International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:5, No:12, 2011

Unsupervised segmentation by hidden Markov chain with bi-dimensional observed process

Abdelali Joumad and Abdelaziz Nasroallah

Abstract—In unsupervised segmentation context, we propose a bi-dimensional hidden Markov chain model $(X, \mathbf{Y})$ that we adapt to the image segmentation problem. The bi-dimensional observed process $\mathbf{Y} = (Y^1, Y^2)$ is such that Y^1 represents the noisy image and Y^2 represents a noisy supplementary information on the image, for example a noisy proportion of pixels of the same type in a neighborhood of the current pixel. The proposed model can be seen as a competitive alternative to the Hilbert-Peano scan. We propose a bayesian algorithm to estimate parameters of the considered model. The performance of this algorithm is globally favorable, compared to the bi-dimensional EM algorithm through numerical and visual data.

Keywords—Image segmentation, Hidden Markov chain with a bi-dimensional observed process, Peano-Hilbert scan, Bayesian approach, MCMC methods, Bi-dimensional EM algorithm.

I. Introduction

HIDDEN Markov models (HMM) are frequently used to study one unobserved variable and uni-dimensional observed process such as hidden Markov fields [13], [28], [8], [5], [30], [40], [10], [27], [35], [17], hidden Markov chain [7], [13], [5], [22], [30], [12], [36], pairwise hidden Markov model [32] or more unobserved variable and uni-dimensional observed process such as triplet hidden Markov model [11]. The multivariate observed process in HMM is used to study unsupervised multi-component images segmentation. Such vectorial image can be obtained for example, from different channels, from several sensors or from image taken at various moments. There are several interesting publications, like Spherically Invariant Random Vector models (SIRV) and copula [15], [21], [24], [33], [34], that treat the multi-dimensional case [22], [24], [23], [31]. One of the main problems in statistical image segmentation using HMM is the adequacy between temporal and spatial context of pixels during the image modeling. This difficulty is usually solved by Markov fields, or by Hilbert-Peano scan in Markov chain case [7], [23]. We propose a hidden Markov chain with a bi-dimensional observed process as another alternative to study this problem. The cited works above concerned with multi-component images use an observed process $\mathbf{Y} = (Y^1, \ldots, Y^M)$ where Hilbert-Peano scan is used to study each image Y^i. Our proposed model consider one noisy image coupled with a noisy supplementary information on the same image (*i.e.* Y^1 represents the noisy image and Y^i for $i = 2, \ldots, M$ represents a noisy supplementary information on Y^1). For simplicity we suppose that $M = 2$ and Y^2 is a noisy proportion of pixels of the same type in a neighborhood of current pixel (*i.e.* Y^2 constructed after calculation of proportion on image and noisy them). The proposed model avoids the use of Hilbert-Peano scan. Also, we propose a bayesian algorithm to estimate parameters of bi-dimensional hidden Markov chain (BHMC) model. We denote this algorithm by (BBHMC). The performance of BBHMC is compared to a bi-dimensional EM algorithm (BEM) through a sample of simulation and visual data.

A. Joumad and A. Nasroallah are with the Department of mathematics, Cadi Ayyad University, Faculty of Sciences Semlalia, B. P. 2390, Marrakesh Morocco e-mails: a.joumad@ucam.ac.ma and nasroallah@ucam.ac.ma

The paper is organized as follows:
In section II, the proposed hidden Markov model BHMC is briefly presented with some necessary hypothesis. Section III is devoted to the BEM algorithm. Section IV is concerned with the parameter estimation problem and the main steps of BBHMC algorithm. In section V, we give an interesting sample of simulation examples that show performance of BBHMC and BEM algorithms: parameter estimations, error rates and CPU times are computed for different scenarios. The efficiency of these two algorithms is examined on a sample of synthetic images in section VI. In the last section we propose a conclusion. A list of references is given in the end.

II. Bi-dimensional hidden Markov model

Consider a random process $(X, \mathbf{Y}) = \{(X_t, \mathbf{Y}_t) \in \Omega \times I\!R^2; t \in IN\}$ defined on a probability space $(\mathcal{X}, \mathcal{F}, P)$, where $\Omega = \{\omega_1, \ldots, \omega_K\}$ is the state space of Markov process X and $K \geq 2$ is a number of classes. The process X is completely characterized by transition parameters $c_{ij} = P(X_t = \omega_i, X_{t+1} = \omega_j)$ for $i, j = 1, \ldots, K$: the initial law $\nu = (\nu_i, 1 \leq i \leq K)$ and the transition matrix $A = (a_{ij})_{1 \leq i,j \leq K}$ are such that

$$\nu_i = P(X_{t_1} = \omega_i) = \sum_{1 \leq j \leq K} c_{ij} \quad \text{and} \quad a_{ij} = \frac{c_{ij}}{\sum\limits_{1 \leq j \leq K} c_{ij}}, \tag{1}$$

where t_1 indicates the initial time of Markov chain X that is supposed homogenous. Furthermore we suppose the following hypothesis:

(H_1) : $\forall (t, t') \in \mathbf{IN^2}$ such that $t \neq t'$, the conditional processes $\mathbf{Y}_t|X$ and $\mathbf{Y}_{t'}|X$ are independent.

