BAYESIAN SEPARATION OF DOCUMENT IMAGES WITH HIDDEN MARKOV MODEL

Feng Su[†][*], Ali Mohammad-Djafari[†]

[†]*Laboratoire des Signaux et Systemes, UMR 8506 (CNRS-Supelec-UPS)*
Supelec, Plateau de Moulon, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France
[*]*State Key Laboratory for Novel Software Technology, Nanjing University, 210093 Nanjing, P. R. China*

Keywords: Blind Source Separation, document image, Bayesian estimation, HMM, MCMC.

Abstract: In this paper we consider the problem of separating noisy instantaneous linear mixtures of document images in the Bayesian framework. The source image is modeled hierarchically by a latent labeling process representing the common classifications of document objects among different color channels and the intensity process of pixels given the class labels. A Potts Markov random field is used to model regional regularity of the classification labels inside object regions. Local dependency between neighboring pixels can also be accounted by smoothness constraint on their intensities. Within the Bayesian approach, all unknowns including the source, the classification, the mixing coefficients and the distribution parameters of these variables are estimated from their posterior laws. The corresponding Bayesian computations are done by MCMC sampling algorithm. Results from experiments on synthetic and real image mixtures are presented to illustrate the performance of the proposed method.

1 INTRODUCTION

Blind source separation (BSS) is an active research topic of signal and image processing in recent years. It considers separating a set of unknown signals from their observed mixtures, with reasonable assumptions of the form of the mixing process: linear or nonlinear, instantaneous or convoluting, under or over determined, noisy or noiseless, and so on. However, in all cases the mixing coefficients remain unknown and have to be estimated as well as original source signals.

Various methods and models have been proposed for BSS task, among which Principal Component Analysis (PCA) seeks orthogonal directions of maximum variance exhibited by the data as source axes, while Independent Component Analysis (ICA) (Hyvarinen et al., 2001), in its basic form, assumes statistical independency of sources and linear mixing process and consists of seeking an inverse linear transformation matrix applying on the data to achieve maximum mutual independency between output components. Both methods exploits basic statistical characteristics of source signals to achieve the separation, which makes them well generalizable and robust in

cases that as few prerequisite assumptions as uncorrelatedness or independency can be made about the source. Some variant algorithms are also proposed to adapt to certain relaxation of model assumptions like nonlinearity or noises (Harmeling, 2003; Almeida, 2005). However, in many other cases, we may find the availability or the needs of various types of prior information to regulate the essentially ill-posed BSS problem. Compared with PCA and ICA, Bayesian framework allows convenient introduction of these prior constraints about the sources and the mixing coefficients, and more important, supports flexible structuring and integrating multiple hierarchical clues for separation purpose.

In the field of image processing, BSS approaches are being widely employed to separate or segment mixed images observed from, for example, satelite and hyper-spectral imaging (Snoussi and Mohammad-Djafari, 2004; Parra et al., 2000; Macias-Macias et al., 2003), medical imaging (Calhoun and Adali, 2006; Snoussi and Calhoun, 2005), and other superimpositions of natural images (Bronstein et al., 2005; Castella and Pesquet, 2004).

This paper focuses on one specific type of im-

ages - document images, where superimposition of two images usually appears as a major type of degradation encountered in digitization (Sharma, 2001) or ancient documents (Drira, 2006). The former degradation usually occurs as artifact during scanning a double-sided document when the text on the backside printing shows through the non-opaque medium and are mixed with the foreside text. Fig.1a shows one such example. The latter cause of text superimposition, usually called bleed-through, can often be observed in old documentations due to ink blurring or penetrating as illustrated by Fig.1b. Other forms of overlapped patterns, like underwriting and watermarks, are also common. Though the actual underlying mixing process may be quite complicated and diverse in various mixture forms, the linear mixing model usually serves as a resonable approximation and benefits analytical and computational simplicity, thus is adopted in most document separation cases.

To separate document image mixtures, the common PCA and ICA algorithms can be used and have shown their effectiveness in detecting independent document features like watermarks, as inspected in (Tonazzini et al., 2004) where each source was considered as random signal sequence in a whole without further internal structuring. The Bayesian framework has also been used before for document separation as in (Tonazzini et al., 2006), where the source is modeled by a Markov Random Field on the pixel values to account for local smoothness inside one object, as well as an extra line process enforcing the discontinuity at object edges.

