

Multichannel Blind Separation and Deconvolution of Images for Document Analysis

Anna Tonazzini, Ivan Gerace, and Francesca Martinelli

Abstract—In this paper, we apply Bayesian blind source separation (BSS) from noisy convolutive mixtures to jointly separate and restore source images degraded through unknown blur operators, and then linearly mixed. We found that this problem arises in several image processing applications, among which there are some interesting instances of degraded document analysis. In particular, the convolutive mixture model is proposed for describing multiple views of documents affected by the overlapping of two or more text patterns. We consider two different models, the interchannel model, where the data represent multispectral views of a single-sided document, and the intrachannel model, where the data are given by two sets of multispectral views of the recto and verso side of a document page. In both cases, the aim of the analysis is to recover clean maps of the main foreground text, but also the enhancement and extraction of other document features, such as faint or masked patterns. We adopt Bayesian estimation for all the unknowns and describe the typical local correlation within the individual source images through the use of suitable Gibbs priors, accounting also for well-behaved edges in the images. This *a priori* information is particularly suitable for the kind of objects depicted in the images treated, i.e., homogeneous texts in homogeneous background, and, as such, is capable to stabilize the ill-posed, inverse problem considered. The method is validated through numerical and real experiments that are representative of various real scenarios.

Index Terms—Blind image deconvolution, blind source separation, document image processing, Markov random fields (MRFs), parameter learning.

I. INTRODUCTION

IMPROVING the human and automatic readability of printed or handwritten documents is a common need in libraries and archives. Indeed, especially when they are ancient, documents are often affected by several types of degradations. One of the most frequent degradation is bleed-through or show-through. Bleed-through occurs when some patterns interfere with the main text due to seeping of ink from the reverse page side. Show-through may appear also in scans of modern,

Manuscript received June 10, 2009; revised October 07, 2009. First published December 18, 2009; current version published March 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Eero P. Simoncelli.

A. Tonazzini is with the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, I-56124 Pisa, Italy (e-mail: anna.tonazzini@isti.cnr.it).

I. Gerace and F. Martinelli are with the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, I-56124 Pisa, Italy, and also with the Dipartimento di Matematica e Informatica, Università degli Studi di Perugia, I-06123 Perugia, Italy (e-mail: ivan.gerace@isti.cnr.it; gerace@dipmat.unipg.it; francesca.martinelli@isti.cnr.it; martinelli@dipmat.unipg.it).

Digital Object Identifier 10.1109/TIP.2009.2038814

well-preserved documents, when the paper is not completely opaque. Removing the bleed-through or show-through patterns from a digital image of a document is not trivial, especially with ancient originals, where interferences of this kind are usually very strong. Indeed, dealing with strong bleed-through degradation is practically impossible by any simple thresholding technique, since the intensities of the unwanted background can be very close to those of the main text. Thus, adaptive and/or structural approaches have to be adopted [1], [2].

Most of the work done in this respect has exploited information from the grayscale front and back pages, usually referred as recto and verso sides [3], [4]. Besides requiring a preliminary registration of the two sides, these techniques are usually expensive, as they are based on steps of segmentation, to identify the bleed-through areas, followed by inpainting of estimated pure background areas [5]. More recently, variational approaches, based on nonlinear diffusion, have been proposed to model and then remove this kind of degradations [6]. In [7], a color scan of a single side is required, but a thresholding technique can only be used in the framework of multiresolution analysis and adaptive binarization. Other authors propose segmentation of the different color clusters in the image via adaptation of the k-means algorithm [8], [9]. In [10], for the grayscale scan of a single-sided document, a classification approach based on a double binary Markov random field (MRF), one for the recto and the other for the verso, is proposed.

However, this problem has mainly been addressed from the point of view of removing all the structured background for obtaining a clean foreground text. Specific situations in which the interfering texts, or some of them, can be of interest themselves have received little attention. Consider, for instance, the cases of underwritings in palimpsests, or stamps, or paper watermarks. Often these patterns represent the most significant information from a cultural and historical point of view, whereas they are usually barely perceivable in the originals. Hence, the goal of document restoration should be their enhancement and recovery. In other cases, even the recovery of the bleed-through/show-through pattern can be of interest *per se*, for instance when the verso scan is not available. In the present approach, based on blind source separation (BSS) techniques, all the overlapped texts are the unknown sources to be recovered and enhanced. Multiple observations of the document, modelled as mixtures of the sources, can be considered as the data set. For instance, multispectral or hyperspectral views of a single document side (e.g., the red, green and blue channels and/or nonvisible channels) can be exploited for the purpose. Alternatively, recto and verso scans, even in grayscale, of a document page can be used as the available data set. This latter setting has application to the large databases of digitized documents that already exist in

many archives and libraries, where the documents have been acquired in grayscale only or as black and white microfilms.

A. Convulsive Blind Source Separation

Blind Source Separation (BSS) and Independent Component Analysis (ICA) techniques [11]–[13] were formerly developed for signal processing problems such as the “cocktail party” problem in audio and speech processing. Recently, these techniques have shown a great potential for solving important image processing and computer vision problems [14], as well. Indeed, in many imaging fields, we have to deal with observations or maps that are mixtures of images with unknown coefficients or kernels. These maps cannot be properly interpreted, unless some strategy is adopted for separately extracting the various component images.

In the literature, the original BSS framework of linear and instantaneous mixtures has been mainly considered. In this formulation, each pixel of the observations solely results from a linear combination of the component values in the corresponding pixel. The linear and instantaneous assumption is often physically grounded, as for instance in the case of radiation sky maps in astrophysics [15]. For the application considered in this paper, i.e., the analysis of ancient documents degraded by the overlapping of two or more texts or patterns, it represents a first approximation of more complex combination phenomena. Within this approximation, in [16], [17] we successfully applied linear ICA techniques, or even second-order statistics methods, to remove interfering artifacts from RGB single-sided documents, recover legible front and reverse text patterns from recto and verso pairs of documents showing bleed-through or show-through, and enhance underwritings in palimpsests from multispectral observations.

Nevertheless, the physical model underlying text overlapping is very complicated, since it derives from complex chemical processes of ink diffusion and paper absorption, which are expected to be nonlinear and nonstationary across the document page. As a fundamental aspect, it is apparent that, in the pixels where two texts are superimposed to each other, the resulting intensity is not the vector sum of the intensities of the two components, but it is likely to be some nonlinear combination of them. For example, in [3], a nonlinear model is derived for the phenomenon of show-through, although a linear approximation has then been proposed to permit tractability. Even adopting a linear approximation and assuming stationarity, we cannot neglect the often strong blur on the text patterns, due to ink diffusion or light spreading through the support, and the noise affecting the scans. This blur can be different from pattern to pattern and from observation to observation, thus making inconsistent the instantaneous assumption.

A step towards a mathematical model that is more adhering to the physics of the problem is to model the observations as convulsive mixtures of the component images. In this formulation, the value of the generic pixel in the observations results from a weighted average of the component values in the supports of the blur kernels. The problem is a special case of multichannel blind deconvolution [18], [19], a classical example being the recovery of the original color of a blurred color image [20], [21]. In that

instance, the imaging system captures multiple views of a single scene or, at least, there is no need to extract two or more individual components from the scene, which is instead the subject of this paper.

Research on blind separation methods from convulsive mixtures has largely been addressed to one dimensional signals, with application in speech and audio processing, in multiaccess digital communication systems and array processing, and in biomedical signal processing. For convulsive mixtures of narrow-band signals, the model is very similar to that of instantaneous linear mixtures, and can be obtained by replacing the real coefficients by complex coefficients corresponding to attenuation and phase. Thus, methods similar to those employed for instantaneous linear mixtures can be used. For convulsive mixtures of wide-band signals, frequency-domain analysis is very popular since, again, all techniques for instantaneous BSS can be applied independently in each frequency bin [22]. Unfortunately, the permutation indeterminacy, which is inherent in BSS, may appear independently in each frequency bin so that the reconstructions can be wrong, unless extra measures are taken to avoid this internal permutation [23]. In the time-based model, separation of convulsive mixtures may be achieved by estimating the coefficient vectors of FIR filters, rather than the only scalar coefficients. Various methods have been proposed, based on second- or higher-order statistics [24], [25], ensemble learning [26], contrast functions [27], and the natural gradient [28]. Some specific works on blind separation of convulsive mixtures of images have regarded, for instance, photography of semireflections, or microscopy and tomography [29]–[31]. For show-through removal from double-sided documents, paper [32] proposes a convulsive BSS formulation, which accounts also for a nonlinearity assumed known and derived from [3]. The solution is based on the total variation stabilizer for the sources. In [33], a maximum likelihood approach is proposed for two nonlinear mixtures of two sources, where the nonlinearity is approximated as quadratic, and a known blur kernel on the interfering pattern is accounted for.

B. Bayesian Convulsive BSS

The Bayesian estimation setup offers a natural and flexible way to approach the integrated solution of two or more problems, and to account for prior knowledge we may have about a problem and its parameters. For instance, autocorrelation constraints have been proved to be effective for achieving stable solutions in many inverse problems, and especially in those dealing with images, where these constraints correspond to natural features of real physical maps and scenes. Bayesian estimation permits to formulate the BSS problem as the joint maximum *a posteriori* (MAP) estimation of the mixing operator and the sources [34]. In [35], we proposed a Bayesian approach to BSS of linear, instantaneous mixtures of images, employing MRF models to describe the image local autocorrelation structure, and showed examples of application to document analysis. In [36], still for linear, instantaneous mixtures, Bayesian estimation is coupled with hidden Markov models for document restoration. This kind of prior for the sources allows also for retaining the independence assumption of ICA, and provides robustness against noise in the data.

