

PAIRWISE MARKOV MODEL APPLIED TO UNSUPERVISED IMAGE SEPARATION

Selwa Rafi, Marc Castella and Wojciech Pieczynski
Département CITI; UMR-CNRS 5157
Institut Telecom; Telecom SudParis
9 rue Charles Fourier
91011 Evry cedex, France

email: selwa.rafi@it-sudparis.eu, marc.castella@it-sudparis.eu, wojciech.pieczynski@it-sudparis.eu

ABSTRACT

The paper deals with blind separation and recovery of a noisy mixture of two binary signals on two sensors. Such a model can be applied in the context of recovery of scanned documents subject to show-through and bleed-through effects. The problem can be considered as a blind source separation one. Due to a complex noise and data structure, it is tackled from the more general approach of Bayesian restoration. The data is assumed to follow a Pairwise Markov Chain model: it generalizes Hidden Markov Chain models but it still allows one to calculate the *a posteriori* distributions of the data. The Expectation-Maximization (EM) and Iterative Conditional Estimation (ICE) methods are considered for parameter estimation, yielding an unsupervised processing. Finally, simulations show the interest of our approach on simulated and real data.

KEY WORDS

Pairwise Markov Chain, blind source separation, image separation, show-through removal.

1 Introduction

In the last years, Blind Source Separation (BSS) has been an active research area [2]. Independently, Pairwise Markov based models have proved their efficiency for recovering hidden discrete data, with application to image segmentation [4, 9]. Inspired from the problem formulation of recovery of scanned documents subject to show-through and bleed-through effects [10, 15, 11, 14, 16], we make in this paper a link between the two approaches. We also highlight in this context the interest of Pairwise Markov Chain (PMC) models, which have been recently introduced in [4], and we study the performance of the corresponding methods.

The PMC model is an extension of Hidden Markov Chain (HMC) models [6, 7, 13]. In both models, the *a posteriori* probability can be calculated, allowing one to implement a Bayesian restoration or segmentation. The interest and the efficiency of the PMC model have been recently illustrated [4] in the case of a scalar observation; in particular, PMC models are able to deal with more complex noise or signal structure.

In this work, for source signals which have a finite number of states, we recast the BSS problem in the more general framework of Bayesian hidden data restoration. In our context of vector observed data, we apply a Marginal Posterior Mode (MPM) restoration technique based on a PMC model. This requires first a parameter estimation which is performed using either the Expectation-Maximization (EM) [9] or Iterative Conditional Estimation (ICE) estimators [4, 3]. In the case of multidimensional data, PMC models are quite involved and we are not aware that it has been considered so far: our work is thus an extension of [4] in the case of a vector observation process. One can expect that PMC models should help dealing with complicated noise structures.

We especially focus on the separation problem for process vectors whose elements take binary values, black and white images. We also deal with the separation problem of real scanned images. The different models (HMC,PMC) are considered in simulations. We first describe in Section 2 the considered models. Then, estimation and restoration techniques are explained in Section 3. Simulation results are provided in Section 4. Finally, Section 5 concludes the work

2 Models and notations

2.1 Mixture model and hidden variables

In many applications, the observations can be modeled as a mixture of unknown sources. More specifically, a series of T samples $\mathbf{x}_t, t \in \{1, \dots, T\}$ of a vector signal is available, where for any t , $\mathbf{x}_t = (x_t^1, \dots, x_t^Q)^T$ is a Q -dimensional vector. These *observations* result from a mixture of N unknown *source signals*. For any sample time $t \in \{1, \dots, T\}$, the source values are stacked in the vector $\mathbf{s}_t = (s_t^1, \dots, s_t^N)^T$. The objective is to restore the unknown sources only from the observed values $\mathbf{x}_t, t \in \{1, \dots, T\}$. In other words, we want to retrieve for any $i \in \{1, \dots, N\}$ the signal $s_t^i, t \in \{1, \dots, T\}$.

In order to propose a solution to the described problem, the process \mathbf{x}_t should depend on \mathbf{s}_t . We only assume a probabilistic dependence and no specific dependence structure. In particular, our method is applicable to the case

where:

$$\mathbf{x}_t = \mathcal{M}(s_t) + \mathbf{b}_t, \quad t \in \{1, \dots, T\} \quad (1)$$

where $\mathcal{M}(\cdot)$ denotes an unknown, linear or nonlinear function and \mathbf{b}_t is an additive noise. The structure of \mathbf{b}_t may be complicated and in conjunction with $\mathcal{M}(\cdot)$ defines the dependency between s_t and \mathbf{x}_t . A model such as 1 occurs when trying to separate two images obtained from a text scanned document subject to ink bleed-through effect.

