BAYESIAN MRF-BASED BLIND SOURCE SEPARATION OF CONVOLUTIVE MIXTURES OF IMAGES

Anna Tonazzini and Ivan Gerace

Istituto di Scienza e Tecnologie dell'Informazione
Area della Ricerca CNR di Pisa
Via G. Moruzzi, 1, I-56124 PISA, Italy
anna.tonazzini@isti.cnr.it

Dipartimento di Matematica e Informatica
Università degli Studi di Perugia
Via Vanvitelli, 1, I-06123 Perugia, Italy
gerace@dipmat.unipg.it

ABSTRACT

This paper deals with the recovery of clean images from a set of their noisy convolutive mixtures. In practice, this problem can be seen as the one of simultaneously separating and restoring source images that have been first degraded by unknown filters, then summed up and added with noise. We approach this problem in the framework of Blind Source Separation (BSS), where the unknown filters, in our case FIR filters in the form of blur kernels, must be estimated jointly with the sources. Assuming the statistical independence of the source images, we adopt Bayesian estimation for all the unknowns, and exploit information about local correlation within the individual sources through the use of suitable Gibbs priors, accounting also for well-behaved edges in the images. We derive an algorithm for recovering the blur kernels that make the estimated sources fit the known properties of the original sources. The method is validated through numerical experiments in a simplified setting, which is however related to real application scenarios.

1. INTRODUCTION

The problem of the joint blind separation and deconvolution of multiple convolutive mixtures of multiple signals/images is known as Multiple Input/Multiple Output (MIMO) blind deconvolution, or multichannel blind deconvolution, and can be formulated in the context of Blind Source Separation (BSS) and Independent Component Analysis (ICA) [1] [5]. The analysis via BSS and ICA of MIMO systems has been first attached in the frequency-domain, since all techniques for instantaneous BSS can be applied independently in each frequency bin [13]. Unfortunately, the permutation indeterminacy, which is inherent in BSS, may make wrong the reconstructions. In the time-based model, various methods have been proposed as well, based on second- and higher-order statistics, neural networks, ensamble learning, and contrast functions [10] [14] [6]. Most of this research has been limited to one dimensional signals, with application in speech and audio processing (e.g. the cocktail party problem), in multiaccess digital communication systems and array processing, and in biomedical signal processing.

More recently, it has been highlighted that also important image processing and computer vision problems can be formulated as instantaneous or convolutive BSS problems [4]. One interesting application is, for instance, the analysis of degraded documents. Indeed, most ancient documents are affected by the overlapping of two or more text patterns, due to seeping or transparency of ink from the reverse side page, which make difficult the legibility by both human and automatic readers. In this case, multiple digital images of the document itself, acquired with different modalities (e.g. multispectral scans) can be modeled as convolutive mixtures of the individual text patterns. BSS techniques can thus be applied to recover legible recto and verso text patterns from documents showing bleed-through or show-through, and underwritings in palimpsests.

In this paper, we propose a general approach to deal with convolutive mixtures of images, when the unknown filtering operators are FIR filters, in the form of blur kernels, and the mixtures are affected by noise, while consider the document analysis application as our case study for the simulations. We adopt a Bayesian estimation formulation, which offers a flexible way to approach the integrated solution of two or more problems, and to account for prior knowledge which can be available. Thus, Bayesian estimation permits to formulate the convolutive BSS problem as the joint estimation of the mixing kernels and the sources. Furthermore, autocorrelation constraints of the individual sources can be naturally enforced through Markov Random Field (MRF) models, in the form of Gibbs priors. These constraints have been proved to be effective for achieving stable solutions in many inverse problems, and especially in those dealing with images, where they correspond to natural features of real physical maps and scenes. MRF models allow for retaining the independence assumption of ICA, and the one we adopt herein has the property of being edge-preserving and of enforcing regularity constraints on the edges themselves. This is an important issue since edges, corresponding to intensity discontinuities due to object boundaries and textures, constitute essential features to be correctly preserved in an image, for analysis and understanding purposes.