Here, we propose to integrate demixing and deblurring, by augmenting Bayesian, MRF-based BSS with a step of blind image deconvolution, and extend the MRF model to account for regularity constraints on the image edges as well. The joint maximization of the posterior distribution with respect to all the parameters can be reformulated as the MAP estimation of the mixing operator alone (both blur kernels and mixing coefficients), where the sources are kept clamped to their MAP estimate, for any status of the mixing. This theoretical scheme would ensure convergence to the global maximum of the posterior. However, to reduce computational complexity, reasonable approximations are derived that allow also for finding a remedy to other drawbacks, such as the unavailability of analytical formulas for the sources viewed as functions of the mixing, and nonconvexity of the priors. This practical scheme ensures stability of the solutions and employs gradient ascent algorithms of the kind of the graduated nonconvexity (GNC) algorithm [37], [38] to update the sources, and simulated annealing (SA) [39] to estimate the blur and mixing coefficients. Through numerical simulations, we will show that accounting for the blur model and adopting proper MRF image models improves the separation process.

The paper is organized as follows. In Section II, our data model will be established, the principles of Bayesian convolutive BSS will be reviewed, and the estimation approach will be formulated. In Section III, the adopted MRF edge-preserving priors will be described, and the estimation algorithm will be derived. Section IV will propose a convolutive BSS formulation for the analysis and restoration of degraded documents, which is our considered case study. Finally, Section V will be devoted to the experimental results, and Section VI to conclusions and remarks.

II. DATA MODEL AND BAYESIAN ESTIMATION

A. Data Generation Model

As shown in Fig. 1, according to the BSS and the image degradation formalism, the data generation model we consider in this paper is given by

$$\begin{aligned} x_i(t) &= \sum_{j=1}^M a_{ij} v_{ij}(t) + n_i(t), \quad i = 1, 2, \dots, N \\ t &= 1, 2, \dots, T \\ v_{ij} &= \mathbf{H}_{ij} s_j, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, M \end{aligned} \quad (1)$$

where $x_i(t)$, $s_j(t)$ and $n_i(t)$ represent the i th observation, j th source, and i th noise or measurement error at location t , respectively. Of course, in imaging, location t stands for the couple of pixel indices. In (1) and through all the paper, $\mathbf{A} = \{a_{ij}\}$ is the $N \times M$ unknown mixing matrix, vectors $s_j = (s_j(1), s_j(2), \dots, s_j(T))^T$, $j = 1, 2, \dots, M$, represent the lexicographically ordered notation of the various sources, $s(t)$ is the column vector of all the M unknown sources at location t , and $\mathbf{s} = (s(1), \dots, s(T))$ is the matrix whose t th column contains the M sources at location t , and whose j th row is the source s_j . These definitions extend to data and noise as well, apart from N that represents the number of observations and then substitutes M . Quantity v_{ij} is the degraded version

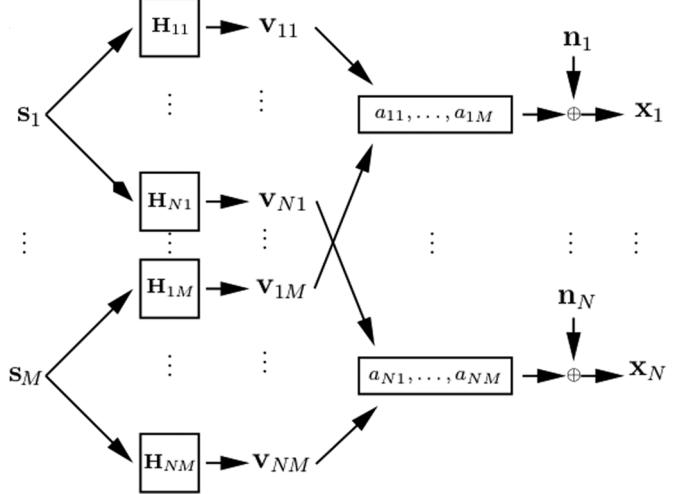


Fig. 1. Data generation model.

of source s_j which contributes to \mathbf{x}_i , and the blur matrix \mathbf{H}_{ij} is the block Toeplitz matrix that performs convolution between a source image, lexicographically ordered, and a blur mask (or kernel) \mathbf{B}_{ij} as a matrix-vector product. Below, the set $\{\mathbf{H}_{ij}, i = 1, \dots, N, j = 1, \dots, M\}$ of all convolution matrices will be indicated with \mathbf{H} .

According to (1), the data images are noisy linear instantaneous mixtures of the blurred sources through matrix \mathbf{A} , but are convolutive mixtures of the ideal, unblurred sources. Thus, the problem becomes one of BSS from noisy convolutive mixtures or, in other words, of joint blind deconvolution and blind separation. In the usual notation, mixing matrix \mathbf{A} does not appear explicitly in (1), since, in a natural way, its coefficients can be incorporated in the kernels, thus assuming the meaning of filter gains. The scale indeterminacy of the instantaneous mixing model translates here in the fact that the sources can only be determined apart from an undetermined kernel. In our case, we choose to distinguish the kernels from the mixing coefficients, due to possible symmetries related to the specific application, which can affect the kernels only, the mixing coefficients only, or even both, but in different ways. Without loss of generality, we can, thus, assume a unitary gain for each kernel.

B. Formulation of the Bayesian Estimation Method

The problem of estimating matrix \mathbf{A} , the blur kernels \mathbf{H} and the deblurred sources \mathbf{s} can be stated as the following joint MAP estimation problem

$$\begin{aligned} (\hat{\mathbf{s}}, \hat{\mathbf{A}}, \hat{\mathbf{H}}) &= \arg \max_{\mathbf{s}, \mathbf{A}, \mathbf{H}} P(\mathbf{s}, \mathbf{A}, \mathbf{H} | \mathbf{x}) \\ &= \arg \max_{\mathbf{s}, \mathbf{A}, \mathbf{H}} P(\mathbf{x} | \mathbf{s}, \mathbf{A}, \mathbf{H}) P(\mathbf{s}) P(\mathbf{A}) P(\mathbf{H}) \end{aligned} \quad (2)$$

where the sets of variables \mathbf{s} , \mathbf{A} and \mathbf{H} are assumed independent of each other, and quantities $P(\mathbf{s}, \mathbf{A}, \mathbf{H} | \mathbf{x})$, $P(\mathbf{x} | \mathbf{s}, \mathbf{A}, \mathbf{H})$, $P(\mathbf{s})$, $P(\mathbf{A})$, and $P(\mathbf{H})$ are the joint posterior distribution, the likelihood function, and the prior distributions, respectively. Assuming independence of the sources, $P(\mathbf{s})$ is given by

$$P(\mathbf{s}) = \prod_{j=1}^M P_j(s_j). \quad (3)$$

Problems like that of (2) are usually approached by means of alternating componentwise maximization with respect to the three sets of variables in turn. We propose the following theoretical scheme:

$$(\hat{\mathbf{A}}, \hat{\mathbf{H}}) = \arg \max_{\mathbf{A}, \mathbf{H}} P(\mathbf{x}|\mathbf{s}(\mathbf{A}, \mathbf{H}), \mathbf{A}, \mathbf{H}) \\ \times P(\mathbf{s}(\mathbf{A}, \mathbf{H})) P(\mathbf{A}) P(\mathbf{H}) \quad (4)$$

where

$$\mathbf{s}(\mathbf{A}, \mathbf{H}) = \arg \max_{\mathbf{s}} P(\mathbf{x}|\mathbf{s}, \mathbf{A}, \mathbf{H}) P(\mathbf{s}). \quad (5)$$

Hence, the original joint MAP estimation is reformulated as the MAP estimation of the whole mixing operator alone, while the sources are kept clamped to their MAP estimate, for any status of the mixing. It is worth saying that the theoretical scheme (4), (5) would converge to the same solution as problem (2) does, that is, to the global maximum of the posterior. Nevertheless, as it will be explained in Section III, its practical implementation poses some difficulties, that we handle through suitable approximations.

It is to be noted that classical ICA for instantaneous mixtures can still be applied to the model of (1) only when each source is affected by the same blur operator in every mixture, i.e., $\mathbf{H}_{ij} = \mathbf{H}_j \forall j, \forall i$. As we will see, for the specific case study of document analysis, this assumption does not hold true, at least in the recto-verso case. However, even if it was applicable, the recovered sources can be, at best, the blurred versions of the original ones, with amplified noise. More likely, since blur can introduce cross-correlation among the sources, separation is harder to be achieved with ICA in this case. On the other hand, available knowledge on the sources, the blur kernels and the mixing cannot be enforced within the ICA approach, which solely relies on source independence.

III. IMAGE MODEL DESCRIPTION AND ESTIMATION ALGORITHM

A. Likelihood and the Model Priors

Considering an independent, white and Gaussian noise with zero mean, the logarithm of the likelihood $P(\mathbf{x}|\mathbf{s}, \mathbf{A}, \mathbf{H})$ is given by

$$\log(P(\mathbf{x}|\mathbf{s}, \mathbf{A}, \mathbf{H})) = -\frac{1}{2} \sum_{t=1}^T (\mathbf{z}(t) - \mathbf{x}(t))^T \Sigma^{-1} (\mathbf{z}(t) - \mathbf{x}(t)) \quad (6)$$

where

$$\mathbf{z}(t) = (z_1(t), z_2(t), \dots, z_N(t))^T, \quad t = 1, 2, \dots, T \\ z_i(t) = \sum_{j=1}^M a_{ij} v_{ij}(t), \quad t = 1, 2, \dots, T, \quad i = 1, 2, \dots, N$$

and Σ is the covariance matrix of the noise, assumed, at present, to be uncorrelated and location-independent, that is $\Sigma = \sigma^2 \mathbf{I}$, being \mathbf{I} the $N \times N$ identity matrix, and σ^2 the noise variance. As it will be clearer later on, in our setting, σ^2 needs not to be known in advance, nor jointly estimated, since it can be considered as incorporated in the model hyperparameters.

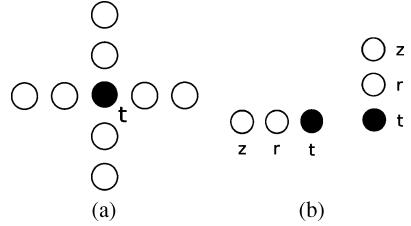


Fig. 2. (a) Neighborhood system; (b) cliques having nonzero potential.