Finally, we will assume that each component of s_t belongs to a finite set and hence the vector s_t also belongs to a finite set denoted by $\Omega \triangleq \{\omega_1, \dots, \omega_K\}$. This occurs when separating only black/white (or gray level) mixtures of images.

In this context, restoring the value of the source vector s_t is actually equivalent to determining to which class in Ω it belongs. We will hence tackle the problem of source separation similarly to the problem of segmentation. In our work, the parameter models are unknown and need to be estimated. Consequently, the problem is said to be *blind* or *unsupervised*.

2.2 Hidden variables models

2.2.1 Temporally iid variables and HMC models

We denote by $\mathbf{S} \triangleq (s_1, \dots, s_T)$ (resp. $\mathbf{X} \triangleq (\mathbf{x}_1, \dots, \mathbf{x}_T)$) the set of all samples of the hidden source process (resp. the observation process). The variables (\mathbf{X}, \mathbf{S}) are generally described by a distribution such that $p(\mathbf{X}, \mathbf{S}) = p(\mathbf{S})p(\mathbf{X} | \mathbf{S})$, where $p(\cdot)$ stands for the probability distribution and $p(\cdot | \cdot)$ for the conditional probability. The following assumptions are among the most common ones:

A1 $p(\mathbf{X} | \mathbf{S}) = \prod_{t=1}^T p(\mathbf{x}_t | s_t)$;

A2 The vector process s_t is a stationary Markov process, that is: $p(\mathbf{S}) = p(s_1) \prod_{t=1}^{T-1} p(s_{t+1} | s_t)$ and $p(s_t, s_{t+1})$ does not depend on t .

When assumptions A1 and A2 hold, the model is referred to as a Hidden Markov Chain model (see Fig. 1(a)) and the distribution reads:

$$p(\mathbf{S}, \mathbf{X}) = p(s_1) \prod_{t=1}^{T-1} p(s_{t+1} | s_t) \prod_{t=1}^T p(\mathbf{x}_t | s_t)$$

One can see that the situation in Equation (1) in Section 2.1 is also described by the above model as soon as s_t and \mathbf{b}_t are mutually independent processes and s_t is a Markov process : in this case, the conditional density $p(\mathbf{x}_t | s_t)$ is given by the density of the noise \mathbf{b}_t . An even simpler situation is obtained in the particular case where assumption A2 is strengthened and the vectors s_t are temporally independent and identically distributed (iid). This case is however very rare in applications and it is generally more realistic to introduce some temporal dependence.

2.2.2 PMC models

It turns out that HMC models present some weaknesses in some situations. For example, \mathbf{S} does not necessarily follow a Markov process or the additive noise \mathbf{b}_t in Section 2.1 may depend on the whole process \mathbf{S} . In both cases, the HMC model is no longer valid and does not describe correctly the data. In the following we propose to use a PMC model [4] which generalizes the HMC model.

In the PMC model, we consider the process $\mathbf{z}_t \triangleq (\mathbf{x}_t, s_t)$ and we assume that $\mathbf{z}_t, t \in \{1, \dots, T\}$ is a Markov Chain (see Fig. 1(b)). This model is strictly more general than HMC [4] because the hidden process s_t is not necessarily Markov. In a PMC, s_t is indeed the marginal of a Markov process. The distribution of $\mathbf{Z} \triangleq (\mathbf{z}_1, \dots, \mathbf{z}_T)$ is given by $p(\mathbf{Z}) = p(\mathbf{z}_1) \prod_{t=1}^{T-1} p(\mathbf{z}_{t+1} | \mathbf{z}_t)$. We will consider only stationary reversible PMC, in which case $p(\mathbf{z}_t, \mathbf{z}_{t+1})$ does not depend on t and the distribution of \mathbf{Z} is entirely defined by:

$$p(\mathbf{z}_t, \mathbf{z}_{t+1}) = p(s_t, s_{t+1})p(\mathbf{x}_t, \mathbf{x}_{t+1} | s_t, s_{t+1}) \quad (2)$$

