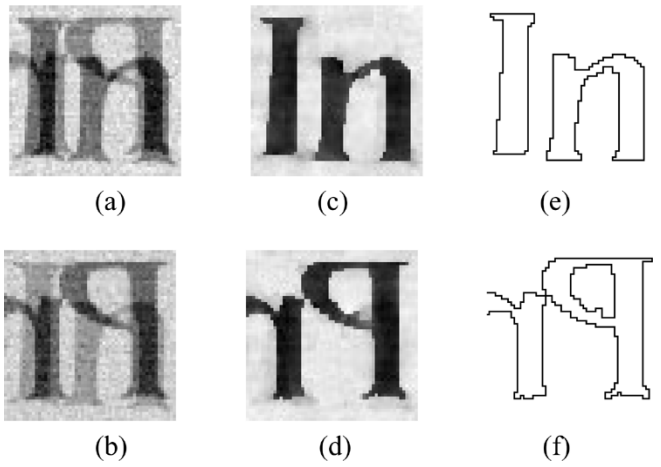


Fig. 5. Real images: (a) first mixture (14-dB SNR); (b) second mixture (14-dB SNR); (c) first demixed image; (d) second demixed image; (e) edge map of the first demixed image; (f) edge map of the second demixed image.

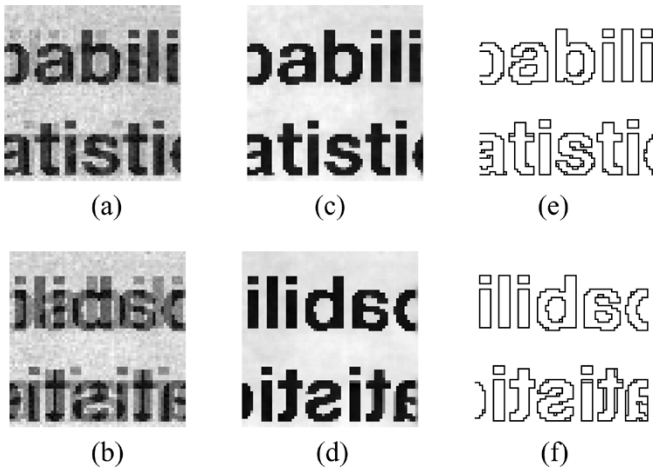


Fig. 6. Real images: (a) first mixture (26-dB SNR); (b) second mixture (26-dB SNR); (c) first demixed image; (d) second demixed image; (e) edge map of the first demixed image; (f) edge map of the second demixed image.

(computed on the images scaled to zero mean and unit variance). This could give an extra reason to the bad performance of FastICA, while, at the same time, highlights the effectiveness of using MRF image models with cross-correlated sources.

Fig. 5 shows the results of another experiment, where we mixed real images and then added a noise of 14-dB SNR. Note that, also in this case, the original sources have a correlation of 0.188.

To further test the robustness of the method against cross-correlated sources, we performed another experiment where we mixed a scan of a real text with its horizontally flipped version. In this case, the correlation between the two sources is 0.22. The results obtained when the additive noise on the mixtures has an SNR of 26 dB are shown in Fig. 6. The mixing matrix was recovered with an rmse of 0.0027 with respect to the original one.

Another set of experiments was aimed at showing a potential application of our method to the analysis of ancient and/or degraded documents. Indeed, in images of ancient documents it is very frequent the presence of two or more overlapped texts. In some cases, textures interfering with the main text are

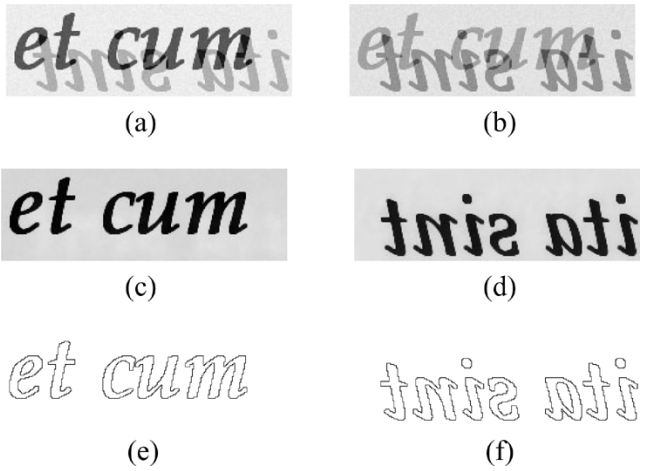


Fig. 7. Show-through image (26-dB SNR): (a) red channel; (b) green channel; (c) first demixed image; (d) second demixed image; (e) edge map of the first demixed image; (f) edge map of the second demixed image.