As per the prior $P(\mathbf{s})$, in the form of (3), we adopt a local autocorrelation model for each source, in the form of generic local smoothness MRF models, augmented to account for information about the regularity features of realistic edge maps. This kind of model is particularly suitable for describing images of texts, both printed and handwritten. In addition, accounting for a well behaved edge process is particularly useful when deblurring must be performed.

In the Gibbs/MRF formalism our priors are given by

$$P_j(\mathbf{s}_j) = \frac{1}{Z_j} \exp \{-U_j(\mathbf{s}_j)\}, \quad j = 1, 2, \dots, M \quad (7)$$

where Z_j is the normalizing constant and $U_j(\mathbf{s}_j)$ is the prior energy in the form of a sum of potential functions, or stabilizers, over the set of cliques of interacting locations. The number of different cliques, as well as their shape, is related to the extent of correlation among the pixels, whereas the functional form of the potentials determines the correlation degree, and various features of the image edges. In particular, we express the regularity of edges by penalizing parallel, adjacent edges. In a previous paper, we adopted a different model, which was an extension of the Ising model for binary image fields, with the aim at segmenting the text characters in highly degraded documents, before the application of an OCR system [40]. However, herein, the goal is the analysis of the document content and its qualitative restoration, so that a continuous-valued image field seemed more appropriate. We consider the neighborhood and clique systems depicted in Fig. 2.

Hence, our $U_j(\mathbf{s}_j)$ is given as

$$U_j(\mathbf{s}_j) = \sum_{t=1}^T \sum_{(r,z) \in N_t} \psi_j((s_j(t) - s_j(r)), (s_j(r) - s_j(z))) \quad j = 1, 2, \dots, M \quad (8)$$

where N_t is the set of the two couples of adjacent locations (r, z) , $z < r$, that, in the 2D grid of pixels, precede location t in horizontal and in vertical. Stabilizers ψ_j have all the same functional form, but possibly different hyperparameters, in order to graduate the constraint strength in dependence of the source considered. We chose the following functions [38], [41]:

$$\psi_j(\xi_1, \xi_2) = \begin{cases} \lambda_j \xi_1^2, & \text{if } |\xi_1| < \theta_j, \\ \alpha_j, & \text{if } |\xi_1| \geq \theta_j, \\ \lambda_j \xi_2^2, & \text{if } |\xi_2| < \theta_j, \\ \alpha_j + \varepsilon_j, & \text{if } |\xi_2| \geq \theta_j. \end{cases} \quad (9)$$

In (9), λ_j is a positive weight, the so-called regularization parameter, quantity $\theta_j = \sqrt{\alpha_j/\lambda_j}$ has the meaning of a

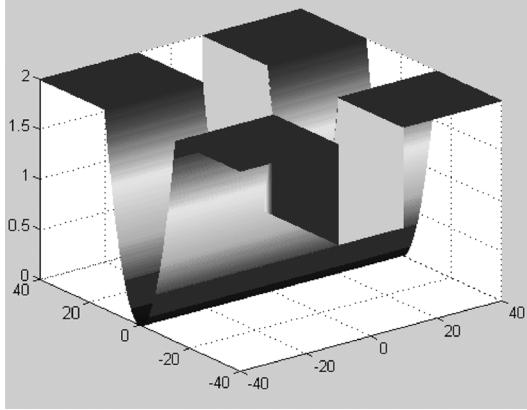


Fig. 3. Shape of stabilizer of (9).

threshold on the intensity gradient above which a discontinuity is expected, whereas $\bar{\theta}_j = \sqrt{(\alpha_j + \varepsilon_j)/\lambda_j}$ is a *suprathreshold*, higher than the threshold, to lower the expectation of an edge when a parallel, close edge is likely to be present. Fig. 3 shows the shape of stabilizer of (9).

As already mentioned, edge regularity constraints are especially useful when, as in this case, deblurring must be achieved. Penalizing close parallel edges, in particular, contrasts the slopes that blur creates in correspondence of jumps in the intensity of the ideal image. Another useful constraint is edge continuation, which means that the image edges, usually corresponding to object boundaries, must form connected, closed lines. This constraint helps in smoothing out the peaks of noise, which would give rise to isolated edges. Our present model is naturally suited to be augmented with the edge continuation constraint, by simply adding to prior energy of (8) a new stabilizer that favors the creation of an edge when a contiguous, aligned edge is likely to be present. This stabilizer should have the same form of (9), except that it should refer to couples of aligned adjacent locations, and the suprathreshold should be lower than the threshold [38].

As per the priors $P(\mathbf{A})$ and $P(\mathbf{H})$, we do not provide here any explicit formula, mainly because the information on these variables can be different from a document to another. On the other hand, this information regards constraints like positivity and ranges of values allowed, so that it cannot be expressed through an analytical prior distribution. This latter is also a reason why we adopted SA to estimate \mathbf{A} and \mathbf{H} .

B. Algorithm for MAP Estimation

Solving the problem of (4) and (5), in view of data model (1), likelihood (6) and priors (7)–(9), can be equivalently reformulated as the minimization of the negative logarithm of the posterior distribution, that is, the posterior energy. This, however, presents some computational difficulties. Indeed, in general, it is not possible to derive analytical formulas for the sources viewed as functions of \mathbf{A} and \mathbf{H} , and the posterior energy is not convex. Thus, a simulated annealing algorithm has to be adopted for the updates of \mathbf{A} and \mathbf{H} , and for each proposal of a new status for each of them, the sources must be computed through numerical estimation. If a correct SA schedule was used and for each proposal of a new status for \mathbf{A} and \mathbf{H} the corresponding

sources were estimated, this scheme would ensure convergence to the global optimum. Nevertheless, its computational complexity would be prohibitive.

To reduce computational costs, while keeping as much as possible the effectiveness of the approach, we acted as follows. First, we chose to update all the \mathbf{A} and \mathbf{H} coefficients at a time. This would be still legitimate, provided that the new status for \mathbf{A} and \mathbf{H} is randomly chosen in a suitably small neighborhood of the current estimate. At each temperature, we let the length of the Markov chains be equal to 20. The initial temperature T_0 was chosen according to the criterion provided in [42], and given by the following formula:

$$T_0 = \frac{\Delta E^+}{\log \left(\frac{n_2}{n_2\chi - n_1(1-\chi)} \right)} \quad (10)$$

where $\chi = 0.5$, n_1 and n_2 , $n_1 + n_2 = 20$, are the number of times that the energy decreases and increases, respectively, and ΔE^+ is the average increase of the energy, when we run the algorithm at any prefixed temperature. According to [39], the law for lowering the temperature has been chosen as follows:

$$\begin{aligned} T^1 &= T_0 \\ T^k &= \mu T^{k-1} \quad k = 2, 3, \dots \end{aligned} \quad (11)$$

with $\mu = 0.95$, and where k indicates here the number of the iteration.

The sources \mathbf{s} are correctly re-estimated, according to (5), for each proposal of a new status for \mathbf{A} and \mathbf{H} . A Metropolis algorithm is then employed to decide about acceptance/rejection of the new configuration for variables $(\mathbf{s}(\mathbf{A}, \mathbf{H}), \mathbf{A}, \mathbf{H})$. Note that, due to the usual small number of mixing and blur coefficients, even a theoretical SA is not particularly cumbersome. The most expensive part of the whole blind separation algorithm is the computation of $\mathbf{s}(\mathbf{A}, \mathbf{H})$, as it entails minimization of the nonconvex posterior energy with respect to the $M \times T$ continuous-valued variables \mathbf{s} . To this end, thanks to the specific image model adopted, we can exploit efficient deterministic nonconvex optimization algorithms, such as the GNC algorithm [37], properly adapted to treat the stabilizer of (9) [43].

The GNC algorithm is based on the minimization, in sequence, of a set of approximations F^p , $p = 2, \dots, 0$, of the original energy function F , where the first approximation F^2 is a convex function and the last approximation F^0 coincides with the original one. In practice, being the likelihood a quadratic function, the GNC approximations only apply to the stabilizers [Fig. 4(a) and (b)]. Although efficient, GNC is still expensive in our case, due to the high number of variables, and the large number of times it has to be executed. Indeed, each GNC minimizes, via gradient descent, 20 approximations of the energy function, using, as starting point, the minimum of the previous approximation. To prevent premature freezing of the solution, in correspondence of the current values of \mathbf{A} and \mathbf{H} , each GNC is initialized with the data.

So far, no strong approximations to the theoretical schedules of both SA and GNC have been adopted. A drastic reduction of the computational costs can, however, be obtained by terminating SA when the reconstructions stabilize, and using only ten steps of gradient descent at each approximation of the

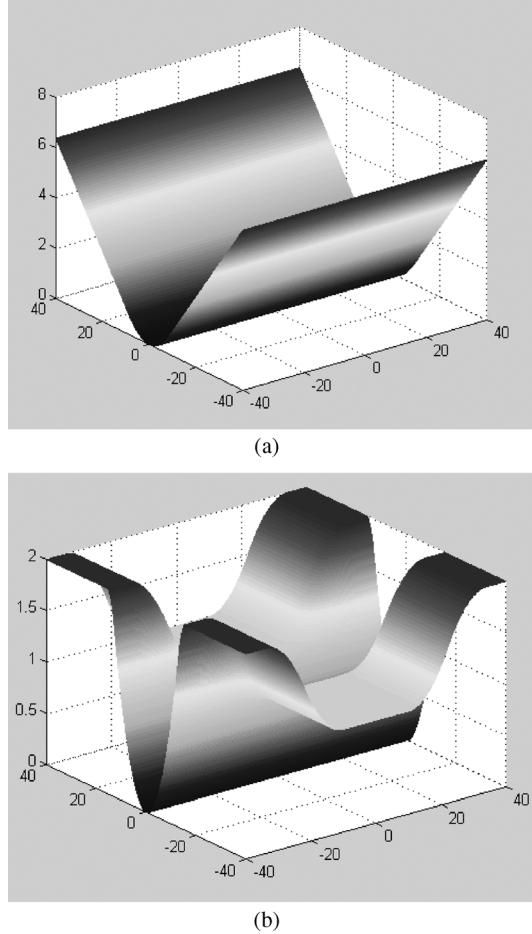


Fig. 4. Two approximations of stabilizer in Fig. 3: (a) first convex approximation ($p = 2$); (b) approximation for $p = 1$.

energy function. Although greatly reduced, the costs are still high, since, at each temperature, a number of GNCs equal to the Markov chain length must be executed.