Note that the PMC model becomes a HMC if the following holds [8]:

$$p(\mathbf{x}_t, \mathbf{x}_{t+1} | s_t, s_{t+1}) = p(\mathbf{x}_t | s_t)p(\mathbf{x}_{t+1} | s_{t+1})$$

2.2.3 Gaussian model and PMC with independent noise

Until now, we have not specified the distribution of the observations conditionally on the hidden process. Gaussian PMC is a simple possibility, in which the conditional distributions $p(\mathbf{x}_t, \mathbf{x}_{t+1} | s_t, s_{t+1})$ are Gaussian. Since the distribution of \mathbf{Z} is entirely given by the distributions in (2), it is characterized by K^2 probability parameters of $p(s_t, s_{t+1})$ and K^2 joint conditional probability densities $p(\mathbf{x}_t, \mathbf{x}_{t+1} | s_t, s_{t+1})$.

As \mathbf{x}_t is a vector in this paper, the number of parameters required to characterize each density $p(\mathbf{x}_t, \mathbf{x}_{t+1} | s_t, s_{t+1})$ may become rapidly large. Since there are K^2 such densities, the problem may become intractable. For simplification we consider the specific model of “independent noise PMC” (see Fig.1(c)) which satisfies:

$$p(\mathbf{x}_t, \mathbf{x}_{t+1} | s_t, s_{t+1}) = p(\mathbf{x}_t | s_t, s_{t+1})p(\mathbf{x}_{t+1} | s_t, s_{t+1}) \quad (3)$$

As shown in [12], in an “independent noise PMC”, \mathbf{S} is not a Markov chain. Hence such a PMC is not a HMC.

We will assume that both $p(\mathbf{x}_t | s_t, s_{t+1})$ and $p(\mathbf{x}_{t+1} | s_t, s_{t+1})$ are Gaussian. When \mathbf{x}_t is of dimension $Q = 2$, $p(\mathbf{x}_t | s_t, s_{t+1})$ and $p(\mathbf{x}_{t+1} | s_t, s_{t+1})$ are defined respectively by the parameters $(\mu_1^{i,j}, \sigma_1^{i,j})$ and $(\mu_2^{i,j}, \sigma_2^{i,j})$. In the following we consider the bidimensional case ($Q = 2$).

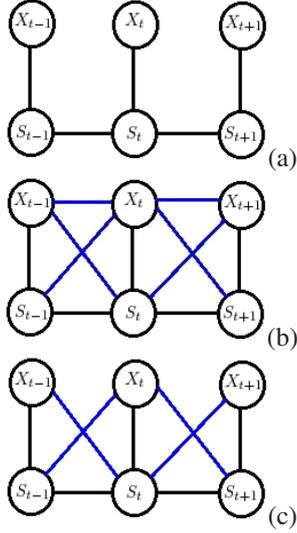


Figure 1. Graphical dependence scheme of: (a) HMC model, (b) PMC model, (c) Independent noise PMC model.

3 Methods

3.1 Bayesian restoration

3.1.1 Maximum Posterior Mode (MPM)

The hidden vector process \mathbf{S} is recovered using the MPM classification method which is based on selecting the class that maximizes the marginal posterior probability. More precisely, the decision rule is:

$$\{\hat{\mathbf{s}}_t = \omega_i\} \iff \{p(\mathbf{s}_t = \omega_i | \mathbf{X}) = \max_j p(\mathbf{s}_t = \omega_j | \mathbf{X})\}$$

3.1.2 Forward-backward procedure

The above restoration method requires to calculate the marginal posterior probability. Let us introduce the following coefficients:

$$\alpha_t(i) = p(\mathbf{s}_t = \omega_i | \mathbf{x}_1, \dots, \mathbf{x}_t)$$

$$\beta_t(i) = \frac{p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T | \mathbf{s}_t = \omega_i, \mathbf{x}_T)}{p(\mathbf{x}_{t+1}, \dots, \mathbf{x}_T | \mathbf{x}_1, \dots, \mathbf{x}_T)}$$

We have then:

$$p(\mathbf{s}_t = \omega_i | \mathbf{X}) = \alpha_t(i)\beta_t(i) \quad (4)$$

For both HMC and PMC models the coefficient $\alpha_t(i)$ and $\beta_t(i)$ can be calculated exactly with an efficient forward-backward procedure [4, 5]. No numerical approximation or integration is required to calculate the marginal posterior probability.