In this contribution, we propose a solution to jointly separate and segment linearly mixed document images. Besides considering the mixture in single grayscale channel, we address the joint separation of multi-channel mixture of multiple sources. In section 2, we give the probability formulation of the problem. In section 3, the algorithm of Bayesian estimation for model parameters is described. In section 4, simulation results of the proposed algorithm are shown on both synthetic and real images.
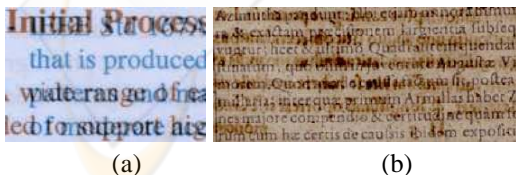

Figure 1: Examples of mixed document images: a) show-through mixture; b) bleed-through mixture.

2 MODEL ASSUMPTION AND FORMULATION

Document images are created by various digitization methods from vast types of documentation. Commonly, a color scanner can be used to produce three different views of one document in the red, green, and blue channels. With detectors working in non-visible wavelengths such as infrared and ultraviolet, even more information channels of data can be obtained, depending on the object of interest in documents.

Given observations of M different mixtures, either in grayscale or multiple channels, our work is thus to obtain N corresponding source images (normally $M > N$) in the same pixel format as the observations.

2.1 Data Model

In this work, the observations are M registered images $(X_i)_{i=1...M}$, which are defined on the same set of pixels $\mathcal{R}$: $X_i = \{x_i(r)\}_{r \in \mathcal{R}}$. The observations are noisy linear instantaneous mixture of N source images $(S_j)_{j=1...N}$ also defined on $\mathcal{R}$, following the data generation model given by:

$$\mathbf{x}(r) = \mathbf{A}\mathbf{s}(r) + \mathbf{n}(r) \qquad r \in \mathcal{R} \qquad (1)$$

where $A = (a_{ij})_{M \times N}$ is the unknown mixing matrix, $\mathbf{n}(r)$ is a set of independent zero-mean white Gaussian noise for each observation with variance $\sigma_\varepsilon^2 = (\sigma_{\varepsilon 1}^2 \dots \sigma_{\varepsilon M}^2)$, $\mathbf{x}(r)$ and $\mathbf{s}(r)$ are the observation and source vector at pixel r respectively. Let $\mathbf{S} = \{\mathbf{s}(r), r \in \mathcal{R}\}$, $\mathbf{X} = \{\mathbf{x}(r), r \in \mathcal{R}\}$, and denote the noise covariance matrix by $\mathbf{R}_\varepsilon = diag[\sigma_{\varepsilon 1}^2 \dots \sigma_{\varepsilon M}^2]$, we have the Gaussian distribution for the observations given the sources and the mixing parameters:

$$p(\mathbf{X}|\mathbf{S}, \mathbf{A}, \mathbf{R}_\varepsilon) = \prod_r \mathcal{N}(\mathbf{A}\mathbf{s}(r), \mathbf{R}_\varepsilon) \qquad (2)$$

2.2 Source Model

We model the distribution of pixel intensity for each source images (and for each color channel) by a Mixture of Gaussians (MoG), whose components correspond to each object type (or class) that appears roughly equal pixel values. For example, the simplest model may consist of two components, one for foreground text and the other for background blank. Furthermore, to allow imposing constraints on distribution of class labels, for every source S_j we represent the class labels by a set of discrete hidden variables $Z_j = \{z_j(r), r \in \mathcal{R}\}$ with $z_j(r)$ taking values from $\{1, \dots, K_j\}$, where K_j is the total number of classes

in image S_j. In the following, we assume all $\{K_j\}$ equal to the same value K.

Given pixel labels, pixels of different classes can be reasonably assumed independent, while concerning the pixels inside a given class, there are usually two choices:

1. We may assume pixel intensities are conditionally independent given their labels;

2. Alternatively, we may explicitly take into account the local dependency between neighboring pixels of same class.

In the first choice, the distribution of pixel r in the jth source is modeled by:

$$p(s_j(r)|z_j(r) = k) = \mathcal{N}(\mu_{jk}, \sigma_{jk}^2) \qquad (3)$$

where μ_{jk} and σ_{jk}^2 are the mean and variance of the kth Gaussian component of the jth source. Assuming independency between different sources and denoting the set of labels corresponding to every source by $\mathbf{Z} = \{Z_j, j = 1 \dots N\}$, we have:

$$p(\mathbf{S}|\mathbf{Z}) = \prod_j \prod_k \prod_{\{r:z_j(r)=k\}} p(s_j(r)|z_j(r) = k)$$

which is by (3) also a Gaussian and spatially separable on r.