IV. DOCUMENT ANALYSIS AND RESTORATION AS A CONVOLUTIVE BSS PROBLEM

As stated since the Introduction, in our approach the restoration of degraded documents is formulated as the problem of recovering and enhancing all the patterns overlapped in the document, by exploiting multiple observations, modelled as convolutive mixtures of the patterns themselves. In this way, we are able to formulate several, different document analysis problems as various instances of the same BSS problem described by (1). This unified formulation, although oversimplified, is versatile and flexible to treat many real scenarios, as it will be shown in the section devoted to the experimental results.

A. Interchannel Model

In the multispectral/hyperspectral single-side scenario, we assume that the available multiple scans (or observation channels, or views) of a document have a vector value $\mathbf{x}(t)$ of N components (of index t in a total of T). Similarly, we assume to have $M \leq N$ superimposed sources represented, at each pixel t , by the vector $\mathbf{s}(t)$. Since we consider images of documents containing homogeneous texts or drawings, we can also reasonably

assume that the color of each source, in its pristine state, i.e., undegraded, is almost uniform, and call a_{ij} the mean reflectance index for the j th source at the i th view. Thus, the source functions $s_j(t)$, $j = 1, 2, \dots, M$ denote the “quantity” of the M patterns that concur to form the color at point t . We assume that each source pattern is affected by its own blur operator, which can be different from channel to channel. In summary, our data model is in perfect accordance with the general data model for convolutive BSS as formulated in (1). We call this model *the interchannel model*.

B. Intrachannel Model

In a second scenario, the available document observations amount to two sets of N multispectral scans of the recto and verso side of a document page. To build a model for our data, we only consider two distinct patterns, one in the clean recto-side and one in the clean verso-side of our page, which appear as overlapped in the N scans. The recovery of these two patterns, i.e., the clean recto and verso images, can then be formulated as the solution of a set of $N, 2 \times 2$ convolutive BSS problems, one for each channel, resulting in the following instance of the model of (1):

$$\begin{aligned} x_r^k(t) &= a_{11}^k v_{11}^k(t) + a_{12}^k v_{12}^k(t) + n_1^k(t) \\ x_v^k(t) &= a_{21}^k v_{21}^k(t) + a_{22}^k v_{22}^k(t) + n_2^k(t) \\ t &= 1, 2, \dots, T, \quad k = 1, 2, \dots, N \\ \mathbf{v}_{ij}^k &= \mathbf{H}_{ij}^k \mathbf{s}_j^k, \quad i = 1, 2, \quad j = 1, 2, \quad k = 1, 2, \dots, N. \end{aligned} \quad (12)$$

In each problem above, $x_r^k(t)$ and $x_v^k(t)$ are the recto and the horizontally flipped verso appearances, respectively, at the k th channel, $s_1^k(t)$ and $s_2^k(t)$ are the reflectance maps at the k th channel associated to the recto and verso pattern, respectively, t is a pixel index, and a_{ij}^k are unknown mixing coefficients. Physically, coefficients a_{ij}^k represent the percentage of ink intensity attenuation of the two patterns in the two sides, due to the transparency of the paper, or ink seeping from the back to the front page. \mathbf{H}_{ij}^k are the block-Toeplitz matrices associated to the filters \mathbf{B}_{ij}^k that model the blur effect.

Note that here the mutual independence of the overall set of sources cannot be assumed, since the different colors of a same pattern are certainly highly correlated. On the other hand, the N problems are independent of each other, that is the overall $2N \times 2N$ system of equations is separable. We call this model *the intrachannel model*.

We observed that the pattern interfering from the back side is always blurred, whereas the same pattern is not, or much less, in the side where it was originally written. Hence, the blur operators are expected to be different, for each source, in the two observations, and for the two sources in a same observation. This, on one hand, means that the inclusion of blurs in the model is necessary to permit the two versions of a same pattern to match from each other. On the other hand, it allows to simplify the problem, by reducing its dimension. Let us consider that, at least in an idealized setting, the two sides have been written with the same ink, same pressure and at two close moments. Then, it is reasonable to assume that, at each channel, the attenuation/smearing of the bleed-through/show-through pattern in the two sides is the same, i.e., $a_{12}^k = a_{21}^k$ and $\mathbf{H}_{12}^k = \mathbf{H}_{21}^k$. For

TABLE I
SUMMARY OF PREFIXED AND ESTIMATED PARAMETERS

	Fig. 6	Fig. 10	Fig. 11	Fig. 12
λ_1	10	1	1	1
α_1	10	2000	1000	8
ε_1	10	2000	1000	8
λ_2	1	1	1	1
α_2	1000	2000	1000	8
ε_2	1000	2000	1000	8
$\hat{\mathbf{A}}$	$\begin{bmatrix} 1 & 0 \\ 0.51 & 0.50 \end{bmatrix}$	$\begin{bmatrix} 0.78 & 0.23 \\ 0.23 & 0.78 \end{bmatrix}$	$\begin{bmatrix} 0.81 & 0.25 \\ 0.25 & 0.81 \end{bmatrix}$	$\begin{bmatrix} 0.81 & 0.17 \\ 0.17 & 0.81 \end{bmatrix}$
$\hat{\mathbf{B}}_{11}$	$\begin{bmatrix} 0.05 & 0.31 & 0.04 \\ 0.06 & 0.08 & 0.26 \\ 0.15 & 0.01 & 0.04 \end{bmatrix}$	$\begin{bmatrix} 0.07 & 0.09 & 0.07 \\ 0.09 & 0.36 & 0.09 \\ 0.07 & 0.09 & 0.07 \end{bmatrix}$	$\mathbf{I}_{3 \times 3}$	$\mathbf{I}_{3 \times 3}$
$\hat{\mathbf{B}}_{12}$	$\mathbf{I}_{3 \times 3}$	$\begin{bmatrix} 0.09 & 0.12 & 0.09 \\ 0.12 & 0.17 & 0.12 \\ 0.09 & 0.12 & 0.09 \end{bmatrix}$	$\begin{bmatrix} 0.02 & 0.15 & 0.02 \\ 0.15 & 0.34 & 0.15 \\ 0.02 & 0.15 & 0.02 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.16 & 0.00 \\ 0.16 & 0.36 & 0.16 \\ 0.00 & 0.16 & 0.00 \end{bmatrix}$
$\hat{\mathbf{B}}_{21}$	$\begin{bmatrix} 0.04 & 0.01 & 0.02 \\ 0.04 & 0.79 & 0.07 \\ 0.02 & 0.03 & 0.09 \end{bmatrix}$	$\begin{bmatrix} 0.09 & 0.12 & 0.09 \\ 0.12 & 0.17 & 0.12 \\ 0.09 & 0.12 & 0.09 \end{bmatrix}$	$\begin{bmatrix} 0.02 & 0.15 & 0.02 \\ 0.15 & 0.34 & 0.15 \\ 0.02 & 0.15 & 0.02 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.16 & 0.00 \\ 0.16 & 0.36 & 0.16 \\ 0.00 & 0.16 & 0.00 \end{bmatrix}$
$\hat{\mathbf{B}}_{22}$	$\begin{bmatrix} 0.23 & 0.01 & 0.03 \\ 0.03 & 0.27 & 0.39 \\ 0.00 & 0.03 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0.07 & 0.09 & 0.07 \\ 0.09 & 0.36 & 0.09 \\ 0.07 & 0.09 & 0.07 \end{bmatrix}$	$\mathbf{I}_{3 \times 3}$	$\mathbf{I}_{3 \times 3}$

similar considerations, it is also expected that $a_{11}^k = a_{22}^k$ and $\mathbf{H}_{11}^k = \mathbf{H}_{22}^k$. Furthermore, as already mentioned, the blur of the bleed-through pattern can be reasonably assumed stronger than that affecting the main text, in both sides. Similarly, the ink intensity of the front text pattern in the recto side should be higher than that of the bleed-through pattern, i.e., $a_{11}^k > a_{12}^k$, with the same relationship holding, reversed, in the verso side.

These properties and symmetry features of the problem can be easily exploited in a direct manner. Thus, the number of kernels and mixing coefficients to be estimated can be reduced to two for each channel, and the conditions on the kernel elements and mixing coefficients can be enforced as constraints, described through the priors $P(\mathbf{H})$ and $P(\mathbf{A})$, in the SA estimation algorithm. In addition, these constraints on the mixing operator prevent the permutation indeterminacy otherwise inherent in BSS methods.

V. DISCUSSION OF THE EXPERIMENTAL RESULTS

In this section, several applications of the method to the analysis and restoration of real document images will be shown, and the results will be discussed in detail. We will consider different instances of the data model, according to the discussion in the previous section, and in dependence of the available data set, the kind of degradation, and the purpose of the analysis. We will first consider the case where the data are constituted of multispectral views of a single-sided document page. The aim is to extract faint or hidden patterns, recover the clean main text when even significant interferences affect the background, or, more in general, extract document features. In this case, the reference model is the general interchannel model of (1). We will then

consider the case where the data are given by the two grayscale scans of the recto and verso side of a document page [model of (12), $N = 1$], with the aim at recovering the clean main texts in the two sides. In some of the experiments performed, we will also provide the results obtained using the FastICA algorithm [44], and discuss its performance with respect to that of the proposed method. Still for comparison purposes, results of a diffusion-based restoration method on some of the examples considered here can also be found in [6].