3.2 Parameter estimation

As we consider the unsupervised context, we now address the problem of parameter estimation. For each class transi-

tion (ω_i, ω_j) we calculated 2 mean vectors of size $(Q \times 1)$ and 2 covariance matrices of size $(Q \times Q)$.

3.2.1 Expectation-Maximization (EM)

A well known parameter estimation method is the iterative Expectation-Maximization algorithm, which can also be applied to PMC [9]. After parameter initialization, each loop of the EM algorithm consists of the two following steps:

- "Expectation" step : we calculate $\psi_t(i, j) = p(\mathbf{s}_t = \omega_i, \mathbf{s}_{t+1} = \omega_j | \mathbf{X})$ the joint probability of being at time t in the class ω_i and at time $(t + 1)$ in the class ω_j knowing the observations.

$$\psi_t(i, j) =$$

$$\frac{\alpha_t(i)p(\mathbf{s}_{t+1} = \omega_j, \mathbf{x}_{t+1} | \mathbf{s}_t = \omega_i, \mathbf{x}_t)\beta_t(i)}{\sum_{(\omega_1, \omega_2) \in \Omega^2} \alpha_t(\omega_1)p(\mathbf{s}_{t+1} = \omega_2, \mathbf{x}_{t+1} | \mathbf{s}_t = \omega_1, \mathbf{x}_t)\beta_t(\omega_2)}$$

- "Maximization" step : we estimate and update the K^2 probability parameters of $p(\mathbf{s}_t, \mathbf{s}_{t+1})$ and K^2 bi-dimensional parameters defining (3) using $\psi_t(i, j)$ as follows:

$$p(\mathbf{s}_t = \omega_i, \mathbf{s}_{t+1} = \omega_j) \leftarrow \frac{1}{T-1} \sum_{t=1}^{T-1} \psi_t(i, j)$$

$$\hat{\mu}_1^{i,j} \leftarrow \frac{\sum_{t=1}^{T-1} \psi_t(i, j) \mathbf{x}_t}{\sum_{t=1}^{T-1} \psi_t(i, j)}$$

$$\hat{\mu}_2^{i,j} \leftarrow \frac{\sum_{t=1}^{T-1} \psi_t(i, j) \mathbf{x}_{t+1}}{\sum_{t=1}^{T-1} \psi_t(i, j)}$$

$$\hat{\Gamma}_1^{i,j} \leftarrow \frac{\sum_{t=1}^{T-1} \psi_t(i, j) (\mathbf{x}_t - \hat{\mu}_1^{i,j})(\mathbf{x}_t - \hat{\mu}_1^{i,j})^T}{\sum_{t=1}^{T-1} \psi_t(i, j)}$$

$$\hat{\Gamma}_2^{i,j} \leftarrow \frac{\sum_{t=1}^{T-1} \psi_t(i, j) (\mathbf{x}_{t+1} - \hat{\mu}_2^{i,j})(\mathbf{x}_{t+1} - \hat{\mu}_2^{i,j})^T}{\sum_{t=1}^{T-1} \psi_t(i, j)}$$

3.2.2 Iterative Conditional Estimation (ICE)

Another way to estimate the model parameters consist in using the ICE algorithm which has shown a remarkable flexibility [4]. It is based on the conditional expectation of any estimator valid when the complete data $\mathbf{Z} = (\mathbf{S}, \mathbf{X})$ is available: all parameters are estimated as soon as samples of \mathbf{X} according to $p(\mathbf{X} | \mathbf{S})$ can be simulated. Concretely after parameter initialization, the ICE algorithm consist in successively sampling \mathbf{X} according to $p(\mathbf{X} | \mathbf{S})$ and updating the parameters. $p(\mathbf{s}_t = \omega_i, \mathbf{s}_{t+1} = \omega_j)$ is obtained the same way as in EM and the other parameters are given by:

$$\hat{\mu}_1^{i,j} \leftarrow \frac{1}{\text{Card}(A^{i,j})} \sum_{t=1}^{T-1} \mathbf{1}_{A^{i,j}} \mathbf{x}_t$$

$$\hat{\mu}_2^{i,j} \leftarrow \frac{1}{\text{Card}(A^{i,j})} \sum_{t=1}^{T-1} \mathbf{1}_{A^{i,j}} \mathbf{x}_{t+1}$$