We propose an estimation strategy for recovering those kernels that, besides satisfying possible a priori information, make the estimated sources fit the known properties of the original sources. Thus, we reformulate the problem as the estimation of the mixing operator alone, based on the source and mixing priors, while the sources are kept clamped to their Maximum A Posteriori (MAP) estimate, for any status of the mixing. From the theoretical scheme, reasonable approximations are derived which allow for reducing the computational complexity, and finding a remedy to other drawbacks, such as the unavailability of analytical formulas for the sources viewed as functions of the kernels, and non-convexity of the priors. These will make the method computational efficient and still effective. In particular, our method is implemented through an iterative scheme where the Maximum Likelihood (ML) estimation of the mixing alternates with the MAP estimation of the sources. This scheme ensures stability of the solutions and employs GNC-like gradient ascent algorithms to update the sources and simulated annealing (SA) to estimate the blur coefficients.

2. PROBLEM FORMULATION AND ESTIMATION ALGORITHM

The data generation model we consider in this paper is given by:

$$x_i(t) = \sum_{j=1}^{N} \left(H_{ij}\mathbf{s}'_j\right)(t) + n_i(t), \qquad t = 1,2,\ldots,T \\ i = 1,2,\ldots,N \qquad (1)$$

where $x_i(t)$, $s_i(t)$ and $n_i(t)$ represents the $i-th$ observation, source, and noise or measurement error at location t, respectively. Of course, in imaging location t stands for the couple of pixel indices. Although not necessary, for simplicity sake, the same number N of measurements and unknown sources has been assumed. In eq. (1) and through all the paper, vectors $\mathbf{s}_i = (s_i(1), s_i(2), ..., s_i(T))$, $i = 1,2,...,N$, represent the lexicographically ordered notation of the various sources, $\mathbf{s}(t)$ is the column vector of all the unknown sources at location t, and $\mathbf{s} = (\mathbf{s}(1), ..., \mathbf{s}(T))$ is the matrix whose t-th column contains the N sources at location t, and whose i-th row is the source $\mathbf{s}_i$. These definitions extend to data and noise as well. Quantity $H_{ij}\mathbf{s}'_j$ is the degraded version of source $\mathbf{s}_j$ which contributes to $\mathbf{x}_i$, where the blur matrix H_{ij}, assumed unknown, is the block Toeplitz matrix that performs convolution between a source image and a blur mask as a matrix-vector product. In the following, the set of all blur matrices will be indicated by $\mathbf{H}$.

The problem of estimating kernels H_{ij} and the deblurred source samples s could be stated as the following joint MAP estimation problem:

$$(\hat{\mathbf{s}}, \hat{\mathbf{H}}) = arg\max_{\mathbf{s},\mathbf{H}} P(\mathbf{s}, \mathbf{H}, |\mathbf{x}) = arg\max_{\mathbf{s},\mathbf{H}} P(\mathbf{x}|\mathbf{s}, \mathbf{H})P(\mathbf{s})P(\mathbf{H}) \qquad (2)$$

where, from the independence assumption, $P(\mathbf{s})$ is given by:

$$P(\mathbf{s}(t)) = \prod_{i=1}^{N} P_i(s_i(t)) \qquad \forall t. \qquad (3)$$

Joint MAP estimation is usually approached by means of alternating componentwise maximization with respect to the two sets of variables in turn [11]. Here, we propose instead the following estimation scheme [8]:

$$\hat{\mathbf{H}} = arg\max_{\mathbf{H}} P\left(\hat{\mathbf{s}}(\mathbf{H})\right)P(\mathbf{H}) \qquad (4)$$

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

The rationale for this particular estimation strategy is the looking for the mixing kernels that, besides satisfying their own known properties, make the estimated sources fit the a priori knowledge we possess about the ideal sources. In this way, the original joint MAP estimation is reformulated as the ML estimation of the mixing alone, based on the source and mixing priors, while the sources are kept clamped to their MAP estimate, for any status of the mixing. The dependence of the mixing from the data is indirectly retained through the sources. A similar strategy was successfully proposed for BSS of images from noisy instantaneous mixtures [7]. In that case, assuming that no prior information is available on the mixing matrix A, our method becomes an extension to the noisy case of the ML approach to noiseless ICA, where the constraint $\hat{\mathbf{s}}(A) = A^{-1}\mathbf{x}$ is substituted by a MAP estimation for $\mathbf{s}$.