In all cases, we assumed that the recto-verso pairs or multispectral views are preregistered. Furthermore, when it applies, we assumed to be able to estimate the kernel sizes off-line, for instance by looking at the slope extent of the character boundaries in the various superimposed patterns. With respect to the MRF hyperparameters, these were found by trial-and-error. However, general guidelines, based on the hyperparameter physical meaning explained in Section III, can be adopted. In particular, the threshold θ is related to the minimum amplitude of the discontinuities that we want to be preserved in the image, whereas the smoothness parameter λ should be higher for higher noise levels and for larger uniform regions in the image. Once λ and θ are determined, also α is automatically defined. As per the suprathreshold $\bar{\theta}$, governed by hyperparameter ε , we have found that, in general, significant improvement of the results can be obtained by its inclusion in the model, but its value is not very critical, so that we always set $\varepsilon = \alpha$. For the values of the prefixed parameters and the estimated parameters we refer to Table I. Since we experimented with real document images, our evaluation of the estimates, as well as the appropriateness of the chosen hyperparameters, including blur mask size, can only be qualitative.

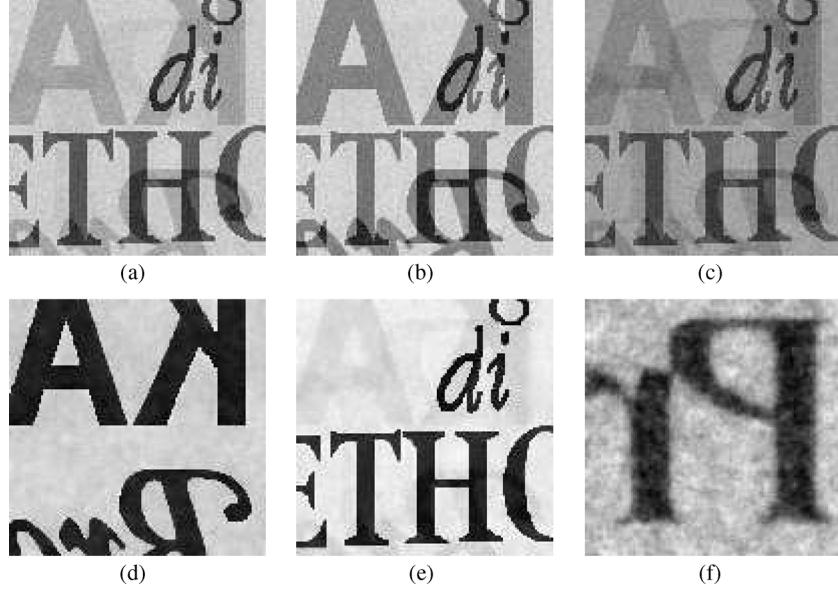


Fig. 5. Separation of text patterns overlapped in a synthetic RGB document: (a) the red channel; (b) the green channel; (c) the blue channel; (d) first separated text; (e) second separated text; (f) third separated text.

A. Interchannel Case

A first numerical experiment is provided to illustrate the method, suggest one of its applications, and quantitatively evaluate its performance. In the example of Fig. 5, we show the RGB components of a single-sided document containing three superposed texts. Although generated synthetically according to (1), these data are representative of a real scenario. In this case, we did not apply any blur to the original source images, that were only overlapped according to the coefficients given by matrix \mathbf{A} in (13) (each row corresponds to a color channel), plus addition of Gaussian noise of variance $\sigma^2 = 49$, $SNR = 16.93$. Although the model is further simplified, the problem here is very challenging, in that the third text pattern is very faint in each color channel (third column), in such a way that it is not perceptible almost at all in the resulting color image.

Hence, the purpose of the analysis of this document is here mainly the recovery of the hidden pattern. Since the two perceptible source patterns in the three observations exhibit similar characteristics, in terms of contrast between text and background and thickness of the characters, we adopted the same hyperparameters $\lambda = 0.5$, $\alpha = \varepsilon = 500$ for both. From inspection of the data alone, it is not possible to foresee the structure of the undetectable third pattern; hence, we were forced to adopt the same hyperparameters above for this pattern as well. The recovered images are shown in Fig. 5(d), (e), and (f), respectively, and the estimated mixing matrix $\hat{\mathbf{A}}$ in (13)

$$\mathbf{A} = \begin{bmatrix} 0.6 & 0.2 & 0.1 \\ 0.5 & 0.4 & 0.05 \\ 0.4 & 0.15 & 0.2 \end{bmatrix}, \quad \hat{\mathbf{A}} = \begin{bmatrix} 0.60 & 0.19 & 0.14 \\ 0.51 & 0.49 & 0.02 \\ 0.40 & 0.15 & 0.39 \end{bmatrix}. \quad (13)$$

As it can be noted, despite the unperfect estimation of the third column of the mixing matrix, the faint third pattern is recovered surprisingly well. However, this appears very noisy. The reason for that lies in the fact that the mixing matrix, having

a quasi-zero column, is ill-conditioned. While for the other two components the information available in the three views allows for overcoming this inconvenience, the information available on the third component is too little, so that noise dominates. A possible remedy could be to increase the smoothness parameter λ associated to this component. However, forcing a stronger smoothness to remove noise could even prevent the detection of the pattern itself.

In a second experiment, shown in Fig. 6, we considered the RGB image of a real-fake palimpsest, that is an artificial palimpsest, generated by hand and then scanned. In particular, this one simulates the colors and the mutual position of the two overlapped texts as they appear in the famous Archimedes Palimpsests [45]. Here, the purpose of the analysis is the recovery of the underwriting, which simulates an older text erased and then overwritten. By inspection of the red, green and blue channels, we observed that in the red and green channels the overwritten text almost disappears, whereas in the blue channel overwriting and underwriting have similar intensities. These observations gave us the possibility to take *a priori* assumptions on the blur masks and the mixing matrix. First of all, we chose to process only the green and blue channels. The reason for this choice is twofold: (i) since our goal is to separate the two only patterns overlapped in the document, two channels are sufficient, provided that the two patterns exhibit a spectral diversity; (ii) the red channel does not carry additional information. Thus, we can reasonably consider that the second source (underwriting) does not contribute to the green channel, and then we can enforce the first row of the mixing matrix to $[1 \ 0]$, and disregard parameter \mathbf{B}_{12} . This helps to reduce the problem dimension. Although feasible, we did not force the second row of the mixing matrix (corresponding to the blue channel) to have equal coefficients, but rather verified this experimentally, and foresaw some blur on the first source (overwriting) as well.



Fig. 6. Separation of overwriting and underwriting from an RGB real-fake palimpsest (from: “Text recovery from the Archimedes Palimpsest”, <http://www.cis.rit.edu/people/faculty/easton/k-12/exercise/index.htm>): (a) the red channel; (b) the green channel; (c) the blue channel; (d) first separated text; (e) second separated text; (f) first ICA output; (g) second ICA output.

The results obtained were the separated text patterns of Fig. 6(d) and (e), and the estimated blur masks and mixing matrix reported in Table I. Note that the mixing matrix correctly exhibits almost equal coefficients in the second row. Note also that $\hat{\mathbf{B}}_{11}$ is the strongest among the three blur masks. This gives reason of the faint appearance of the first pattern in the data [Fig. 6(b)], and helps to refocus and increase the contrast in the recovered pattern [Fig. 6(d)], despite having fixed to 1 the related mixing coefficient. Among the two remaining masks, the stronger is $\hat{\mathbf{B}}_{22}$, which explains the more blurred appearance of the second pattern in the blue channel. Observe also that the estimated blur masks are asymmetric. This may be due to some polarization during the writing of the texts (e.g., higher pressure in some direction), or to different spreading characteristics of the pigments in the ink. Another reason could be that, being likely the degradation nonstationary, the algorithm looks for global blur masks that are an average of the many local blur masks.

As it is apparent, the recovered underwritten pattern presents gaps in the areas where the overwritten text occludes it. This is one of the main disadvantages of a linear data model, which is unable to describe the color saturation effect. To properly manage occlusions, a nonlinear model should be devised, for instance by accounting for some known nonlinear function within a convolutive model, as done in [32]. In the pure linear assumption, the gaps can be avoided only when the pattern to be recovered is dominant over the others at least in one of the available channels, which is not the case with this example.

Fig. 6(f) and (g) shows the outputs of the FastICA algorithm, which provides a satisfactory separation of the two texts, although no full separation was achieved. The estimated mixing matrix was

$$\hat{\mathbf{A}}_{ICA} = \begin{bmatrix} 0.61 & 0.06 \\ 0.80 & 1.00 \end{bmatrix}.$$

We should point out that, when applying an ICA algorithm for instantaneous mixtures, we are implicitly assuming that each source is affected by the same blur (eventually an impulse, i.e., no blur) in each channel. We believe that this assumption is not verified in this case, since it is apparent that, e.g., the first pattern is more blurred in the blue channel than in the green channel. As a result, besides blur and amplified noise, some traces of the interfering patterns are still visible in the separations, and a subsequent deblurring is not sufficient to remove them. In other words, the imperfect separation achieved by ICA is just due to the fact that individual blurs on the patterns are not accounted for during separation, so that a pattern in an observation channel cannot match with the same pattern in the other observation channel. This effect is somewhat similar to an imperfect registration of the two channel maps prior separation. Nevertheless, the mixing coefficients estimated by ICA are reasonable, and give reason of the different intensity of the two patterns in the two views, thus making possible an increase of the contrast. We obtained the same effect by assuming a nonimpulsive kernel \mathbf{B}_{11} on the first source as well, while keeping to 1 the corresponding mixing coefficients.

For completing the experiment, we run a nonblind deconvolution on the second FastICA output, by using the blur mask $\hat{\mathbf{B}}_{22}$ previously estimated. Results corresponding to different values of the hyperparameters are shown in Fig. 7. In all cases, we observed that a good deblurring has been achieved with the blur mask estimated with our method. However, to obtain the removal of the residual interference, we had to set the regularization parameter λ to very high values (in Fig. 7 the value of λ increases from left to right), with the side adverse effect to attenuate the character contrast.