$$\hat{\Gamma}_1^{i,j} \leftarrow \frac{\sum_{t=1}^{T-1} \mathbf{1}_{A^{i,j}} (\mathbf{x}_t - \hat{\mu}_1^{i,j})(\mathbf{x}_t - \hat{\mu}_1^{i,j})^T}{\text{Card}(A^{i,j})}$$

$$\hat{\Gamma}_2^{i,j} \leftarrow \frac{\sum_{t=1}^{T-1} \mathbf{1}_{A^{i,j}} (\mathbf{x}_{t+1} - \hat{\mu}_2^{i,j})(\mathbf{x}_{t+1} - \hat{\mu}_2^{i,j})^T}{\text{Card}(A^{i,j})}$$

where $A^{i,j}$ denotes the set of indices $1 \leq t < T$ for which $(\mathbf{s}_t, \mathbf{s}_{t+1}) = (\omega_i, \omega_j) \in \Omega^2$.

4 Simulation Results

Similarly to the previous section, $Q = 2$ observations are available in our simulations. The vector hidden process \mathbf{s}_t is also of dimension $N = 2$ and it consists of two processes $s_t^i, i \in \{1, 2\}$ which are assumed to take binary values ± 1 . Equivalently, we consider that \mathbf{s}_t belongs to the set $\Omega = \{(\pm 1, \pm 1)\}$ composed of four classes. This experimental setting corresponds to the considered application of bleed-through effect in scanned documents.

4.1 Simulated processes

4.1.1 Data generation

We first tested our method on simulated processes with $T=2000$ samples. The process \mathbf{s}_t was either iid distributed (table 2) or was a Markov chain (table 1, case HMC) with transition parameters:

$$a_{ij} = \begin{pmatrix} 0.8 & 0.1 & 0.05 & 0.05 \\ 0.1 & 0.8 & 0.05 & 0.05 \\ 0.1 & 0.05 & 0.8 & 0.05 \\ 0.1 & 0.05 & 0.05 & 0.8 \end{pmatrix}$$

The two sources have been mixed linearly with matrix $\mathbf{M} = \begin{pmatrix} 0.8 & 0.7 \\ 0.7 & 0.8 \end{pmatrix}$ and a noise has been added. The global model can hence be expressed by the following equation:

$$\mathbf{x}_t = \mathbf{M}\mathbf{s}_t + \mathbf{b}_t \quad (5)$$

The noise \mathbf{b}_t is independent of \mathbf{s}_t and presents either of the following characteristics:

- **iid noise:** In this situation, chains are corrupted with an independent noise (with variance equal to 0.4), the global model obtained follows HMC model.
- **PMC noise:** a more complex situation is to consider a noise following a PMC model in which the dependency between successive observations conditionally on the successive hidden sources $p(\mathbf{x}_t, \mathbf{x}_{t+1} | \mathbf{s}_t, \mathbf{s}_{t+1})$ is given by a Gaussian distribution defined with $2(K^2)$ variance parameters.

4.1.2 Restoration results

The MPM restoration is completely unsupervised: all parameters are unknown and are estimated by the algorithm (either EM or ICE). The simulations results are collected in the following tables :

Method	iid noise (s^1, s^2)	PMC noise (s^1, s^2)
PMC-ICE	(12.3, 11.8)	(29.6, 31.9)
PMC-EM	(11.3, 11.7)	(29.7, 31.3)
HMC	(12.2, 11.3)	(35.6, 38.2)

Table 1. Misclassification rates for Markov Chains in%

For the first column of the table 1, the data follows the HMC model. We can observe that PMC model has approximately the same behaviour as HMC model. However, when the data does no longer follow an HMC model (second column), the PMC algorithm consistently performs better than HMC. This proves that PMC model is more general than the HMC model.

Method	iid noise (s^1, s^2)	PMC noise (s^1, s^2)
PMC-ICE	(22.7, 22.9)	(34.7, 32.4)
PMC-EM	(20.0, 19.4)	(34.6, 32.5)
HMC	(20, 8, 20.4)	(35.5, 36.4)

Table 2. Misclassification rates for iid Chains in %

For the table 2 especially in the second column, the data follows neither HMC model nor the PMC model. The structure of the global process is unknown. The results show that the PMC model is more advantageous than the HMC model.