due to seeping or transparency of the ink from the reverse side, and should be removed to improve both human and automatic readability. In other cases, e.g., in palimpsests, faint traces of an underwriting, that is an older text that has been erased to reuse the support, usually parchment, should be enhanced and extracted to make it readable to interested scholars. In both situations, if an approximated linear model is adopted to describe the phenomenon of overlapping texts, BSS techniques can be thought as the appropriate methodology for recovering the separated patterns from multiple observations, in different visible or nonvisible bands, of the document itself [30]. Furthermore, since text patterns, at least in their pristine state, are constituted of homogeneous patches (the characters) with sharp edge boundaries, the MRF models described here could represent a very efficient and feasible way to describe the source features.

In a first example, we considered a synthetic image generated so as to simulate a real bleed-through or show-through effect. We mixed two texts by a 3×2 matrix, added a noise realization to obtain a 26-dB SNR, and considered the three mixtures as the red, green and blue channels of a color image. We then used the red and green channels [shown in Fig. 7(a) and (b)] as inputs for separating the two texts. With our method, the separation gave the results shown in Fig. 7(c)–(f).

In the second example, we considered the color image of a real-fake palimpsest, depicted in grayscale in Fig. 8(e). In particular, this one simulates the colors and the mutual position of the two overlapped texts as they appear in the famous Archimedes palimpsest [31]. Note that the underwriting is the lighter text. By processing the green and blue channels of the color image, we obtained the two separations shown in Fig. 8(a) and (b), and the related edge maps shown in Fig. 8(c) and (d). In this case, the linear model is inadequate in correspondence of the areas where the overwriting completely masks the underwriting, so that this latter was not satisfactorily recovered in those areas. However, a simple global thresholding is sufficient to significantly improve the text readability, as shown in Fig. 8(f).

Besides document restoration and analysis, another possible application, that we are going to investigate, is the analysis of remote sensed data, where our method could be effectively used to classify different geological and vegetation regions.

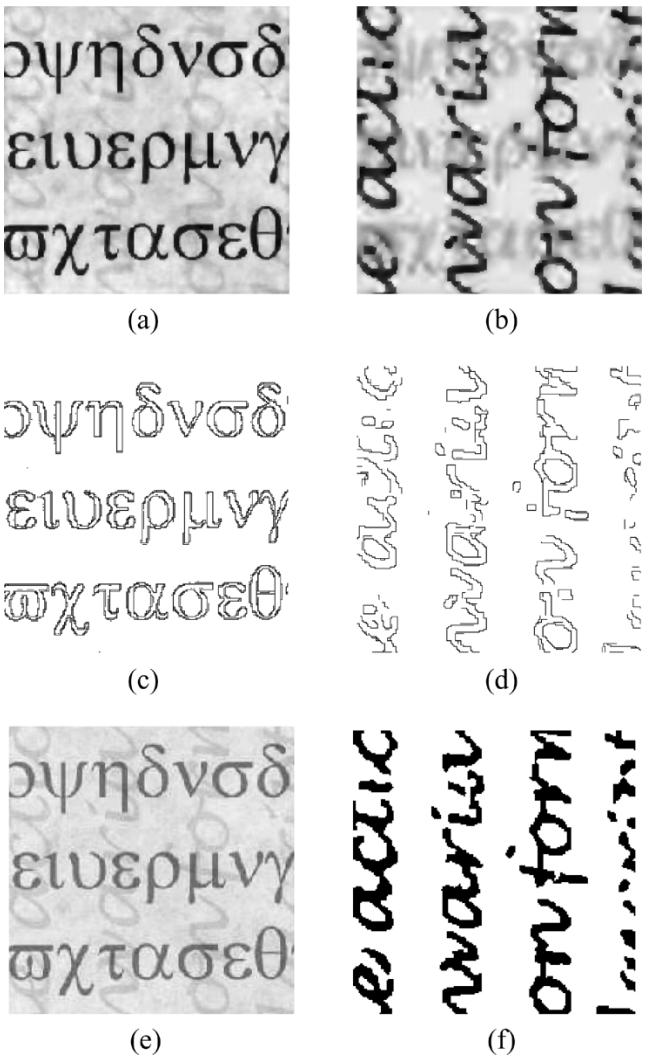


Fig. 8. Real-fake palimpsest: (a) first demixed image; (b) second demixed image; (c) edge map of the first demixed image; (d) edge map of the second demixed image; (e) grayscale version of the original color image; (f) recovered underwriting after thresholding.