This experiment also gives us the opportunity to further comment on the hyperparameter meaning. The first result of Fig. 7(a) has been obtained by employing $\lambda = 0.5$ and $\alpha = \varepsilon = 100$, which gives a low threshold $\theta = 20$. In this

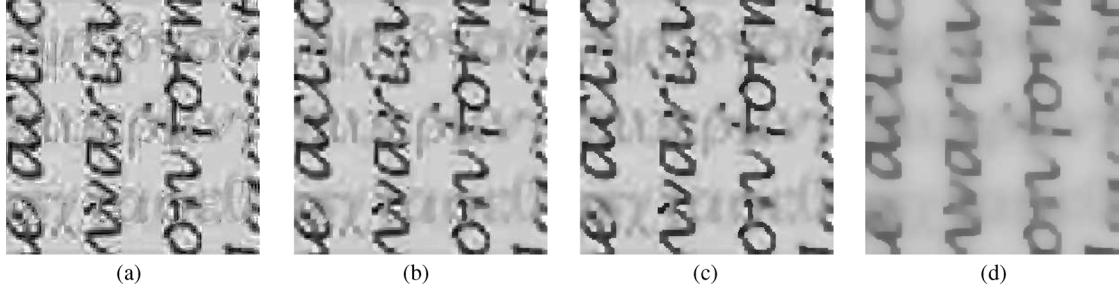


Fig. 7. Nonblind deconvolution of the second ICA output of Fig. 6(g), for different values of the hyperparameters.

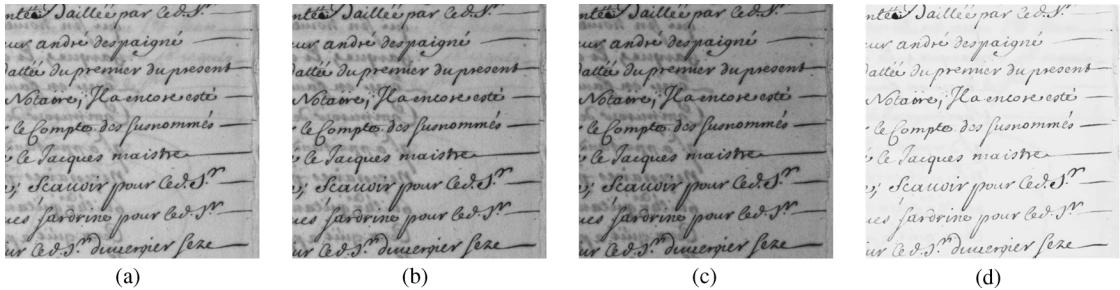


Fig. 8. Recovery of the clean main text from a degraded RGB manuscript [from the acquisition campaign of Project Isyreadet (<http://www.isyreadet.net>)]: (a) the red channel; (b) the green channel; (c) the blue channel; (d) the restored manuscript.

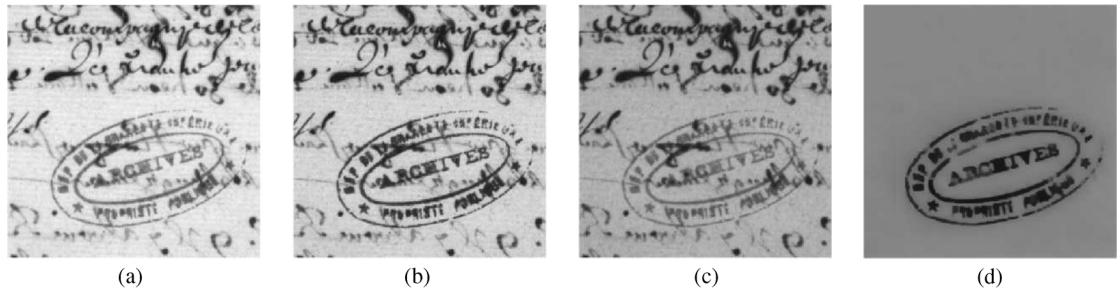


Fig. 9. Application to the segmentation of the document contents [from the acquisition campaign of Project Isyreadet (<http://www.isyreadet.net>)]: extraction of a stamp: (a) the red channel; (b) the green channel; (c) the blue channel; (d) the stamp.

conditions, most of the image discontinuities are preserved, including those corresponding to the unwanted interferences. In the second test, we thus kept fixed the value of λ to 0.5 and increased the threshold to $\theta = 63$, by employing a higher $\alpha = 1000$ [Fig. 7(b)]. We can notice that the interferences, although reduced, are still present, whereas the deblurring effect begins to diminish. Indeed, a too high threshold does not allow to locate edges in the slopes that blur causes in the image. We then attempted to smooth out the residuals by increasing the regularization parameter λ , for an intermediate threshold value $\theta = 31$. The results corresponding to $\lambda = 2$ and $\lambda = 10$ are shown in Fig. 7(c) and (d), respectively. It is apparent that, for this image, no pairs of λ and α exist that are able to cope with the opposite needs to smooth out the interferences and refocus the main text.

In the third experiment, we considered a real RGB manuscript affected by a significant bleed-through distortion. As it is possible to notice, the bleed-through pattern is strong and very blurred in all the three available channels [see Fig. 8(a)–(c)], whereas the main text is quite well contrasted. The aim of our analysis is here the recovery of the clean foreground text. Since we are not interested in the recovery of the bleed-through pat-

tern, at least in principle we could disregard the blur kernels from the model. Nevertheless, since our approach is based on the looking for a total match between the various patterns in the different views, we must still include and estimate blur operators, to cope with possible different extents of blur on the bleed-through pattern, in the different channels. Choosing two among the three channels as data set, we recovered the restored manuscript shown in Fig. 8(d).

The goal of the successive experiment was the extraction of a stamp from an RGB manuscript, shown in Fig. 9. This is an example of the possible application of the method to the segmentation of the document contents or the extraction of *beyond text* features for indexation purposes. Since there are three patterns overlapped in the document, at least in principle, we should exploit all the three red, green and blue channels. When, as in this case, the interest is in the recovery of only one of the three patterns, two views can be sufficient. We then processed the green and blue channels, and obtained the result shown in Fig. 9(d).

B. Intrachannel Case

In the next experiments, we considered the intrachannel data model of (12), with $N = 1$, with the aim at recovering clean

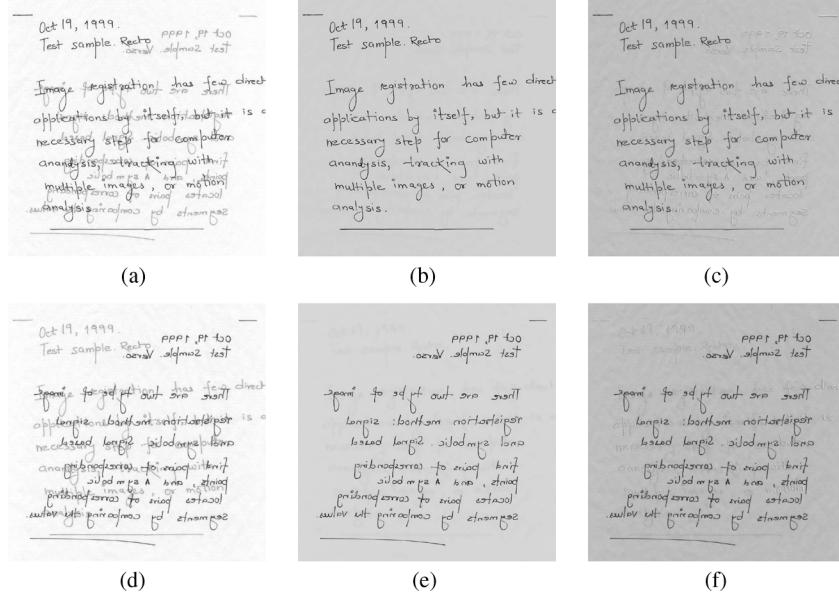


Fig. 10. Restoration of the grayscale recto and verso sides of a real document (from: <http://www.site.uottawa.ca/~edubois/documents>): (a) recto side; (b) restored recto side; (c) first ICA output; (d) horizontally flipped verso side; (e) horizontally flipped restored verso side; (f) second ICA output.

main texts in both the front and reverse side of grayscale documents affected by bleed-through/show-through. We assumed the reliability of the symmetry conditions discussed in the previous section for this scenario. Thus, we estimated both a symmetric mixing matrix and two blur masks, which symmetrically affect the two main texts in the two sides. For the manuscript in Fig. 10 [46], it is apparent that the above symmetries can be extended to the MRF hyperparameters as well, in that character font and ink intensities are equal in the two sides. Finally, we also enforced a reasonable circular symmetry for the blur masks, to reduce the number of parameters. We obtained the mixing matrix and blur kernels reported in Table I, and the clean recto and verso maps shown in Fig. 10(b) and (e). Note also that, since we allow for the coefficients a_{11} and a_{22} to be less than 1, this results in a darkening of the background, but of the text as well, which is desirable, in general. Fig. 10(c) and (f) shows the quite satisfactory results of the FastICA algorithm, although a full separation is not achieved. Once again, we believe that this is due to the fact that no individual blur for each source in each channel was accounted for, so that methods based on pixel-by-pixel matching cannot perform well.