We deduce from tables 1 and 2 that the PMC model is a generalization of HMC model. The results obtained for PMC with EM and ICE algorithms are comparable.

4.2 Noisy mixture of real images

In this section, we present results regarding the application of Gaussian PMC model in image separation. Similarly to [4, 9, 1] we transform the images into chains using the Hilbert-Peano scanning of the image (see Fig.2).

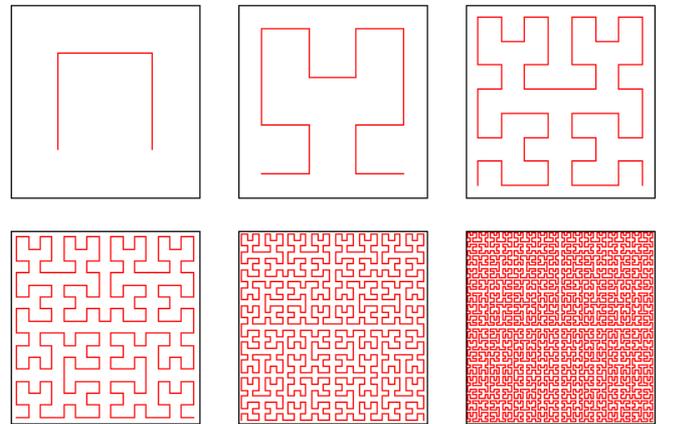


Figure 2. Hilbert-Peano curves and scannings for images of respective size: $2 \times 2, 4 \times 4, 8 \times 8, 16 \times 16, 32 \times 32, 64 \times 64$.

We considered the two black and white images (each color corresponding to $+1$ or -1) showed in Fig.3. Similarly to the previous experiment, they have been mixed according to (5) with the same matrix \mathbf{M} as previously.

We have considered two different noise which are described in Sections 4.2.1 and 4.2.2 respectively. Images are

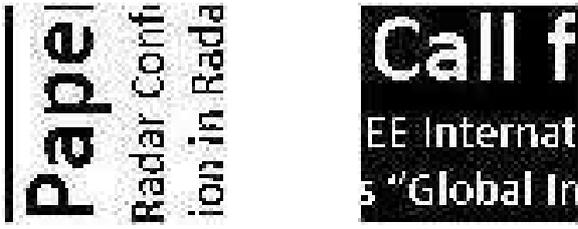


Figure 3. Original images (128×128 pixels)

recovered using MPM restoration and the parameters have been estimated with the ICE algorithm.

4.2.1 Mixture of numerical images affected with noise following a PMC model

In this experiment, we take the images in Fig. 3 that we mix with M and we have added a noise following a Gaussian PMC model as in Section 4.1.1. The observations are in Fig. 4.



Figure 4. Mixed images affected with a PMC noise

The MPM restoration with PMC model on Fig.5(b) gives results better than results obtained with HMC on Fig.5(a). It is due to the inability of HMC model to take into account the dependency between noise and sources.

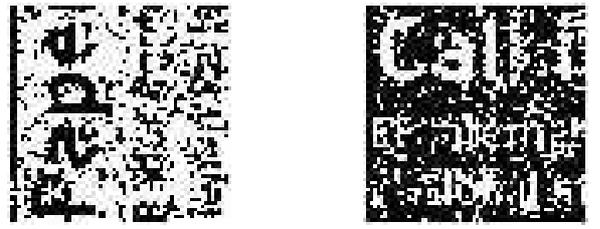
In the following section we consider a correlated noise model which more realistically corresponds to the situation encountered in the show-through and bleed-through effects.

4.2.2 Mixture of numerical images affected with a correlated noise

In this section, we consider the same mixture as the previous experiment. Each noise component is now correlated and has been generated according to:

$$b(r) = \frac{1}{1+4a} [\varepsilon(r) + a \sum_{i=1}^4 \varepsilon(r_i)]$$

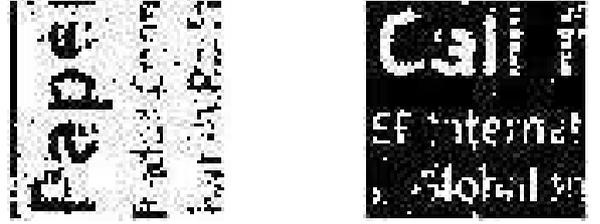
where r denotes an image pixel, $\varepsilon(r)$ is an iid image Gaussian noise, $\varepsilon(r_i)$ $i = 1, \dots, 4$ are the noise value at four neighbors of r in the image and a is a given parameter.