VI. Conclusion

We propose a Bayesian formulation of the linear BSS problem; this allows us to adopt an EM strategy to solve the problem as one of MAP estimation. The data model adopted is the classical noisy linear mixture with unknown coefficients and additive Gaussian noise. Our source model is a pair of interacting MRFs, which permits the mutual independence and the local autocorrelation of the sources to be enforced.

To make our EM approach feasible, we adopted a mean-field approximation. The status of the mean field at convergence approximates the MAP estimates of the source maps (intensity process) and of the regions where the spatial autocorrelation within the individual sources is broken (line process). Our procedure can, thus, be adopted to perform joint parameter and source estimation and segmentation, which is helpful in subsequent classification and/or recognition tasks. This could be used, for example, in text analysis and remote sensed area classification.

Besides the additional advantage of yielding edge maps, our extensive experimentation with both synthetic and real images has confirmed the beneficial effect of introducing spatial correlation into the source model, which had been already observed in the case of one-dimensional signals. In particular, we verified the robustness of the method in dealing with noise, even space variant. Furthermore, for sources whose characteristics closely fit the model we adopted (e.g., text images), independence is no more a strict requirement, and cross-correlated sources can be separated as well.

A refinement of the strategy to automatically select the hyperparameters and introduce constraints on the interactions between the edges, as well as a further experimentation on images from different applications is expected for the future.

References

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2002.

[2] M. Burger and O. Scherzer, "Regularization methods for blind deconvolution and blind source separation," *Math. Control Signals Syst.*, vol. 14, pp. 358–383, 2001.

[3] S. Geman and D.Geman , "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–740, Jun. 1984.

[4] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solutions of ill-posed problems in computational vision," *J. Amer. Stat. Assoc.*, vol. 82, pp. 76–89, 1987.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B*, vol. 39, pp. 1–38, 1977.

[6] A. J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, pp. 1129–1159, 1995.

[7] T. Lee, M. Lewicki, and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources," *Neural Comput.*, vol. 11, pp. 409–433, 1999.

[8] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, pp. 157–192, 1999.

[9] S. Amari and A. Cichocki, "Adaptive blind signal processing—neural network approaches," *Proc. IEEE*, vol. 86, pp. 2026–2048, 1998.

[10] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, Jun. 1999.

[11] ——, "Gaussian moments for noisy independent component analysis," *IEEE Signal Process. Lett.*, vol. 6, no. 2, pp. 145–147, Feb. 1999.

[12] H. Attias, "Independent factor analysis," *Neural Comput.*, vol. 11, pp. 803–851, 1999.

[13] E. Moulines, J. F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, 1997, pp. 3617–3620.

[14] E. Kuruoglu, L. Bedini, M. T. Paratore, E. Salerno, and A. Tonazzini, "Source separation in astrophysical maps using independent factor analysis," *Neural Netw.*, vol. 16, pp. 479–491, 2003.

[15] K. Knuth, "Bayesian source separation and localization," in *Proc. SPIE Bayesian Inference for Inverse Problems*, 1998, pp. 147–158.

[16] S. E. Lee and S. J. Press, "Robustness of Bayesian factor analysis estimates," *Commun. Stat.—Theory Meth.*, vol. 27, pp. 1871–1893, 1998.

[17] A. Mohammad-Djafari, "A Bayesian approach to source separation," presented at the 19th Int. Workshop on Maximum Entropy and Bayesian Methods, 1999.

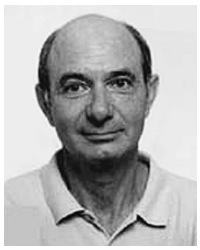
[18] A. Tonazzini, L. Bedini, E. E. Kuruoglu, and E. Salerno, "Blind Separation of Time-Correlated Sources From Noisy Data," Tech. Rep. IEI-CNR TR-42-2001, 2001.

[19] L. Bedini, I. Gerace, E. Salerno, and A. Tonazzini, "Models and algorithms for edge-preserving image reconstruction," in *Advances in Imaging and Electron Physics*, P. W. Hawkes, Ed. New York: Academic, 1996, vol. 97, pp. 86–189.