In another example, we considered a recto-verso pair drawn from the Google Book Search dataset [47] [Fig. 11(a) and (d)]. In this case, we observed a good resolution of the main text in both sides, so that we fixed $\mathbf{B}_{11} = \mathbf{B}_{22} = \mathbf{I}$. On the other hand, our edge-preserving image model is inherently able to remove little amounts of blur even without explicitly including a blur kernel in the data model. We obtained the restored images of Fig. 11(b) and (e). Observe that the recto side of the document contains not only the main text and the show-through pattern, but also some handwriting and some weak extra texts. Similarly, the verso side contains some stains. These extra patterns are not fully removed in the final, restored images. This, from our point of view, is a confirmation of the good performance of our method, since, from the adopted model, with two channels available, we can only separate two superposed patterns. It

is worth saying again that in our approach the goal of restoration is not the indiscriminate removal of the structured background, but, rather, the artifacts are removed in a selective way, depending on their origin. For instance, in the present case the aim is to remove all the strokes coming from the reverse side, whereas all other patterns belonging to the front side can and should be preserved. This, while removing the spurious information and then improving readability, permits to maintain the original appearance of an old manuscript or document, which often contains the imprints of its history and authenticity. For comparison, Fig. 11(c) and (f) shows the result of the application of FastICA, which exhibits performance similar to that in the case of Fig. 10.

Fig. 12(a) and (c) shows unmatched details of the recto and verso sides of a very complex real manuscript affected by show-through [46]. The complexity of this document image lies in the fact that the original recto and verso scans were in the jpeg format and that the text characters are very thin. Thus, on one hand, the image model adopted should be able to “regularize” enough to remove the artifacts introduced by the jpeg compression, even more amplified by the matrix inversion inherent with the linear mixing data model considered. On the other hand, it should preserve the fine details in the image. A good compromise has been found using $\lambda = 1$ and $\alpha = 8$, which make a very low threshold for capturing the fine details, but a relatively high smoothness degree. Here the value of $a_{12} = a_{21}$ resulted smaller than those estimated in the other experiments, which is in accordance with the lower intensity of the interfering pattern. The restored images are shown in Fig. 12(b) and (d). Note the preservation of the paper texture and folding.

A last, experiment for the recto-verso case is shown in Fig. 13. Here the document is a printed text affected by a significant show-through distortion [Fig. 13(a) and (d)]. In particular, the show-through pattern partially overlaps with the main foreground text, in both sides. With our approach, we obtained the restored images of Fig. 13(b) and (e). It is apparent that in the

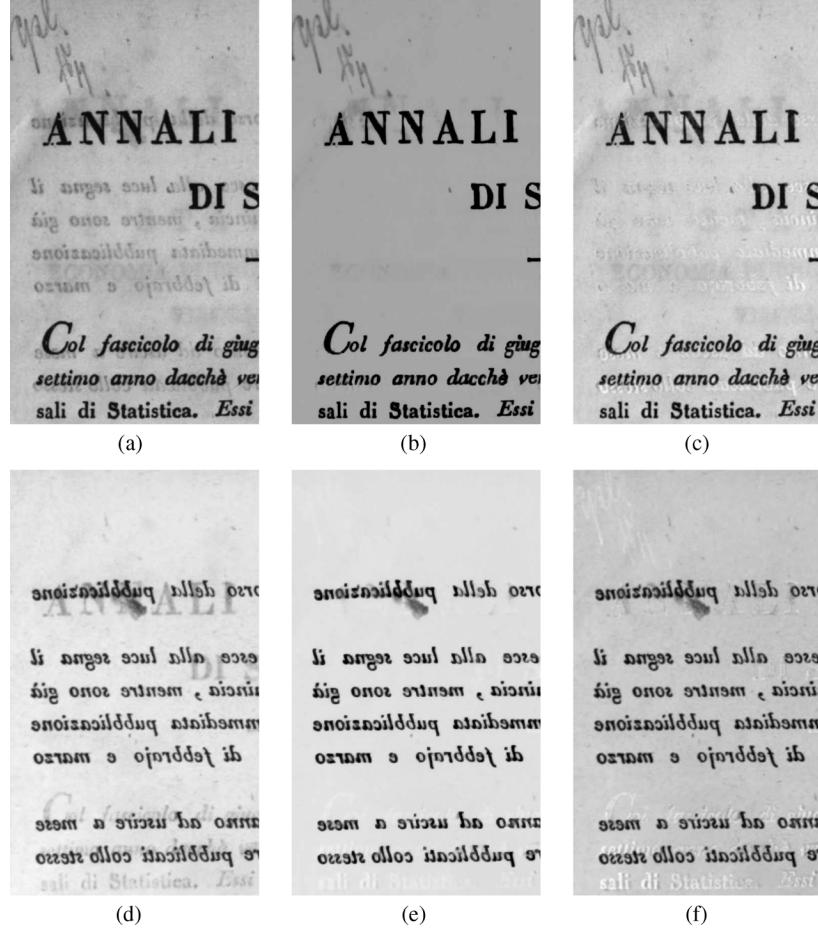


Fig. 11. Restoration of the grayscale recto and verso sides of a real document (from the Google Book Search dataset): (a) recto side; (b) restored recto side; (c) first ICA output; (d) horizontally flipped verso side; (e) horizontally flipped restored verso side; (f) second ICA output.

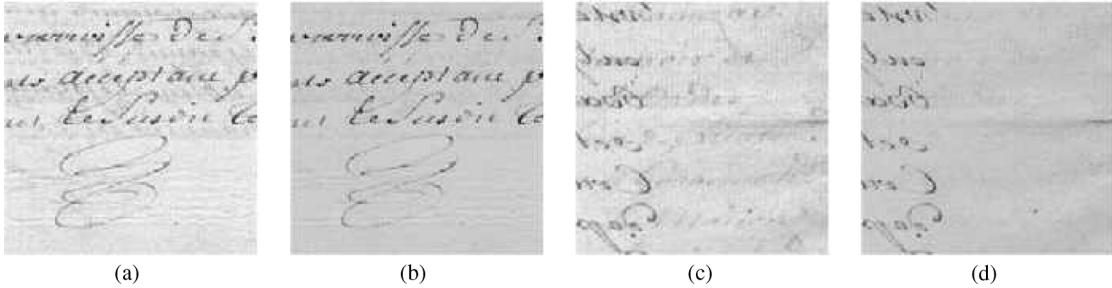


Fig. 12. Unmatched details of the restored recto and verso sides of a real manuscript (from <http://www.site.uottawa.ca/~edubois/documents>): (a) recto side; (b) restored recto side; (c) horizontally flipped verso side; (d) horizontally flipped restored verso side.

areas where the show-through pattern overlaps the main text, the ink intensity is lower than that in the data. On the other hand, the show-through pattern is almost perfectly cancelled from the background area. As already highlighted, this effect is due to the stationary linear model adopted, which, in particular, cannot describe the color saturation effect. Note that the same effect is less evident in the previous experiments, where, on the contrary, some faint traces of the verso strokes remain in the background as well. We deem this is due to an excellent estimation of the average mixing matrix, which can rely on a very large show-through pattern. A nonstationary model would have probably resulted with the same, correct mixing matrix in the background area, and a zero mixing matrix in the occlusion areas.

However, the data and the image model are able to prevent full gaps in the characters.

VI. CONCLUSION

We proposed a Bayesian formulation for the joint blind source separation and restoration of noisy mixtures of degraded images. We considered edge-preserving MRF image models to describe local spatial autocorrelation, and reformulated the original MAP estimation of all the variables as the MAP estimation of the mixing alone, while the sources are kept clamped to their MAP estimate, for any status of the mixing. This scheme is implemented in the practice by a SA algorithm where updates of the mixing operator alternates with MAP estimations of

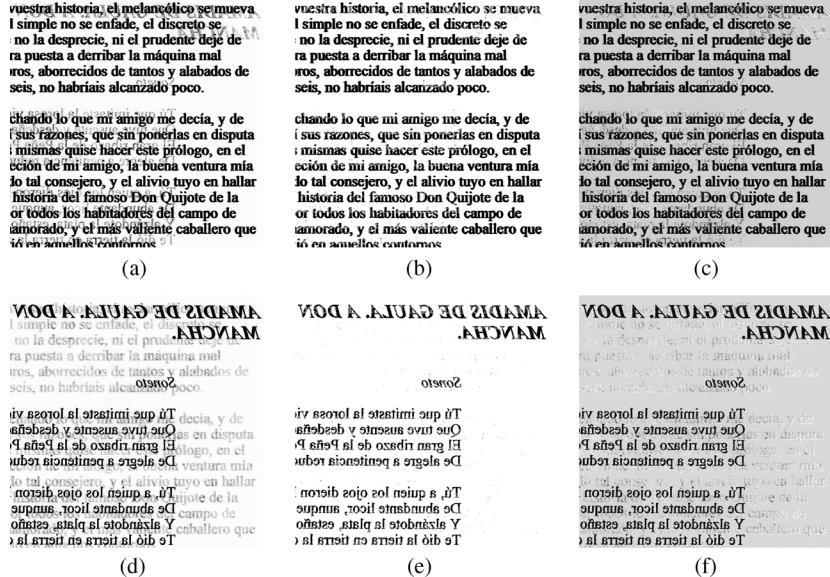


Fig. 13. Restoration of the grayscale recto and verso sides of a real printed document (original scans by the authors): (a) recto side; (b) restored recto side; (c) first ICA output; (d) horizontally flipped verso side; (e) horizontally flipped restored verso side; (f) second ICA output.

the sources, performed through GNC-like gradient ascent algorithms. We presented results of the application of this technique to several instances of analysis and restoration of degraded documents, characterized by the overlapping of two or more texts, or artifacts such as bleed-through or show-through. Our method involves a linear data model, where multimodal observations of the object are seen as mixtures of all the patterns to be extracted, but we introduced a blur model to account for the typical degradations due to unpredictable phenomena of smearing, diffusion and fading of the ink, and to the optical blur and noise. This is necessary to make the various patterns to match in the data, and, at the same time, permits the refocusing of the texts extracted. Although blur identification has been largely investigated for the case of the blind restoration of images, the problem is extremely more complicated in the present case where the degraded images are also mixed. This also in consideration of the fact that the blur can differ from source to source, and from mixture to mixture. The possibility of separating and removing interfering patterns in documents is crucial for improving readability, both by a human reader and by an OCR system. As discussed, the method could also be used for the enhancement of traces of erased texts, as in ancient palimpsests, and is flexible, in general, for application to a large range of instances of content analysis and restoration of images of ancient documents or other artworks, such as photos, videos and paintings, of high interest for the preservation and access of our cultural heritage.