(H_2) : $\forall t \in \mathbf{IN}$, $\mathbf{Y}_t|X$ and $\mathbf{Y}_t|X_t$ have the same gaussian distribution.

So $\forall t \in \mathbf{IN}$, $\mathbf{Y}_t|X_t = \omega_i \sim \mathcal{N}(\mu_i, \Sigma_i)$, where the density function is given by

$$f_i(\mathbf{y}_t) = \frac{1}{2\pi|\Sigma_i|^{\frac{1}{2}}} exp\left(-\frac{1}{2}(\mathbf{y}_t - \mu_i)' \Sigma_i^{-1} (\mathbf{y}_t - \mu_i)\right),$$

with $\mu_i = (\mu_i^1, \mu_i^2)'$ and $\Sigma_i = \begin{pmatrix} \sigma_{11}^i & \sigma_{12}^i \\ \sigma_{21}^i & \sigma_{22}^i \end{pmatrix}$, $i = 1, \ldots, K$, are respectively the mean vector and variance-covariance matrix of bi-normal distribution, and $'$ is the transpose operator. The process $(X, \mathbf{Y})$ is called bi-dimensional hidden Markov chain (BHMC), where X is the unobserved process and $\mathbf{Y} = (Y^1, Y^2)$ is the observed one. For a realization of $(X, \mathbf{Y})$

$$(x, \mathbf{y}) = ((\omega_{i_1}, \mathbf{y}_1), \ldots, (\omega_{i_n}, \mathbf{y}_n)),$$

the likelihood is given by

$$L(x, \mathbf{y}; \Theta) = \nu_{i_1} \prod_{m=2}^{n} a_{i_{m-1} i_m} \prod_{m=1}^{n} f_{i_m}(\mathbf{y}_m)$$

where $\Theta = (\nu, A, (\mu_1, \Sigma_1), \ldots, (\mu_K, \Sigma_K))$. The size of Θ is K^2+7K (*i.e.* K for ν, K^2 for A, $2K$ for $(\mu_i)_{i=1,\ldots,K}$ and $4K$ for $(\Sigma_i)_{i=1,\ldots,K}$). Using the fact that each row of A and ν are probability distributions and the symmetry of Σ, this size can be reduced to K^2+5K-1. We suppose that block components $\nu, A, (\mu_i, \Sigma_i), i = 1, \ldots, K$ of Θ are independent.

The classification problem is thus the estimation of the unobserved process realization $X = x$ from the observation $\mathbf{Y} = \mathbf{y}$. The Bayesian Maximum A Posteriori (MAP) and Maximum Posterior Mode (MPM) can be easily computed for the BHMC model. In this paper we will use the MPM method based on minimizing the mean ratio of wrongly classified points. Formally, for each t, the state ω_j that maximizes the a posteriori probability is such that

$$P(X_t = \omega_j \mid \mathbf{Y} = \mathbf{y}) = \max_{1 \le i \le K} P(X_t = \omega_i \mid \mathbf{Y} = \mathbf{y})$$

The computation of the posterior marginal distribution is feasible thanks to the "Forward-Backward" method.

Finally, the problem is: given the bi-dimensional observed process $\mathbf{Y}$, we have to estimate the global parameter Θ.

III. BEM algorithm

Generally, the maximum likelihood estimator has good statistical properties. The Expectation-Maximization (EM) algorithm [2], [26], [37] is one of the most popular method to approximate parameter Θ maximizing the likelihood $L(x, \mathbf{y}; \Theta)$. Given an initial guess $\Theta^{(0)}$, the EM algorithm consists on generating iteratively a sequence $(\Theta^{(q)})_{q \in IN}$ that globally converges thanks to the theorem 10.5.2 in [25]. The two principal steps of EM are:

E-step: Determine $Q(\Theta, \Theta^{(q)}) = E_{\Theta^{(q)}}(\ln(L(x, \mathbf{y}; \Theta))|\mathbf{y})$.

M-step: Choose $\Theta^{(q+1)}$ to be the value that maximizes $Q(\Theta, \Theta^{(q)})$.

To adapt the EM algorithm to the hidden Markov chain with a bi-dimensional observed process, we need the following notations and definitions.