17.99%

(a)

18.06%



7.32%

(b)

8.76%

Figure 5. Separated images and missclassification rate with (a) MPM based on HMC and ICE, (b) MPM based on PMC and ICE. The misclassification rate is given by the percentage of the restored image pixels that are different from the original image pixels.

The hidden data has been recovered using MPM method with HMC and PMC model. The results are respectively represented in Fig. 7(a) and Fig. 7(b).

The MPM restoration with the PMC algorithm (Fig. 7(b)) is better than the restoration based on HMC algorithm (Fig. 7(a)). Like in the case of simulated chains, the algorithm based on PMC shows a better performance in image separation than the method based on HMC model. It is due to its capacity to consider the characteristic of all elements on the mixture and it proves that PMC are more general than HMC.



Figure 6. Mixed images affected with a correlated noise, $a = 0.7$

4.3 Real scanned images with show-through effect

In this section we consider the problem of show-through and bleed-through effects in scanned document [10, 15, 11, 14, 16]. In this situation, we propose to apply our algorithm which allows a good separation automatically without any

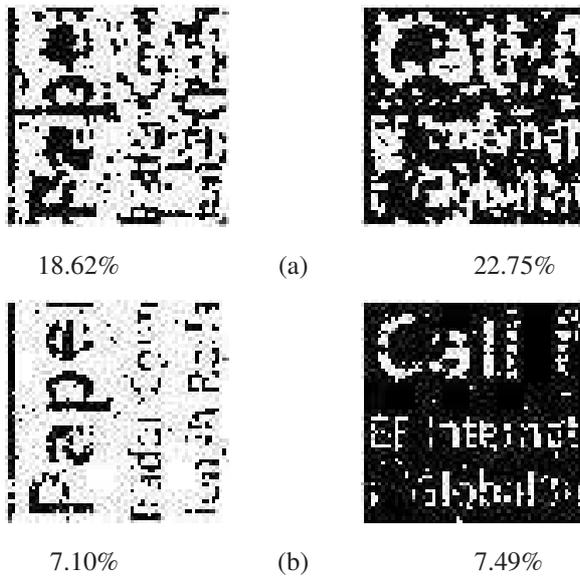


Figure 7. Separated images and missclassification rate with (a) MPM based on HMC and ICE, (b) MPM based on PMC and ICE.

manual intervention. We compare the performance of the PMC algorithm with respect to HMC model. In both algorithms (PMC and HMC) the estimation of the parameters is done using ICE. Fig. 8 shows the recto and verso scan of the same real document, the show-through phenomenon is clearly observed. Fig. 9(a) represents the images restored by HMC model, and in Fig. 9(b) the images are obtained by restoration based on the PMC model.

From these figures, we can conclude that the HMC model is not able to separate correctly the handwriting at each side in the document contrary to PMC model which yielding interesting results. Further investigations on how to apply PMC for the bleed-through problem seem promising.

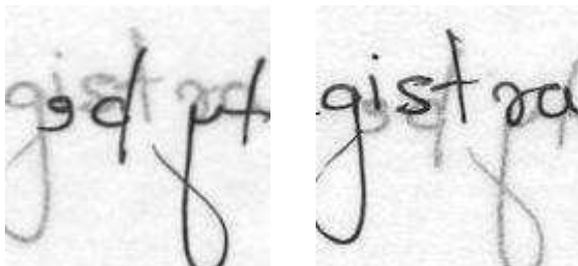


Figure 8. Real scanned document with show-through effect (from://www.site.uottawa.ca/ edubois/documents).