As per the prior P($\mathbf{s}$), in the form of eq. (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. Accounting for an edge process, and especially a well-behaved one, is particularly useful for deconvolution, when blur must be removed. In the Gibbs/MRF formalism our priors are given by:

$$P_i(\mathbf{s}_i) = \frac{1}{Z_i} \exp\left\{-U_i(\mathbf{s}_i)\right\} \qquad (6)$$

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, 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, while the functional form of the potentials determines the correlation strength, and various features of the image edges. In our case we express the regularity of edges by penalizing parallel, adjacent edges, and chose $U_i(\mathbf{s}_i)$ as:

$$U_i(\mathbf{s}_i) = \sum_t \sum_{(r,z)\in N_t} \psi_i\left((s_i(t) - s_i(r)), (s_i(r) - s_i(z))\right) \qquad (7)$$

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. As stabilizers ψ_i, all having same functional form but possible different hyperparameters, in order to graduate the constraint strength in dependence of the source considered, we chose the following functions [2]:

$$\psi_i(\xi_1, \xi_2) = \begin{cases} \begin{cases} \lambda_i \xi_1^2 & \text{if} |\xi_1| < \theta \\ \alpha_i & \text{if} |\xi_1| \geq \theta \end{cases} & \text{if} |\xi_2| < \theta \\ \begin{cases} \lambda_i \xi_1^2 & \text{if} |\xi_1| < \bar{\theta} \\ \alpha_i + \varepsilon_i & \text{if} |\xi_1| \geq \bar{\theta} \end{cases} & \text{if} |\xi_2| \geq \theta. \end{cases} \qquad (8)$$

In eq. (8), λ_i is a positive weight, the so-called regularization parameter, the quantity $\theta = \sqrt{\alpha_i/\lambda_i}$ has the meaning of a *threshold* on the intensity gradient above which a discontinuity is expected, while $\bar{\theta} = \sqrt{(\alpha_i + \varepsilon_i)/\lambda_i}$ is a *suprathreshold*, higher than the threshold, to lower the expectation of an edge when a parallel, close edge is likely to be present. 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 eq. (7) a new stabilizer which favors the creation of an edge when a contiguous, aligned edge is likely to be present. This stabilizer should have the same form of eq. (8), except that it should refer to couples of aligned adjacent locations, and the suprathreshold should be lower than the threshold.

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

$$log\left(P(\mathbf{x}|\mathbf{s},\mathbf{H})\right) = -\frac{1}{2}\sum_{t}\left(\mathbf{H}\mathbf{s}(t)-\mathbf{x}(t)\right)'\Sigma_t^{-1}\left(\mathbf{H}\mathbf{s}(t)-\mathbf{x}(t)\right) \tag{9}$$

where Σ is the covariance matrix of the noise, assumed, in general, to be location-dependent, and $\mathbf{Hs}(t)$ indicates the column vector of the degraded sources at location t. Problem in eqs. (4)-(5), in view of data model of eq. (1), likelihood of eq. (9) and priors of eqs. (6)-(8), can be reformulated as:

$$\hat{\mathbf{H}} = arg\min_{\mathbf{H}}\sum_{i}U_i\left(\hat{\mathbf{s}}_i(\mathbf{H})\right) - logP(\mathbf{H}) \tag{10}$$

$$\hat{\mathbf{s}}(\mathbf{H}) = arg\min_{\mathbf{s}}\frac{1}{2}\sum_{t}\left(\mathbf{H}\mathbf{s}(t)-\mathbf{x}(t)\right)'\Sigma_t^{-1}\times$$
$$\times\left(\mathbf{H}\mathbf{s}(t)-\mathbf{x}(t)\right)+\sum_{i}U_i\left(\mathbf{s}_i\right). \tag{11}$$

The solution of problem in eqs. (10)-(11) presents some computational difficulties. Indeed, in general it is not possible to derive analytical formulas for the sources viewed as functions of $\mathbf{H}$, and the priors are not convex. Thus, a simulated annealing (SA) algorithm has to be adopted for the updating of $\mathbf{H}$ and the sources must be computed through numerical estimation. If a correct SA schedule was used and for each proposal of a new status for $\mathbf{H}$ the corresponding sources were estimated, this scheme would ensure convergence to the global optimum. Nevertheless, its computational complexity would be prohibitive. However, some reasonable approximations can be adopted to reduce the complexity of the original problem, while keeping the effectiveness of the approach. First of all, due to the usual small number of coefficients for the blur kernels, SA is not particularly cumbersome in this case. On the other hand, based on the feasible assumption that small changes in $\mathbf{H}$ do not affect too much the sources, these can be updated only after significant modifications, e.g. at the end of a complete visitation of all the filter coefficients. Furthermore, though the posterior is non-convex as well, the image models we adopted allow for performing the MAP source estimation through efficient deterministic non-convex optimization algorithms, such as the Graduated Non-Convexity (GNC) algorithm [3]. A GNC-like algorithm for the specific stabilizer in eq. (8) was derived in [2], in the case of image denoising. In [9], the same algorithm has been extended to account for images degraded by a linear operator. In this form, the algorithm is suitable to be applied to our present separation problem, where the linear operator is given by the combination of the blur operators. The whole blind separation algorithm reduces thus to an alternating scheme governed by an external simulated annealing for the estimation of $\mathbf{H}$, according to eq. (10), interrupted at the end of each Metropolis cycle, to perform an update of the sources $\mathbf{s}$, according to eq. (11). In all the experiments performed using this scheme, stabilization of the solutions was always observed.