For $t \in IN$ and $\omega_i, \omega_j \in \Omega$, we define: $\psi_t(i,j) := P(X_t = \omega_i, X_{t+1} = \omega_j | \mathbf{Y} = \mathbf{y})$ and $\chi_t(i) := P(X_t = \omega_i | \mathbf{Y} = \mathbf{y})$. These probabilities can be expressed according to the forward-backward probabilities $\alpha_t^*(.)$ and $\beta_t^*(.)$ used in [26] as follows: $\psi_t(i,j) = \frac{\alpha_t^*(i) a_{ij} f_j(\mathbf{y}_{t+1}) \beta_{t+1}^*(j)}{\sum_{l=1}^{K} f_l(\mathbf{y}_{t+1}) \sum_{j=1}^{K} \alpha_t^*(j) a_{jl}}$ and $\chi_t(i) = \sum_{j=1}^{K} \psi_t(i,j)$, where $\alpha_t^*(i) = P(X_t = \omega_i | \mathbf{Y}_1 = \mathbf{y}_1, \ldots, \mathbf{Y}_t = \mathbf{y}_t)$ and $\beta_t^*(i) = \frac{P(\mathbf{Y}_{t+1}=\mathbf{y}_{t+1}, \ldots, \mathbf{Y}_n=\mathbf{y}_n | X_t = \omega_i)}{P(\mathbf{Y}_{t+1}=\mathbf{y}_{t+1}, \ldots, \mathbf{Y}_n=\mathbf{y}_n | \mathbf{Y}_1=\mathbf{y}_1, \ldots, \mathbf{Y}_t=\mathbf{y}_t)}$. The numerator and denominator in $\beta_t^*(i)$ are conditional joint density functions. These notations are usually used for simplicity.

Applying the two steps of EM algorithm to the proposed model BHMC (the obtained algorithm is noted BEM), we obtain the following equations: given estimations at iteration q, those at iteration $(q+1)$ are explicitly as follows:

For $i, j = 1, \ldots, K$,

$$\nu_i^{(q+1)} = \frac{1}{n} \sum_{t=1}^{n} \chi_t^{(q)}(i), \qquad a_{ij}^{(q+1)} = \frac{\sum_{t=1}^{n-1} \psi_t^{(q)}(i,j)}{\sum_{t=1}^{n-1} \chi_t^{(q)}(i)}, \tag{2}$$

$$\mu_i^{(q+1)} = \frac{\sum_{t=1}^{n} \chi_t^{(q)}(i) \mathbf{y}_t}{\sum_{t=1}^{n} \chi_t^{(q)}(i)} \tag{3}$$

and

$$\Sigma_i^{(q+1)} = \frac{\sum_{t=1}^{n} \chi_t^{(q)}(i) (\mathbf{y}_t - \mu_i^{(q+1)})(\mathbf{y}_t - \mu_i^{(q+1)})'}{\sum_{t=1}^{n} \chi_t^{(q)}(i)}. \tag{4}$$

IV. Hidden Markov chain and bayesian approach

As usual in bayesian estimation problems, there are three key issues that have to be addressed: The choice of the prior distribution, the specification of the distribution prior parameters and the evaluation of the posterior distribution.

For our proposed model, we suppose that the rows of the transition matrix associated to the hidden process are statistically independent and following a Dirichlet prior (*i.e.* for $i = 1, \ldots, K$, $(a_{i1}, \ldots, a_{iK}) \sim D_K(\alpha_{i1}, \ldots, \alpha_{iK})$, where α_{ij} are hyper-parameters).

We will use iteratively Metropolis accept-reject and Gibbs sampler algorithms [4], [20], [38] to infer the BHMC parameters. For this aim we consider a sample of r independent trajectories of the unobserved process X that we also note $X = (X_{it_k})_{1 \le i \le r, 1 \le k \le r_i}$ and $\mathbf{Y} = (\mathbf{Y}_{it_k})_{1 \le i \le r, 1 \le k \le r_i}$ are r trajectories of the bi-dimensional observed process $\mathbf{Y}$, where r_i designates the length of the i^{th} trajectory, X_{it_k} is the state of the processus X at time t_k for the trajectory i and $\mathbf{Y}_{it_k}$ is the observed processus at time t_k for the trajectory i.

For simplicity, we denote the conditional distribution at x given all the others and the prior distribution at x by $\pi(x \mid .)$ and $\pi(x)$ respectively.

A. Full conditional distribution for A

Using (1), for $i, j = 1, \ldots, K$ the estimation of a_{ij} can be deduced from those of c_{ij}. So it's being sufficient to determine the full conditional distribution of c_{ij}.
For $u, v = 1, \ldots, K$,

$$\begin{aligned}\pi(c_{uv}|.) &\propto \pi(c_{uv})\pi(\mathbf{Y} \mid X = x, \nu, A, \mu, \Sigma)\\ &\quad \pi(X = x|\nu, A, \mu, \Sigma)\\ &\propto \pi(c_{uv})\pi(X = x|\nu, A),\end{aligned}$$

where $\mu = (\mu_i; 1 \leq i \leq K)$, $\Sigma = (\Sigma_i; 1 \leq i \leq K)$ and the symbol $\propto$ indicates proportionality up to a constant that does not depend on the considered variable.
On the other hand, including the fact that X is an homogenous Markov chain and $x = (x_{ik})_{1\leq i\leq r, 1\leq k\leq r_i}$ is a realization of r independent trajectories of X, we have

$$\begin{aligned}&P(X = x|\nu, A)\\ =\ & \prod_{i=1}^{r} P(X_{it_1} = x_{i1} \mid \nu)\\ & \prod_{k=2}^{r_i} P(X_{it_k} = x_{ik}|X_{it_{k-1}} = x_{i(k-1)}, \nu, A)\\ =\ & \prod_{i=1}^{r} \nu_{x_{i1}} \prod_{k=2}^{r_i} \frac{P(X_{it_{k-1}} = x_{i(k-1)}, X_{it_k} = x_{ik})}{\nu_{x_{i1}}}\\ =\ & \prod_{i=1}^{r} \nu_{x_{i1}}^{2-r_i} \prod_{k=2}^{r_i} P(X_{it_{k-1}} = x_{i(k-1)}, X_{it_k} = x_{ik}).\end{aligned}$$