However, some relevant problems remain to be explored. The most critical one is the devising of a more comprehensive model for the phenomenon of interfering patterns in documents. This should account for nonlinearities, in order to efficiently describe the color saturation effect and the occlusions, and nonstationarity of the degradation operator, which is particularly noticeable in the bleed-through effect. Other problems regard the MRF hyperparameter selection, which is currently performed in a heuristic way, based on the visual inspection of the mixtures. However, our estimation scheme could in principle be added with an extra step where the hyperparameters are jointly estimated. This is especially important to manage

nonstationary images or nonstationary noise showing a large variability, where it is expected that also the hyperparameters should be considered space-varying, and, thus, cumbersome to select by trial-and-error.

REFERENCES

- [1] G. Leedham, S. Varma, A. Patankar, and V. Govindaraju, "Separating text and background in degraded document images—A comparison of global thresholding techniques for multi-stage thresholding," in *Proc. 8th Int. Workshop on Frontiers in Handwriting Recognition*, 2002, pp. 244–249.
- [2] C. L. Tan, R. Cao, and P. Shen, "Restoration of archival documents using a wavelet technique," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 10, pp. 1399–1404, Oct. 2002.
- [3] G. Sharma, "Show-through cancellation in scans of duplex printed documents," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 736–754, May 2001.
- [4] E. Dubois and A. Pathak, "Reduction of bleed-through in scanned manuscript documents," in *Proc. IS&T Image Processing, Image Quality, Image Capture Systems Conf.*, 2001, pp. 177–180.
- [5] P. Dano, "Joint Restoration and Compression of Document Images With Bleed-Through Distortion," Master's thesis, Ottawa-Carleton Institute for Electrical and Computer Engineering, School of Information Technology and Engineering, Univ. Ottawa, Ottawa, ON, Canada, Jun. 2003.
- [6] R. F. Moghaddam and M. Cheriet, "Low quality document image modeling and enhancement," *Int. J. Document Anal. Recognit.*, vol. 11, no. 4, pp. 183–201, Mar. 2009.
- [7] H. Nishida and T. Suzuki, "A multiscale approach to restoring scanned color document images with show-through effects," in *Proc. 7th Int. Conf. Document Analysis and Recognition*, 2003, vol. 1, pp. 584–588.
- [8] Y. Leydier, F. LeBourgeois, and H. Emptoz, "Serialized unsupervised classifier for adaptive color image segmentation: Application to digitized ancient manuscripts," in *Proc. Int. Conf. Pattern Recognit.*, 2004, pp. 494–497.
- [9] F. Drida, F. LeBourgeois, and H. Emptoz, "Restoring ink bleed-through degraded document images using a recursive unsupervised classification technique," in *Proc. 7th Workshop Document Analysis Systems*, 2006, pp. 38–49.
- [10] C. Wolf, Document Ink Bleed-Through Removal With Two Hidden Markov Random Fields and a Single Observation Field, Laboratoire d'Informatique en Images et Systèmes d'Information, INSA de Lyon, Tech. Rep. RR-LIRIS-2006-019, Nov. 2006.
- [11] S. Amari and A. Cichocki, "Adaptive blind signal processing-neural network approaches," *Proc. IEEE*, vol. 86, no. 10, pp. 2026–2048, Oct. 1998.

- [12] 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.
- [13] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
- [14] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2002.
- [15] 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.
- [16] A. Tonazzini, L. Bedini, and E. Salerno, "Independent component analysis for document restoration," *Int. J. Document Anal. Recognit.*, vol. 7, pp. 17–27, 2004.
- [17] A. Tonazzini, E. Salerno, and L. Bedini, "Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique," *Int. J. Document Anal. Recognit.*, vol. 10, pp. 17–25, Jun. 2007.
- [18] W. Zhu, N. Galatsanos, and A. Katsaggelos, "Regularized multichannel restoration using cross-validation," *Graph. Models Image Process.*, vol. 57, no. 1, pp. 38–54, 1995.
- [19] M. Kang, "Generalized multichannel image deconvolution approach and its applications," *Opt. Eng.*, vol. 37, no. 11, pp. 2953–2964, Nov. 1998.
- [20] Y. He, K. H. Yap, L. Chen, and L. P. Chau, "A novel hybrid model framework to blind color image deconvolution," *IEEE Trans. Syst., Man, Cybern. A*, vol. 38, no. 4, pp. 867–880, Apr. 2008.
- [21] M. K. Ng, W. C. Kwanand, and R. Chan, "A fast algorithm for high-resolution color image reconstruction with multisensors," *Lecture Notes Comput. Sci.*, vol. 1988, pp. 615–627, 2001.
- [22] K. Kokkinakis and A. K. Nandi, "Multichannel blind deconvolution for source separation in convolutive mixtures of speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 1, pp. 202–212, Jan. 2006.
- [23] S. Araki, S. Makino, R. Mukai, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," in *Proc. ICA*, 2001, pp. 132–137.
- [24] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 120–134, Jan. 2005.
- [25] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-gaussian processes using higher order statistics and inverse filter criteria," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 658–672, Mar. 1997.
- [26] J. Miskin and D. MacKay, "Ensemble learning for blind image separation and deconvolution," in *Adv. Independent Compon. Anal.*, M. Girolami, Ed. New York: Springer Verlag, 2000.
- [27] C. Simon, P. Loubaton, and C. Jutten, "Separation of a class of convolutive mixtures: A contrast function approach," *Signal Process.*, vol. 81, pp. 883–887, 2001.
- [28] S. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and speech separation using causal fir filters," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 92–104, Jan. 2005.
- [29] S. Shwarts, Y. Schechner, and M. Zibulevsky, "Blind separation of convolutive image mixtures," *Neurocomputing*, vol. 71, no. 10–12, pp. 2164–2179, 2008.
- [30] M. Castella and J.-C. Pesquet, "An iterative blind source separation method for convolutive mixtures of images," *Lecture Notes Comput. Sci.*, vol. 3195, pp. 922–929, 2004.
- [31] E. Be'ery and A. Yeredor, "Blind separation of superimposed shifted images using parameterized joint diagonalization," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 340–353, Mar. 2008.
- [32] B. Ophir and D. Malah, "Show-through cancellation in scanned images using blind source separation techniques," in *Proc. Int. Conf. Image Processing*, 2007, vol. III, pp. 233–236.
- [33] F. Merrikh-Bayat, M. Babaie-Zadeh, and C. Jutten, "A nonlinear blind source separation solution for removing the show-through effect in the scanned documents," presented at the 16th European Signal Processing Conf. (EUSIPCO), Aug. 2008.
- [34] K. Knuth, "Bayesian source separation and localization," in *Proc. SPIE Bayesian Inference for Inverse Problems*, 1998, pp. 147–158.
- [35] A. Tonazzini, L. Bedini, and E. Salerno, "A markov model for blind image separation by a mean-field EM algorithm," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 473–482, Feb. 2006.
- [36] F. Su and A. Mohammad-Djafari, "Bayesian separation of document images with hidden markov models," in *Proc. 2nd Int. Conf. Computer Vision Theory and Applications*, 2007, pp. 151–156.
- [37] A. Blake and A. Zissermann, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
- [38] L. Bedini, I. Gerace, and A. Tonazzini, "A deterministic algorithm for reconstructing images with interacting discontinuities," *CVGIP: Graph. Models Image Process.*, vol. 56, no. 2, pp. 109–123, 1994.
- [39] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 229, no. 4598, pp. 671–680, 1983.
- [40] A. Tonazzini, S. Vezzosi, and L. Bedini, "Analysis and recognition of highly degraded printed characters," *Int. J. Document Anal. Recognit.*, vol. 6, pp. 236–247, 2004.
- [41] A. Tonazzini and I. Gerace, "Bayesian MRF-based blind source separation of convolutive mixtures of images," presented at the 13th European Signal Processing Conf. EUSIPCO, Sep. 2005.
- [42] E. Aarts and J. Korst, *Simulated Annealing and Boltzman Machines*. New York: Wiley, 1989.
- [43] A. Boccuto, M. Discepoli, I. Gerace, R. Pandolfi, and P. Pucci, "A GNC algorithm for deblurring images with interacting discontinuities," in *Proc. VI SIMAI*, Jul. 2002, pp. 296–310.
- [44] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, Mar. 1999.
- [45] Chester F. Carlson Center for Imaging Science, Rochester Inst. Technol., Rochester, NY [Online]. Available: www.cis.rit.edu/people/faculty/easton/k-12/exercise/index.htm
- [46] School of Information Technology and Engineering, Univ. Ottawa, Ottawa, ON, Canada [Online]. Available: <http://www.site.uottawa.ca/edubois/documents>
- [47] Google, Book Search Dataset 2007.



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

She is a researcher at the Signals and Images Laboratory of the Istituto di Scienze e Tecnologie dell'Informazione, Italian National Research Council (CNR), Pisa. She cooperated in several projects for basic and applied research on image processing and computer vision and is the coauthor of over 80 scientific papers. Her present interests are in applied inverse problems, multichannel image restoration and reconstruction, document analysis and recognition, blind source separation, and computational biology.



Ivan Gerace received the degree in computer science from the University of Pisa, Italy, in 1992, and the Ph.D. degree in computational mathematics and operational research from the University of Milan, Italy, in 1999.

He has been an Assistant Professor of numerical analysis since May 2000 at the University of Perugia, Italy. His research interests include image processing, numerical linear algebra, and computational complexity.



Francesca Martinelli was born in Umbertide, Italy, in 1980. She received the degree in mathematics from the University of Perugia, Italy, in 2005, and the Ph.D. degree in mathematics and computer science for information and knowledge management from the University of Perugia in 2009.

Currently, she is a Postdoctoral Fellow at the Istituto di Scienze e Tecnologie dell'Informazione, Italian National Research Council (CNR), Pisa, Italy. Her research interests are in blind image deconvolution, document analysis and recognition, and blind source separation.