In the second choice, the local dependency can be accounted by extra smoothness constraints, like the mean value, between neighboring pixels. We first assign a binary valued contour flag $q_j(r)$ for every pixel r of every source j, which is deterministicly computed by:

$$q_j(r) = \begin{cases} 1 & \text{if} \quad z_j(r') = z_j(r), \forall r' \in \mathcal{V}(r) \\ 0 & \text{else} \end{cases}$$

where $\mathcal{V}(r)$ denotes the neighbor sites of the site r.

Then, based on the value of the contour flag and possibly current values of the neighboring pixels, the distribution of intensity of individual pixel is formulated as:

$$p(s_j(r)|z_j(r) = k, s_j(r'), r' \in \mathcal{V}(r)) = \mathcal{N}(\bar{s}_j(r), \bar{\sigma}_j^2(r)) \qquad (4)$$

with,

$$\bar{s}_j(r) = q_j(r)\mu_{jk} + (1 - q_j(r))\frac{1}{|\mathcal{V}_{jk}(r)|} \sum_{r' \in \mathcal{V}_{jk}(r)} s_j(r')$$

$$\bar{\sigma}_j^2(r) = q_j(r)\sigma_{jk}^2 + (1 - q_j(r))\sigma_j^2$$

where $\mathcal{V}_{jk}(r)$ denotes the intersection of $\mathcal{V}(r)$ with the site set $\mathcal{R}_{jk} = \{r : z_j(r) = k\}$, σ_j^2 is the *a prior* variance of pixel values inside a region. Eqn.(4) states that at the contour pixel intensities follow the Gauss

distribution whose parameters are determined by the class labels as (3), while inside a region the distribution parameters are computed from the neighboring pixels. Note that under this assumption, $p(\mathbf{S}|\mathbf{Z})$ is no longer separable on r, but with the *parallel Gibbs sampling* scheme proposed in (Feron and Mohammad-Djafari, 2005), it can still be simulated efficiently.

As a commonly observed property of visual objects, pixels belonging to the same object usually connect to each other in a neighborhood of space, forming several connected regions of uniformly classified pixels, for instance, the multiple components constituting a text. By class labels defined earlier, this implies regional smoothness of the spatial distribution of class labels. This can be naturally modeled by *a prior* Potts Markov Random Field for every label process $z_j(r)$:

$$p(z_j(r), r \in \mathcal{R}) \propto \exp\left[\beta_j \sum_{r \in \mathcal{R}} \sum_{r' \in \mathcal{V}(r)} \delta(z_j(r) - z_j(r'))\right] \qquad (5)$$

The parameter β reflects the degree of smoothing interactions between pixels and controls the expected size of the regions. In our work, all $\{\beta_j\}$ are assumed equal and assigned an empirical value within $[1.5, 2.0]$.

2.3 Multiple Channels

When multi-channel image data are considered, there are multiple options for the processing model. We can perform separation of sources independently in each channel and by some measures merge the results in the end. Or, we may consider joint demixing for all channels. In the latter case, the mixing model can still have more alternatives:

a) all channels are equally mixed with the same mixing matrix;

b) the mixing occurs separately in each channel with different mixing matrices;

c) cross-channel mixing is assumed to be present.

In the case of a), samples from different channels of the same observation can be concatenated for estimation of the mixing coefficients, which is similar to the monochrome case. In the case of c), an expanded mixing matrix $\mathbf{A}_{ML \times NL}$ (supposing L channels) is used for all channels of all sources.

In this work, we assume the model b), where the mixing in different channels are mutually independent and with their own separate coefficients. Thus, in RGB color format, the sources

and observations are actually $\{S_j^r, S_j^g, S_j^b\}_{j=1\ldots N}$ and $\{X_i^r, X_i^g, X_i^b\}_{i=1\ldots M}$. Correspondingly, there are $\{\mathbf{A}^r, \mathbf{R}_\varepsilon^r, \mu_{jk}^r, \ldots \mathbf{A}^b, \mathbf{R}_\varepsilon^b, \mu_{jk}^b\}$ and so on. But for each source S_j, only one classification field Z_j is maintained and shared by all channels, as a natural way to enforce the common segmentation among different channels. This two-level hierarchical source model, which also facilitates introducing segmentation constraints like discontinuity and local regional dependency, is the main difference with the work of (Tonazzini et al., 2006), where a one-level MRF modeling of sources is defined on the single-channel pixel intensities along with an explicit binary edge process.