So

$$\begin{aligned}\pi(c_{uv}|.) &\propto \pi(c_{uv}) \prod_{i=1}^{r} \nu_{x_{i1}}^{2-r_i}\\ &\quad \prod_{k=2}^{r_i} P(X_{it_{k-1}} = x_{i(k-1)}, X_{it_k} = x_{ik}).\end{aligned}$$

To select the prior distribution $\pi(c_{uv})$, $Gamma(\alpha_{uv}, 1)$ is a natural choice since it leads to a Dirichlet distribution supposed for each row of the transition matrix A. The generation of the univariate full conditional distribution for c_{uv} can be obtained by a Metropolis accept-reject algorithm.
Remark.1. Simulation from a Dirichlet distribution $D_K(\delta_1, \ldots, \delta_K)$ can be drawn thanks to the following result: if $\xi_1, \ldots, \xi_K$ are independent with ξ_i having a $Gamma(\delta_i, 1)$ distribution, then

$$\left(\frac{\xi_1}{\sum_{i=1}^{K}\xi_i}, \frac{\xi_2}{\sum_{i=1}^{K}\xi_i}, \ldots, \frac{\xi_K}{\sum_{i=1}^{K}\xi_i}\right) \sim D_K(\delta_1, \ldots, \delta_K)$$

B. Full conditional distribution for ν

The full conditional initial distribution of Markov chain is given by:

$$\pi(\nu|.) \propto \pi(\nu) \prod_{i=1}^{r} \nu_{x_{i1}},$$

which can be written as

$$\pi(\nu|.) \propto \pi(\nu) \prod_{k=1}^{K} \nu_k^{r_k(1)},$$

where $r_k(1) := \sum_{i=1}^{r} \mathbb{1}_{\{x_{i1}=\omega_k\}}$ is the number of visits to state ω_k at initial time ($\mathbb{1}_E$ is the indicator function of set E).
A conjugate prior for $\nu = (\nu_1, \ldots, \nu_K)'$ is a Dirichlet distribution $D_K(\delta_1, \ldots, \delta_K)$ with known hyper-parameters $\delta_k > 0,\ k = 1, \ldots, K$ $\left(i.e.\ \pi(\nu) \propto \prod_{k=1}^{K} \nu_k^{\delta_k - 1}\right)$.
This choice is justified by the fact that $\nu_1, \ldots, \nu_K$ are parameters of a multinomial distribution. So we obtain

$$\pi(\nu|.) \propto \prod_{k=1}^{K} \nu_k^{r_k(1)+\delta_k-1},$$

which leads to

$$\pi(\nu|.) \sim D_K(\delta_1 + r_1(1), \ldots, \delta_K + r_K(1)).$$

C. Full conditional distribution for μ_j and Σ_j

We suppose that Σ_j for $j = 1, \ldots, K$, are independent and μ_j depends only on Σ_j, $j = 1, \ldots, K$, the corresponding prior distributions are chosen to be

$$\mu_j|\Sigma_j \sim \mathcal{N}_2(\tau_j, \Sigma_j) \quad \text{and} \quad \Sigma_j \sim Inv - Wishart(\beta_j, W_j^{-1}),$$

where the hyper-parameters τ_j, β_j and W_j^{-1} are supposed known.
Let's consider the following definitions and notations: $N_i^j := \{k = 1, \ldots, r_i / x_{ik} = \omega_j\}$ is the set of times of visits to state ω_j in trajectory i, $N_j = \sum_{i=1}^{r} \mid N_i^j \mid$, $\overline{\mathbf{y}}_j = \frac{1}{N_j} \sum_{i=1}^{r} \sum_{k\in N_i^j} \mathbf{y}_{ik}$ and $S_j = \sum_i \sum_{k\in N_i^j} (\mathbf{y}_{ik} - \overline{\mathbf{y}}_j)(\mathbf{y}_{ik} - \overline{\mathbf{y}}_j)'$, where $\mid N_i^j \mid$ is the cardinality of the finite set N_i^j and $\mathbf{y} = (\mathbf{y}_{ik})_{1\leq i\leq r, 1\leq k\leq r_i}$ is the given realization of $\mathbf{Y}$.
So, for $j = 1, \ldots, K$,

$$\begin{aligned}\pi(\Sigma_j, \mu_j|.) &\propto \pi(\Sigma_j, \mu_j)\pi(\mathbf{Y}|X = x, \Sigma, \mu)\\ &\propto \pi(\Sigma_j)\pi(\mu_j|\Sigma_j)\pi(\mathbf{Y}|X = x, \Sigma, \mu).\end{aligned}$$