5 Conclusion

In this paper we have presented the application of the recent PMC Model to signal and image separation. The main

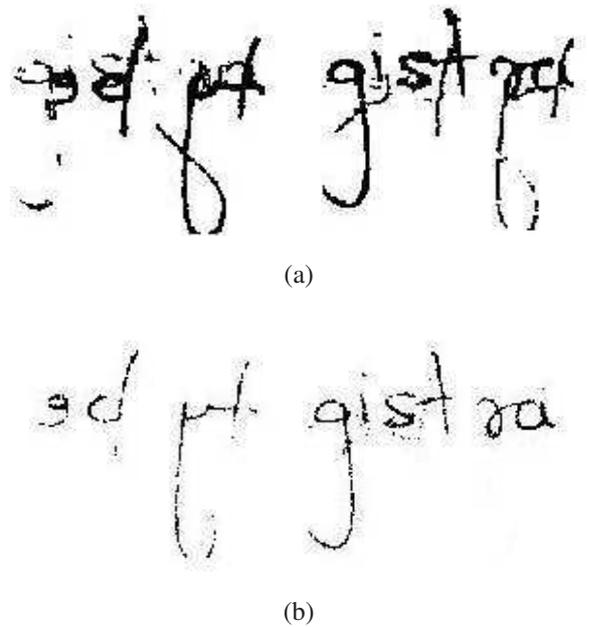


Figure 9. Images separated with (a) MPM based on HMC and ICE, (b) MPM based on PMC and ICE.

contribution was to extend PMC algorithm to multidimensional observed data. The PMC model is richer and more general than the classical HMC model and allows one to take into account more complicated noise structures. We have first validated the performance of the proposed algorithm in the case of simulated iid sources and Markov chains sources. We have applied our method to image separation for both synthetic and real images. We have used Bayesian restoration techniques in the context of PMC, and then we have illustrated that PMC models are more efficient than HMC models. We have shown that using the proposed method, the problem of source separation can be solved even in delicate situations.

References

- [1] Btissam Benmiloud and Wojciech Pieczynski. Estimation des paramètres dans les chaînes de markov cachées et segmentation d'images. *Traitement du signal*, 12.
- [2] Pierre Comon and Christian Jutten. *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. 2010.
- [3] Jean Pierre Delmas. An equivalence of the EM and ICE algorithm for exponential family. *IEEE Trans.Signal Processing*, 45:2612–2615, Oct. 1997.
- [4] Stéphane Derrode and Wojciech Pieczynski. Signal and image segmentation using pairwise markov.

IEEE Transaction on Signal Processing, 52(9), Sept. 2004.

- [5] Pierre A. Devijver. Baum's forward-backward algorithm revisited. *Pattern Recognition Letters*, 3:369–373, 1985.
- [6] Yariv Ephraim and Neri Merhav. Hidden markov processes. *IEEE Trans. Inform. Theory*, 48:1518–1569.
- [7] Timo Koski. Hidden markov models for bioinformatics. *Kluwer Academic Publishers*, 2001.
- [8] Pierre Lanchantin, Jérôme Lapuyade-Lahorgue, and Wojciech Pieczynski. Unsupervised segmentation of randomly switching data hidden with non-gaussian correlated noise. *Signal Processing*, 2(91):163–175.
- [9] Pierre Lanchantin and Wojciech Pieczynski. Unsupervised non stationary image segmentation using triplet markov chains. *Advanced Concepts for Intelligent Vision Systems (ACVIS.04)*, Sept. 2004.
- [10] Farnood Merrikh-Bayat, Massoud Babaie-Zadeh, and Christian Jutten. A nonlinear source separation solution for removing show-through effect in the scanned documents. *16th European Signal Processing Conference EUSIPCO.08*, Aug. 2008.
- [11] Boaz Ophir and David Mallah. Show-through cancellation in scanned images using blind source separation techniques. *IEEE Int. Conf. on Image Processing ICIP-07*, 3:233–236, 2007.
- [12] Wojciech Pieczynski. Multisensor triplet markov chains and theory of evidence. *International Journal of Approximate Reasoning*, 45(1):1–16, 2007.
- [13] Lawrence R. Rabiner. A tutorial on hidden markov models and selected application in speech recognition. *Proc. IEEE*, 77:257–286, Feb. 1989.
- [14] Gaurav Sharma. Cancellation of show-through in duplex scanning. *Proc. IEEE Int. Conf. Image Processing*, 2:609–612, Sept. 2000.
- [15] Anna Tonazzini, Emanuele Salerno, Matteo Mochi, and Luigi Bedini. Bleed-through removal from degraded documents using a colour decorrelation method. *Proc. Document Analysis Systems VI: 6th International Workshop, Springer-Verlag GmbH, LNCS*, 3163:229–240, 2004.
- [16] Christian Wolf. Document ink bleed-through removal with two hidden markov random fields and a single observation field. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(3):431–447, 2010.