3 BAYESIAN ESTIMATION OF MODEL PARAMETERS

The unknown variables we want to estimate in the models given above are $\{\mathbf{S}, \mathbf{Z}, \mathbf{A}, \theta\}$, θ representing all hyperparameters. The Bayesian estimation approach consists of deriving the posterior distribution of all the unknowns given the observation and then based on this distribution, employing appropriate estimators such as Maximum A Posteriori (MAP) or the Posterior Means (PM) for them. With our model assumptions, this posterior distribution can be expressed as:

$$p(\mathbf{S}, \mathbf{Z}, \theta | \mathbf{X}) \propto p(\mathbf{X} | \mathbf{S}, \mathbf{A}, \mathbf{R}_\varepsilon) p(\mathbf{S} | \mathbf{Z}, \theta_s) p(\mathbf{Z}) p(\theta) \tag{6}$$

where, $\theta_s = \{(\mu_{jk}, \sigma_{jk}^2), j = 1 \ldots N, k = 1 \ldots K\}$ and $\theta = \{\mathbf{A}, \mathbf{R}_\varepsilon, \theta_s\}$.

3.1 Prior Assignments for Model Parameters

According to the linear mixing model and all Gaussian assumptions, we choose corresponding conjugate priors for model hyperparameters.

- Gaussian for source means

$$\mu_{jk} \sim \mathcal{N}(\mu_{k0}, \sigma_{k0}^2)$$

- Inverse Gamma for source variances

$$\sigma_{jk}^2 \sim I\mathcal{G}(\alpha_{k0}, \beta_{k0})$$

- Inverse Wishart for noise covariance

$$\mathbf{R}_\varepsilon^{-1} \sim \mathcal{W}_i^l(\alpha_{\varepsilon_0}, \beta_{\varepsilon_0})$$

In this work, we assign uniform prior to $\mathbf{A}$ for simplicity and no preference of the mixing coefficients, while in other cases prior distributions like Gamma may be used to enforce positivity.

3.2 Estimation by MCMC Sampling

Given the joint *a posteriori* distribution (6) of all unknown variables, we use the Posterior Means as the estimation for them. Since direct integration over $\mathbf{z}$ is intractable, MCMC methods are employed in the actual Bayesian computations. In our work, a Gibbs sampling algorithm is used to generate a set of samples for every variable to be estimated, according to its full-conditional *a posteriori* distribution given all other variables fixed to their current values. Then, after certain burn-in runs, sample means from further iterations are used as the Posterior Means estimation for the unknowns. The algorithm takes the form:

Repeat until converge,
1. simulate $\mathbf{S}' \sim p(\mathbf{S} | \mathbf{Z}, \theta, \mathbf{X})$
2. simulate $\mathbf{Z}' \sim p(\mathbf{Z} | \mathbf{S}', \theta, \mathbf{X})$
3. simulate $\theta' \sim p(\theta | \mathbf{Z}', \mathbf{S}', \mathbf{X})$

Below we give the expressions of related conditional probability distributions.

- Sampling $\mathbf{Z} \sim p(\mathbf{Z} | \mathbf{X}, \mathbf{S}, \theta) \propto p(\mathbf{X} | \mathbf{Z}, \theta) p(\mathbf{Z})$:

$$\begin{aligned} p(\mathbf{X} | \mathbf{Z}, \theta) &= \prod_r p(\mathbf{x}(r) | \mathbf{z}(r), \theta) \\ &= \prod_r \mathcal{N}(\mathbf{A}\mathbf{m}_{\mathbf{z}(r)}, \mathbf{A}\Sigma_{\mathbf{z}(r)}\mathbf{A}^t + \mathbf{R}_\varepsilon) \end{aligned}$$

where, $\mathbf{m}_{\mathbf{z}(r)} = [\mu_{1z_1(r)}, \ldots, \mu_{Nz_N(r)}]^t$ and $\Sigma_{\mathbf{z}(r)} = diag[\sigma_{1z_1(r)}^2, \ldots, \sigma_{Nz_N(r)}^2]$.