On the other hand, by hypothesis (H_1) and (H_2) in section II, we have

$$\begin{aligned}&\pi(\mathbf{Y}|X = x, \Sigma, \mu)\\ \propto\ & \prod_{i=1}^{r} \prod_{k=1}^{r_i} \pi(\mathbf{Y}_{it_k}|X_{it_k} = x_{ik}, \Sigma_{x_{ik}}, \mu_{x_{ik}})\\ \propto\ & \prod_{i=1}^{r} \prod_{k=1}^{r_i} |\Sigma_{x_{ik}}|^{\frac{-1}{2}} e^{-\frac{1}{2}(\mathbf{Y}_{it_k} - \mu_{x_{ik}})' \Sigma_{x_{ik}}^{-1} (\mathbf{Y}_{it_k} - \mu_{x_{ik}})}.\end{aligned}$$

Therefore, the full conditional distribution of (Σ_j, μ_j) is

given by

$$\begin{aligned}
& \pi(\Sigma_j, \mu_j|.) \\
\propto\ & \pi(\Sigma_j)\pi(\mu_j|\Sigma_j) \\
& \prod_{i=1}^{r}\prod_{k\in\mathbf{N}_i^j} |\Sigma_j|^{\frac{-1}{2}} e^{-\frac{1}{2}(\mathbf{Y}_{it_k}-\mu_j)'\Sigma_j^{-1}(\mathbf{Y}_{it_k}-\mu_j)} \\
\propto\ & \pi(\Sigma_j)\pi(\mu_j|\Sigma_j)|\Sigma_j|^{\frac{-N_j}{2}} \\
& e^{-\frac{1}{2}[tr(\Sigma_j^{-1}S_j)+N_j(\overline{\mathbf{y}}_j-\mu_j)'\Sigma_j^{-1}(\overline{\mathbf{y}}_j-\mu_j)]} \\
\propto\ & |\Sigma_j|^{\frac{-(N_j+\beta_j+3)}{2}} \\
& e^{-\frac{1}{2}[tr(\Sigma_j^{-1}(W_j^{-1}+S_j))+\frac{N_j}{N_j+1}(\overline{\mathbf{y}}_j-\tau_j)(\overline{\mathbf{y}}_j-\tau_j)']} \\
\times\ & e^{-\frac{1}{2}[(\mu_j-\frac{\tau_j+N_j\overline{\mathbf{y}}_j}{N_j+1})'(\frac{\Sigma_j}{N_j+1})^{-1}(\mu_j-\frac{\tau_j+N_j\overline{\mathbf{y}}_j}{N_j+1})]}.
\end{aligned}$$

Thereby

$$\pi(\mu_j|.) \sim \mathcal{N}_2\left(\frac{\tau_j+N_j\overline{\mathbf{y}}_j}{N_j+1}, \frac{\Sigma_j}{N_j+1}\right)$$

and

$$\pi(\Sigma_j|.) \sim Inv-Wishart(\beta_j+N_j, W(S_j,W_j)),$$

with

$$W(S_j,W_j) = W_j^{-1} + S_j + \frac{N_j}{N_j+1}(\overline{\mathbf{y}}_j-\tau_j)(\overline{\mathbf{y}}_j-\tau_j)'.$$

Remark.2. As argued in [9], [25], often there is no compelling reason to choose the previous priors, except for their simplicity, but the restrictive aspect of conjugate priors can be attenuated by using hyper-priors on the hyper-parameters. Those hyper-priors can be chosen amongst so-called non-informative (or vague) priors to attenuate the impact on the resulting inference.
Now, the joint prior distribution for Θ can be written as:

$$\begin{aligned}
& \pi(\Theta) \\
=\ & \pi(\nu_1,\ldots,\nu_K, A, \mu_1,\ldots,\mu_K, \Sigma_1,\ldots,\Sigma_K) \\
=\ & \pi(\nu_1,\ldots,\nu_K)\pi(A)\pi(\Sigma_1,\ldots,\Sigma_K) \\
& \pi(\mu_1,\ldots,\mu_K|\Sigma_1,\ldots,\Sigma_K) \\
=\ & \prod_{j=1}^{K}\pi(\nu_j)\prod_{i=1}^{K}\pi(a_{i1},\ldots,a_{iK}) \\
& \prod_{j=1}^{K}\pi(\mu_j|\Sigma_j)\prod_{j=1}^{K}\pi(\Sigma_j) \\
\propto\ & \prod_{j=1}^{K}\nu_j^{\delta_j-1}\prod_{i,j=1}^{K}a_{ij}^{\alpha_{ij}-1}\prod_{j=1}^{K}|\Sigma_j|^{-\frac{1}{2}} \\
& e^{-\frac{1}{2}(\mu_j-\tau_j)'\Sigma_j^{-1}(\mu_j-\tau_j)} \\
\times\ & \prod_{j=1}^{K}|\Sigma_j|^{-\frac{1}{2}(\beta_j+3)}e^{-\frac{1}{2}\mathbf{tr}(\Sigma_j^{-1}W_j^{-1})}
\end{aligned}$$

This joint prior distribution is obtained using independence of block components of Θ supposed in section II.