3. EXPERIMENTAL RESULTS: A DOCUMENT ANALYSIS CASE

The performance of the Bayesian blind separation method from noisy convolutive mixtures of images described in this paper was tested on synthetic images that simulate the scans of couples of recto and verso pages of documents exhibiting the bleed-through or show-through effect. As already said, these distortions are very frequent in ancient documents, due to the kind of ink used, which seeps and smears through the paper support, but can also be present in scans of modern documents, when the paper is not completely opaque.

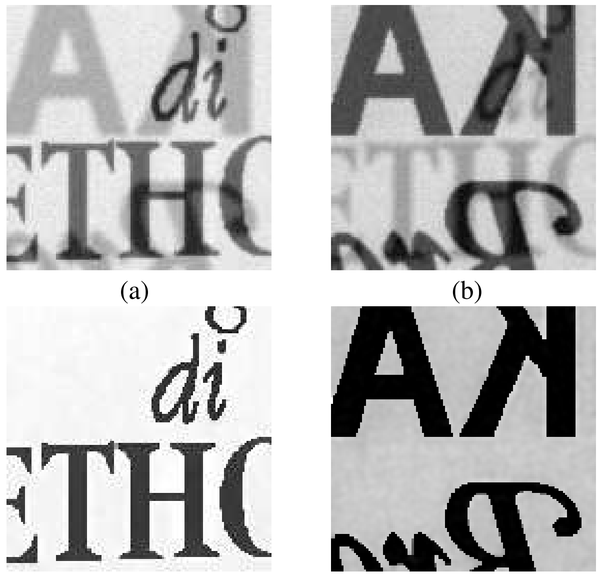
The problem of recovering clean recto and verso images of such documents can be seen as a 2×2 case of the convolutive BSS problem formulated in eq. (1), where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the grayscale images obtained by scanning, respectively, the recto and the horizontally flipped verso page of a document, and $\mathbf{s}_1$ and $\mathbf{s}_2$ represent the clean main texts in the front and back page, respectively. In general, kernels H_{ij} are filters modeling the blur effect due to the digital acquisition process, ink fading and other degradations. In particular, for $i\neq j$, H_{ij} model also the often strong smearing and attenuation of the ink seeping from the back to the front page, and vice versa. This ink smearing is mainly due to the chemical composition of the ink, and the characteristics of the transmission medium (paper or parchment). Usually, these kernels are not known, both in size and coefficients, so that we should estimate them jointly with the sources. In this specific application, however, some physical constraints on the mixing process can be exploited, which allow us to simplify the problem, i.e. to further reduce its dimension. For instance, providing that the two pages have been written with the same ink, same pressure and at two close moments, it is reasonable to assume that the attenuation/smearing of the bleed-through pattern in the two pages is the same, i.e. $H_{12}=H_{21}$. For similar considerations, it is also expected that $H_{11}=H_{22}$. Furthermore, the intensity of the front text pattern in the recto page should be higher than that of the bleed-through pattern, i.e. the sum of the coefficients of the blur mask associated with H_{12} should be lower than that of the blur mask associated with H_{11}, with the same relationship holding, reversed, in the verso page.

Neglecting the blur effect, and considering only the attenuation coefficients, this problem reduces to an instantaneous BSS problem, and we showed that the above symmetry features can be indirectly exploited to efficiently solve the problem via symmetric whitening [12].