Notice $p(\mathbf{Z}) = \prod_{j=1}^N p(\mathbf{z}_j)$, and as mentioned earlier, $p(\mathbf{z}_j)$ takes the form of Potts MRF as (5). An inner Gibbs sampling is then used to simulate $\mathbf{z}_j$ with the likelihood $p(\mathbf{x}(r) | \mathbf{z}(r), \mathbf{A}, \theta)$ marginalized over all configurations of $\{\mathbf{z}_{j'}(r), j' \neq j\}$.

- Sampling $\mathbf{S} \sim p(\mathbf{S} | \mathbf{X}, \mathbf{Z}, \theta)$:

$$\begin{aligned} p(\mathbf{S} | \mathbf{X}, \mathbf{Z}, \theta) &\propto p(\mathbf{X} | \mathbf{S}, \mathbf{A}, \mathbf{R}_\varepsilon) p(\mathbf{S} | \mathbf{Z}, \theta) \\ &= \prod_r \mathcal{N}(\mathbf{m}_s^{apost}(r), \mathbf{R}_s^{apost}(r)) \end{aligned}$$

$$\begin{cases} \mathbf{R}_s^{apost}(r) &= \left[\mathbf{A}^t \mathbf{R}_\varepsilon^{-1} \mathbf{A} + \Sigma_{\mathbf{z}(r)}^{-1}\right]^{-1} \\ \mathbf{m}_s^{apost}(r) &= \mathbf{R}_s^{apost}(r)\left[\mathbf{A}^t \mathbf{R}_\varepsilon^{-1} \mathbf{x}(r) + \Sigma_{\mathbf{z}(r)}^{-1} \mathbf{m}_{\mathbf{z}(r)}\right] \end{cases}$$

- Sampling $\mathbf{R}_\varepsilon$:

$$p(\mathbf{R}_\varepsilon | \mathbf{X}, \mathbf{S}, \mathbf{A}) \propto p(\mathbf{X} | \mathbf{S}, \mathbf{A}, \mathbf{R}_\varepsilon) p(\mathbf{R}_\varepsilon)$$

Considering we assign an inverse Wishart distribution to $p(\mathbf{R}_\varepsilon)$, which is conjugate prior for the likelihood (2), $\mathbf{R}_\varepsilon$ is *a posteriori* sampled from:

$$\begin{cases} \mathbf{R}_\varepsilon^{-1} \sim \mathcal{W}_i^l(\alpha_\varepsilon, \beta_\varepsilon) \\ \alpha_\varepsilon = \frac{1}{2}(|\mathcal{R}| - n), \beta_\varepsilon = \frac{1}{2}|\mathcal{R}|(\mathbf{R}_{xx} - \mathbf{R}_{xs}\mathbf{R}_{ss}^{-1}\mathbf{R}_{xs}^t) \end{cases}$$

where, the sample statistics $\mathbf{R}_{xx} = \frac{1}{|\mathcal{R}|}\sum_r \mathbf{x}_r \mathbf{x}_r^t$, $\mathbf{R}_{xs} = \frac{1}{|\mathcal{R}|}\sum_r \mathbf{x}_r \mathbf{s}_r^t$, $\mathbf{R}_{ss} = \frac{1}{|\mathcal{R}|}\sum_r \mathbf{s}_r \mathbf{s}_r^t$.

- Sampling $\mathbf{A} \sim p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{R}_\varepsilon)$:

$$p(\mathbf{A}|\mathbf{X},\mathbf{S},\mathbf{R}_\varepsilon) \propto p(\mathbf{X}|\mathbf{S},\mathbf{A},\mathbf{R}_\varepsilon)p(\mathbf{A})$$

Given uniform or Gaussian prior for $\mathbf{A}$, the posterior distribution of $\mathbf{A}$ is a Gaussian:

$$\begin{cases} Vec(\mathbf{A}) \sim \mathcal{N}(\mu_A, \mathbf{R}_A) \\ \mu_A = Vec(\mathbf{R}_{xs}\mathbf{R}_{ss}^{-1}), \mathbf{R}_A = \frac{1}{|\mathcal{R}|}\mathbf{R}_{ss}^{-1} \otimes \mathbf{R}_\varepsilon \end{cases}$$

where $\otimes$ is the Kronecker product and $Vec(.)$ represents the column-stacking operation.