D. A bayesian algorithm for BHMC

We propose a bayesian algorithm for bi-dimensional hidden Markov chain model, principally based on the classical Gibbs sampler and Metropolis accept-reject algorithms. For given hyper-parameters $\alpha_{ij}, \tau_j, \delta_j, \beta_j$ and W_j^{-1} ; $i,j = 1,\ldots,K$, the main steps of BBHMC are the following:

The BBHMC Algorithm

E0. Select starting values (iteration $q=0$),
$\Theta^{(q)} = (\nu^{(q)}, a_{ij}^{(q)}, \mu_i^{(q)}, \Sigma_i^{(q)}, i,j = 1,\ldots,K)$
and compute $c_{ij}^{(q)}$ using (1).

E1. Iteration $(q+1)$

E1.1 Generate a sample
$x^{(q)} = \{x_{ik}^{(q)}, i = 1,\ldots,r,\ k = 1,\ldots,r_i\}$

E1.2 Generate a transition matrix
$A^{(q+1)} = (a_{ij}^{(q+1)})_{1\le i,j\le K}$
according to the following accept-reject algorithm:
for every couple (i,j)

E1.2.1 Generate a candidate
$b_{ij} \sim \mathcal{N}(c_{ij}^{(q)}, 1)$ and $u \sim \mathcal{U}_{(0,1)}$
such that b_{ij} and u are independent

E1.2.2 Compute $\prod(b_{ij})$ and $\prod(c_{ij}^{(q)})$ by

$$\prod(b_{ij}) = \prod_{l=1}^{r}\prod_{k=1}^{r_l}[c^{(q)}_{x_{l(k-1)}x_{lk}}\mathbb{1}_{\{(x_{l(k-1)},x_{lk})\neq(i,j)\}} + b_{ij}\mathbb{1}_{\{(x_{l(k-1)},x_{lk})=(i,j)\}}]$$

$$\prod(c_{ij}^{(q)}) = \prod_{l=1}^{r}\prod_{k=1}^{r_l} c^{(q)}_{x_{l(k-1)}x_{lk}}$$

E1.2.3 Compute
$\rho_{ij}^{(q+1)} = \min\left(1, \frac{g(b_{ij})\prod(b_{ij})}{g(c_{ij}^{(q)})\prod(c_{ij}^{(q)})}\right)$
where g is the density function of $Gamma(\alpha_{ij}, 1)$

E1.2.4 Compute
$c = c_{ij}^{(q)}\mathbb{1}_{\left\{u\ge\rho_{ij}^{(q+1)}\right\}} + b_{ij}\mathbb{1}_{\left\{u<\rho_{ij}^{(q+1)}\right\}}$
if $c = c_{ij}^{(q)}$, we reject b_{ij} and return to **E1.2.1**
else we accept b_{ij}: $c_{ij}^{(q+1)} = c$

E1.2.5 Compute $a_{ij}^{(q+1)}$ (by formula in (1))

E2. Compute $N_j, \overline{\mathbf{y}}_j, S_j$ and $W(S_j,W_j)$ for $j = 1,\ldots,K$ (from $x^{(q)}$)

E3. For $j = 1,\ldots,K$,

E3.1 Generate

$$\pi(\mu_j^{(q+1)}|.) \sim \mathcal{N}_2\left(\frac{\tau_j+N_j\overline{\mathbf{y}}_j}{N_j+1}, \frac{\Sigma_j^{(q)}}{N_j+1}\right)$$

E3.2 Generate

$$\pi(\Sigma_j^{(q+1)}|.) \sim Inverse-Wishart(\beta_j+N_j, W(S_j,W_j))$$

E3.3 Generate

$$\pi(\nu^{(q+1)}|.) \sim D_K(\delta_1+r_1(1),\ldots,\delta_K+r_K(1))$$

V. Numerical simulation

We interest to the performance of BEM and BBHMC algorithms by simulating different noisy Markov chains, and we compare the estimations obtained from the noisy data to those computed from the complete data. These last will be called

Empirical Values and will be denoted (EV). The computation of the empirical values c^*_{ij}, μ^*_j and Σ^*_j for $i, j = 1, \ldots, K$ makes itself by the following formulas:

$$c^*_{ij} = \frac{1}{r(r_r - 1)} \sum_{l=1}^{r} \sum_{k=1}^{r_l - 1} \mathbb{1}_{\{(X_{lk}, X_{l(k+1)}) = (w_i, w_j)\}},$$

$$\mu^*_j = \frac{1}{N_j} \sum_i \sum_{k \in N^j_i} \mathbf{y}_{it_k},$$

$$\Sigma^*_j = \frac{1}{N_j} \sum_i^r \sum_{k \in N^j_i} (\mathbf{y}_{it_k} - \mu^*_j)(\mathbf{y}_{it_k} - \mu^*_j)^{'}$$

In the following, we present the parameter initializations and the data scenarios, that has been used by many authors [7], [14], [29], after adapting them to the bi-dimensional case.