[20] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 4, pp. 576–586, Apr. 1995.
[21] ——, "Close-form solution and parameter selection for convex minimization-based edge-preserving smoothing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 916–932, Aug. 1998.
[22] L. Bedini, I. Gerace, and A. Tonazzini, "A GNC algorithm for constrained image reconstruction with continuous-valued line processes," *Pattern Recognit. Lett.*, vol. 15, pp. 907–918, 1994.
[23] D. Shulman and J. Y. Herve, "Regularization of discontinuous flow fields," in *Proc. Workshop Vis. Motion*, 1989, pp. 81–86.
[24] L. Bedini, E. Salerno, and A. Tonazzini, "Blind source separation from noisy data using Bayesian estimation and Gibbs priors," in *Proc. IASTED Sig. Image Processing Conf.*, 2001, pp. 108–112.
[25] D. Chandler, *Introduction to Modern Statistical Mechanics*. Oxford, U.K.: Oxford Univ. Press, 1987.
[26] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. Signal Process.*, vol. 40, no. 11, pp. 2570–2583, Nov. 1992.
[27] ——, "The mean field theory in EM procedures for blind Markov random field image restoration," *IEEE Trans. Image Process.*, vol. 2, no. 1, pp. 27–40, Jan. 1993.
[28] A. Tonazzini, L. Bedini, E. E. Kuruoglu, and E. Salerno, "Blind separation of auto-correlated images from noisy mixtures using MRF models," in *Proc. ICA*, Nara, Japan, Apr. 2003, pp. 675–680.
[29] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 5, pp. 401–412, May 1991.
[30] A. Tonazzini, L. Bedini, and E. Salerno, "Independent component analysis for document restoration," *Int. J. Doc. Anal. Recognit.*, vol. 7, pp. 17–27, Mar. 2004.
[31] R. L. Easton. Simulating Digital Image Processing used for the Archimedes Palimpsest. [Online]. Available: http://www.cis.rit.edu/people/faculty/easton/k-12/index.htm
[32] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
[33] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–685, 1989.
[34] J.-M. Morel and S. Solimini, "Variational methods in image segmentation with 7 image processing experiments," in *Progress in Nonlinear Differential Equations and Their Applications*. Basel, Switzerland: Birkhäuser, 1994, vol. 14.
[35] A. Tonazzini and L. Bedini, "Degradation identification and model parameter estimation in discontinuity-adaptive visual reconstruction," *Adv. Imag. Electron. Phys.*, vol. 120, pp. 193–284, 2002.

Anna Tonazzini received the degree in mathematics (cum laude) from the University of Pisa, Pisa, Italy, in 1981.

In 1984, she joined the Istituto di Scienza e Tecnologie dell'Informazione of the Italian National Research Council (CNR), Pisa, where she is currently a Researcher with the Signals and Images Laboratory. She has cooperated in special programs for basic and applied research on image processing and computer vision and is the coauthor of over 60 scientific papers. Her present interests are in inverse problems theory, image restoration and reconstruction, document analysis and recognition, independent component analysis, neural networks, and learning.



Luigi Bedini received the degree in electronic engineering (cum laude) from the University of Pisa, Pisa, Italy, in 1968.

Since 1970, he has been a Researcher with the Italian National Research Council, Istituto di Scienza e Tecnologie dell'Informazione, Pisa. From 1971 to 1989, he was an Associate Professor of system theory at the Computer Science Department, University of Pisa. His interests have included modeling, identification and parameter estimation of biological systems applied to noninvasive diagnostic techniques. At present, his research interests are in the fields of digital signal processing, image reconstruction, and neural networks applied to image processing. He is the coauthor of more than 80 scientific papers.

Emanuele Salerno received the degree in electronic engineering from the University of Pisa, Pisa, Italy, in 1985.

In September 1987, he joined the Italian National Research Council (CNR), Department of Signal and Image Processing, Information Processing Institute (now Institute of Information Science and Technologies, ISTI, Signals and Images Laboratory), Pisa, where he has been working on applied inverse problems, image reconstruction and restoration, microwave nondestructive evaluation, and blind signal separation. He has been assuming various responsibilities in research programs in nondestructive testing, robotics, numerical models for image reconstruction and computer vision, and neural network techniques in astrophysical imagery.

Dr. Salerno is a member of the Italian society for information and communication technology (AICT-AEIT).