- Sampling (μ_{jk},σ_{jk}^2):

With (Z,S) sampled in earlier steps and conjugate priors assigned, the means μ_{jk} and the variances σ_{jk}^2 can be sampled from respective posteriors as follows:

$$\begin{cases} \mu_{jk}|\mathbf{s}_j,\mathbf{z}_j,\sigma_{jk}^2 &\sim& \mathcal{N}(m_{jk}, v_{jk}^2) \\ m_{jk} &=& v_{jk}^2\left(\frac{\mu_{k0}}{\sigma_{k0}^2} + \frac{1}{\sigma_{jk}^2}\sum_{r\in\mathcal{R}_k^{(j)}}\mathbf{s}_j(r)\right) \\ v_{jk}^2 &=& \left(\frac{n_k^{(j)}}{\sigma_{jk}^2} + \frac{1}{\sigma_{k0}^2}\right)^{-1} \end{cases}$$

and,

$$\begin{cases} \sigma_{jk}^2|\mathbf{s}_j,\mathbf{z}_j,\mu_{jk} &\sim& I\mathcal{G}(\alpha_{jk},\beta_{jk}) \\ \alpha_{jk} &=& \alpha_{k0} + \frac{n_k^{(j)}}{2} \\ \beta_{jk} &=& \beta_{k0} + \frac{1}{2}\sum_{r\in\mathcal{R}_k^{(j)}}(\mathbf{s}_j(r) - \mu_{jk})^2 \end{cases}$$

where, label region $\mathcal{R}_k^{(j)} = \{r : \mathbf{z}_j(r) = k\}$ and the region size $n_k^{(j)} = |\mathcal{R}_k^{(j)}|$.

4 SIMULATION RESULTS

For evaluating the performance of the proposed algorithm, we use both synthetic and real images in the test. The synthetic images were generated according to the model setting that each source is composed of pixels of two classes (text and background) and two source images are linearly mixed in every color channel independently to produce two observation images. This was done in three steps:

1. Two binary ($K_{j=1,2} = 2$) text image were scanned from real documents or created by graphic tools. They were used as the class labels $Z_{j=1,2}$ for each source;

2. With known means and variances for pixel value of each class, the source images were generated according to (3);

3. For each color channel, a random selected $\mathbf{A}_{2\times2}$ was used to mix the sources and finally white Gaussian noises $\mathbf{R}_\varepsilon$ were added (SNR=20dB).
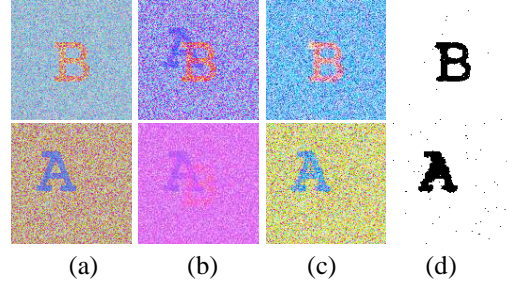

Figure 2: Separation of synthetic image mixtures: a) original sources; b) image mixtures; c) demixed sources; d) classification labels.

Fig.2 shows the synthetic image mixtures, demixed sources and the label fields.

The real image for test was scanned from a duplex printed paper, where show-through causes the superimposition of text. The separation result is shown in Fig.3.
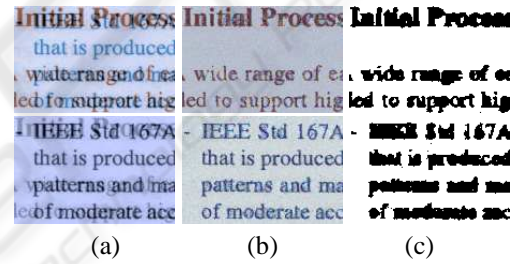

Figure 3: Separation of real show-through image mixtures: a) image mixtures; b) demixed sources; c) classification labels.

For comparison, we also employed the FastICA algorithm (Hyvarinen, 1999) on the sample images with typical parameter set. The results on the show-through examples of Fig.3 are shown in Fig.4. All three channels of the two observed mixtures were used as inputs simultaneously to the ICA algorithm. The two demixed sources can be found in two of six independent components (IC) outputed, while the other four output ICs usually contain unintended noise-like signals, which, along with the permutability property of the ICA algorithm, bring difficulties to reconstructing color representation of the sources. On the other hand, when less color channels are exploited in demixing, we noticed that the separation result does not necessarily degrade or improve, owing to the possible presence of cross-channel correlations.