A. Parameter initializations

For $i, j = 1, \ldots, K$ and $d, d^{'} = 1, 2$

• Markov chain parameters:

$$\nu_i = \frac{1}{K}, \quad a_{ii} = 0.5 \quad \text{and} \quad a_{ij} = \frac{1}{2(K-1)} \quad \text{for} \quad i \neq j$$

• Noisy parameters:

Mean vector:

If $K = 2p$, for $1 \leq k \leq p$

$$\mu^d_{k+1} = m^d - 0.5(p-k)\sqrt{\sigma_{dd}}$$

$$\mu^d_{K-k} = m^d + 0.5(p-k)\sqrt{\sigma_{dd}}$$

If $K = 2p + 1$, for $1 \leq k \leq K$

$$\mu^d_{k+1} = m^d - 0.5(0.5K - k)\sqrt{\sigma_{dd}},$$

Variance-covariance matrix:

$$\sigma^i_{dd} = \sigma_{dd}, \quad \sigma^i_{dd'} = c_i.\sigma_{dd'} \quad \text{for} \quad d \neq d^{'}$$

where $c_i = 0.1$ for numerical simulation and $c_i = 0.9$ for image segmentation. The mean vector $(m^1, m^2)^{'}$ and the variance-covariance matrix $\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ are computed using a generated trajectory of X.

• Hyper-parameters:

For $1 \leq i, j \leq K$, $\alpha_{ij} = 2$, $\tau_i = (2, 2)^{'}$, $\beta_i = 1$, $\delta_i = 1$, and $W_i^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

B. Markov chain and noisy data scenarios

While considering a sample of size $n = 1000$ and a number of classes $K = 2$, there are two factors of wish we study the influence on the performance of BEM and BBHMC algorithms: the noise and the Markov chain.

• For the noise factor, we consider the following three scenarios:

(MD): means discriminating noise $\mu_1 \neq \mu_2$ and $\Sigma_1 = \Sigma_2$

(VD): variances discriminating noise $\mu_1 = \mu_2$ and $\Sigma_1 \neq \Sigma_2$

(MVD): means and variances discriminating noise $\mu_1 \neq \mu_2$ and $\Sigma_1 \neq \Sigma_2$.

The noise level depends strongly on the following two parameters:

$$\Delta = (\mu^1_2 - \mu^1_1)^2 + (\mu^2_2 - \mu^2_1)^2 \quad \text{and} \quad \rho = \rho^2_2/\rho^2_1,$$

where

$$\rho_2 = \sigma^2_{12}/\sqrt{\sigma^2_{11}\sigma^2_{22}} \quad \text{and} \quad \rho_1 = \sigma^1_{12}/\sqrt{\sigma^1_{11}\sigma^1_{22}}.$$

We fix the parameters values for each scenario as follows:

MD: $\mu_1 = (1, 1)^{'}$, $\mu_2 = (1, 2)^{'}$ and $\Sigma_1 = \Sigma_2 = \begin{pmatrix} 2 & 0.9 \\ 0.9 & 3 \end{pmatrix}$.

These values correspond to $(\Delta, \rho) = (1, 1)$.

VD: $\mu_1 = \mu_2 = (1, 1)^{'}$, $\Sigma_1 = \begin{pmatrix} 0.25 & 0.2 \\ 0.2 & 4 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 5 & 5.3 \\ 5.3 & 9 \end{pmatrix}$.

These values correspond to $(\Delta, \rho) = (0, 1.56)$.

MVD: $\mu_1 = (1, 1.5)^{'}$, $\mu_2 = (2, 1)^{'}$, $\Sigma_1 = \begin{pmatrix} 0.25 & 0.9 \\ 0.9 & 4 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 5 & 3.5 \\ 3.5 & 9 \end{pmatrix}$.

These values correspond to $(\Delta, \rho) = (1.25, 0.30)$.

• For the Markov chain factor, we consider the following two scenarios:

C1: $a_{11} = a_{22} = 0.98$.

C2: $a_{11} = a_{22} = 0.80$.

C. Implementation parameters

There are several parameters which arise when implementing MCMC algorithms [39]. The algorithm must be run for a heating period, called burn-in, which depends on initial values of parameters. The aim of this period is to reduce the effect of initial values on posterior inference. Visual inspection of plots of the algorithm output $(\Theta^{(q)}, q = 0, 1, \ldots, T)$ with T is the global length of iterations, is the most obvious and commonly used method for determining the burn-in period. For the burn-in period length T_0 of BBHMC algorithm, we take $T_0 = 1000$ for numerical simulations and $T_0 = 7000$ for image segmentation. After this period, we start the collection of estimations for $T_1 = 1000$ supplementary iterations. So, the total number of iterations is $T = T_0 + T_1$ and the estimation $\widehat{\Theta}$ of Θ is given by the MMSE estimator $\left(i.e.\ \widehat{\Theta} = \frac{1}{T - T_0 + 1} \sum_{q=T_0}^{T} \Theta^{(q)} \right)$.