In the approach described here, the symmetry features of the problem can be easily exploited in a direct manner. Thus, the number of kernels to be estimated can be reduced to two, and the conditions on the kernel elements can be enforced as well as constraints in the SA estimation algorithm. These constraints can be described through the prior $P(\mathbf{H})$. With respect to the kernel sizes, we assumed to can estimate them off-line, for instance looking at the slope extent of the character boundaries in the two superimposed patterns.

A typical experiment among the ones we carried out is shown in Figure 1. To obtain the recto and verso images of Figures 1(a) and 1(b), the original, clean text of the front (back) page was blurred with a 5×5 Gaussian-like blur mask h_1 whose element sum is 0.8, while the original, clean text of the back (front) page was first horizontally flipped, and then blurred with a still 5×5 but more uniform mask h_2, with element sum equal to 0.3, in order to obtain stronger blur and attenuation for the bleed-through patterns with respect to the foreground text patterns. Both the recto and verso images were then added with a space-invariant white Gaussian noise (SNR=26 dB).

As prior for $\mathbf{H}$, we considered the positivity of the blur mask coefficients, and assumed to know that the sum of the elements of h_2 is lower than that of the elements of h_1. These constraints are easily enforced through SA. With respect to the source priors, since the two ideal text patterns have similar characteristics, we assumed the same stabilizer of eq. (8) for each of them, but adopted different parameters, to account for the different scale of the characters and the different contrast between text and background. With parameters $\lambda = 0.8$, $\alpha = 250$ and $\varepsilon = 250$ for the front text image, and $\lambda = 0.5$, $\alpha = 160$ and $\varepsilon = 160$ for the back text image, we obtained the results shown in Figures 1(c) and 1(d). The good performance of our method in this synthetic but realistic case is apparent, both for separation and deconvolution purposes. In particular, we can appreciate the good reconstruction of the fine scale characters in the front text image of Figure 1(c). Experiments on real documents are planned for the near future.



(a) (b) (c) (d)

Figure 1: Joint separation and deconvolution of overlapped text patterns in documents affected by bleed-through distortion: (a) first mixture simulating the recto image of a document page; (b) second mixture (horizontally flipped verso image); (c) first estimated source, representing the clean text of the front page; (d) second estimated source, representing the clean, horizontally flipped, text of the back page.

REFERENCES

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

[2] L. Bedini, I. Gerace and A. Tonazzini, "A deterministic algorithm for reconstructing images with interacting discontinuities", Graph. Models Image Proc., vol. 56, pp. 109–123, 1994.

[3] A. Blake and A. Zisserman, Visual Reconstruction. Cambridge (MA): MIT Press, 1987.

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

[5] P. Comon, "Independent Component Analysis, a new concept?", Signal Processing, vol. 36, pp. 287-314, 1994.

[6] P. Comon, "Contrasts for multichannel blind deconvolution", IEEE Signal Processing Letters, vol. 3(7), pp. 209-211, 1996.

[7] I. Gerace, F. Cricco and A. Tonazzini, "An extended Maximum Likelihood approach for the robust blind separation of autocorrelated images from noisy mixtures", LNCS, vol. 3195, pp. 954-961, 2004 (Proc. ICA 2004).

[8] I. Gerace, R. Pandolfi and P. Pucci, "A new estimation of blur in the blind restoration problem", in Proc. ICIP 2003 p. 4.

[9] I. Gerace, P. Pucci, A. Boccuto, M. Discepoli and R. Pandolfi, "A new technique for restoring blurred images with convex first approximation", submitted, 2004.

[10] H.-L. Nguyen Thi and C. Jutten, "Blind source separation for convolutive mixtures", Signal Processing, vol. 45, pp. 209-229, 1995.

[11] A. Tonazzini, L. Bedini, E. Kuruoglu and E. Salerno, "Blind separation of auto-correlated images from noisy mixtures using MRF models", in Proc. ICA 2003, pp. 675-680.

[12] A. Tonazzini, E. Salerno and L. Bedini, "Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique", Int. J. on Document Analysis and Recognition, to appear, 2005.

[13] L. Tong, G. Xu, B. Hassibi and T. Kailath, "Blind channel identification based on secon order statistics: a frequency-domain approach", IEEE Trans. on Information Theory, vol. 41(1), pp. 329-334, 1995.

[14] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-gaussian processes using higher order statistics and inverse filter criteria", IEEE Trans. on Signal Processing, vol. 45(3), pp. 658-672, 1997.