The MCMC computation involved in the proposed Bayesian separation method is time-consuming. For the example image of 300x240 pixels in Fig.3, which is small relative to ordinary document sizes and resolutions, the typical computation time of the experi-

(a) two ICs corresponding to the demixed sources
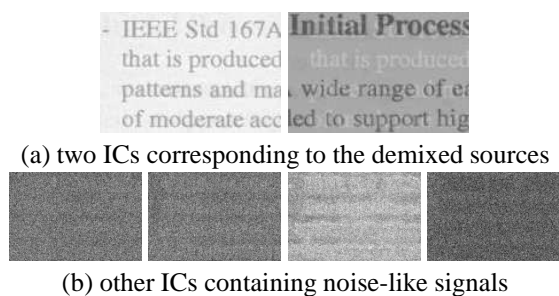

(b) other ICs containing noise-like signals

Figure 4: Separation results by ICA.

mental implementation can come up to hours without specific optimizations. However, various computing alternatives such as Mean Field and variational approximation can be exploited to achieve higher efficiency.

5 CONCLUSION

We proposed a Bayesian approach for separating noisy linear mixture of document images. For source images, we considered a hierarchical model with the hidden label variable $\mathbf{z}$ representing the common classification of objects among multiple color channels, and a Potts-Markov prior was employed for the class labels imposing local regularity constraints. We showed how Bayesian estimation of all unknowns of interest can be computed by MCMC sampling from their posterior distributions given the observation. We then illustrated the feasibility of the proposed algorithm on joint separation and segmentation by tests on sample images.

REFERENCES

Almeida, L. B. (2005). Separating a real-life nonlinear image mixture. *Journal of Machine Learning Research*, 6:1199–1232.

Bronstein, A. M., Bronstein, M. M., Zibulevsky, M., and Zeevi, Y. Y. (2005). Sparse ICA for blind separation of transmitted and reflected images. *Intl. Journal of Imaging Science and Technology (IJIST)*, 15:84–91.

Calhoun, V. D. and Adali, T. (2006). Unmixing fMRI with independent component analysis. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):79–90.

Castella, M. and Pesquet, J.-C. (2004). An iterative blind source separation method for convolutive mixtures of images. *Lecture Notes in Computer Science*, 3195/2004:922–929.

Drira, F. (2006). Towards restoring historic documents degraded over time. In *Second International Con-*
ference on Document Image Analysis for Libraries, pages 350–357.

Feron, O. and Mohammad-Djafari, A. (2005). Image fusion and unsupervised joint segmentation using HMM and MCMC algorithms. *Journal of Electronic Imaging*, 14(2).

Harmeling, S. (2003). Kernel-based nonlinear blind source separation. *Neural Computation*, 15(5):1089–1124.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634.

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons, Inc., New York.

Macias-Macias, M., Garcia-Orellana, C. J., Gonzalez-Velasco, H., and Gallardo-Caballero, R. (2003). Independent component analysis for cloud screening of meteosat images. In *International Work-conference on Artificial and Natural Neural Networks (LNCS 2687/2003)*, volume 2687, pages 551–558.

Parra, L., Spence, C., Ziehe, A., Muller, K.-R., and Sajda, P. (2000). Unmixing hyperspectral data. In *Advances in Neural Information Processing Systems*, volume 12, pages 942–948.

Sharma, G. (2001). Show-through cancellation in scans of duplex printed documents. *IEEE Transactions on Image Processing*, 10(5):736–754.

Snoussi, H. and Calhoun, V. D. (2005). Bayesian blind source separation for brain imaging. In *IEEE International Conference on Image Processing (ICIP) 2005*, volume 3, pages 581–584.

Snoussi, H. and Mohammad-Djafari, A. (2004). Fast joint separation and segmentation of mixed images. *Journal of Electronic Imaging*, 13:349–361.

Tonazzini, A., Bedini, L., and Salerno, E. (2006). A markov model for blind image separation by a mean-field EM algorithm. *IEEE Transactions on Image Processing*, 15(2):473–482.

Tonazzini, A., Salerno, E., Mochi, M., and Bedini, L. (2004). Blind source separation techniques for detecting hidden texts and textures in document images. In *ICIAR 2004, LNCS 3212*, pages 241–248, Berlin. Springer-Verlag.