For BEM algorithm, we use an automatic stopping criteria based on the following distances (inspired to [7])

$$D^2(\mu^{(q)}) = \frac{1}{K} \sum_{k=1}^{K} c_k [(\mu_k^{1^{(q)}} - \frac{1}{q+1} \sum_{l=0}^{q} \mu_k^{1^{(l)}})^2 + (\mu_k^{2^{(q)}} - \frac{1}{q+1} \sum_{l=0}^{q} \mu_k^{2^{(l)}})^2]$$

$$D^2(\sigma^{(q)}_{ij}) = \frac{1}{K} \sum_{k=1}^{K} c_k \left[\sigma_{ij}^{k^{(q)}} - \frac{1}{q+1} \sum_{l=0}^{q} \sigma_{ij}^{k^{(l)}} \right]; \ 1 \leq i, j \leq 2$$

The stoping of iterations is a function of quality estimation of parameters and stoutness of MPM method. The BEM

algorithm is stopped when $D^2(\mu^{(q)}) < \varepsilon$ and $D^2(\sigma_{ij}^{(q)}) < \varepsilon$ for $i,j = 1,2$. We take $(\varepsilon, c_k) = (10^{-4}, 1)$ for numerical simulations and $(\varepsilon, c_k) = (10^{-2}, 10^{-3})$ for image segmentation.

D. Numerical results

This subsection deals with numerical estimation for parameters and misclassification rates obtained by BEM and BBHMC, compared with those obtained by EV. These algorithms are tested on combinations of **C1** and **C2** with **MD**, **VD** and **MVD**. The BEM and BBHMC algorithms use the same initializations. The obtained estimations are presented in Tables II to V. In Table I, we present the misclassification rates of pixels computed by MPM method.

From Table I, we remark that BBHMC has globally a small advantage with respect to BEM mainly for **C2+MD** and **C2+MVD** scenarios witch are qualified as worst cases.

Concerning estimations of ν and A, resumed in Tables II and III, the BEM algorithm takes advantage with respect to BBHMC algorithm with a stability advantage of estimations for the last. It seems that this stability is due to the estimations form (*i.e.* estimators obtained by the law large number).

Regarding Tables IV and V, the estimations of μ and Σ obtained by BEM and BBHMC are very closed to those obtained by EV.

TABLE I
ERROR RATES FOR EACH CLASS (ER_1 AND ER_2) GIVEN BY EV, BEM AND BBHMC FOR DIFFERENT DATA SCENARIOS, WHERE ER_i IS THE MISCLASSIFICATION RATE OF CLASS i. THE MEAN CPU TIME FOR BEM AND BBHMC IS 7.5

Data	Error rates(%)	EV	BEM	BBHMC
C1+MD	ER_1	2.4	2.4	2.1
	ER_2	3.3	3.3	3.5
C1+VD	ER_1	0.3	0.1	0.1
	ER_2	1.8	1.8	1.8
C1+MVD	ER_1	0.1	1.9	1.9
	ER_2	0.4	0.4	0.4
C2+MD	ER_1	22.6	23.7	22.6
	ER_2	18.5	19.0	18.5
C2+VD	ER_1	6.7	7.1	6.9
	ER_2	12.7	12.9	12.7
C2+MVD	ER_1	4.8	4.2	4.8
	ER_2	12.9	16.0	12.9

TABLE II
ESTIMATIONS OF COMPONENT ν_1 BY EV, BEM AND BBHMC FOR DIFFERENT DATA SCENARIOS. THE STANDARD DEVIATION IS GIVEN BETWEEN BRACKETS FOR BBHMC ESTIMATIONS

Data	EV	BEM	BBHMC
C1+MD	0.522	0.521	0.490 (0.410×10^{-2})
C1+VD	0.522	0.523	0.490 (0.410×10^{-2})
C1+MVD	0.522	0.519	0.490 (0.410×10^{-2})
C2+MD	0.475	0.477	0.493 (0.021×10^{-2})
C2+VD	0.475	0.482	0.493 (0.021×10^{-2})
C2+MVD	0.475	0.501	0.493 (0.021×10^{-2})

TABLE III
ESTIMATIONS OF a_{11} AND a_{22} BY EV, BEM AND BBHMC FOR DIFFERENT DATA SCENARIOS. THE STANDARD DEVIATION IS GIVEN BETWEEN BRACKETS FOR BBHMC ESTIMATIONS

Data		EV	BEM	BBHMC
C1+MD	a_{11}	0.980	0.980	0.977 (0.11×10^{-3})
	a_{22}	0.979	0.976	0.970 (0.29×10^{-3})
C1+VD	a_{11}	0.980	0.977	0.977 (0.11×10^{-3})
	a_{22}	0.979	0.979	0.970 (0.29×10^{-3})
C1+MVD	a_{11}	0.980	0.978	0.977 (0.11×10^{-3})
	a_{22}	0.979	0.979	0.970 (0.29×10^{-3})
C2+MD	a_{11}	0.791	0.821	0.796 (0.33×10^{-3})
	a_{22}	0.809	0.835	0.814 (0.46×10^{-3})
C2+VD	a_{11}	0.791	0.796	0.796 (0.33×10^{-3})
	a_{22}	0.809	0.808	0.814 (0.46×10^{-3})
C2+MVD	a_{11}	0.791	0.800	0.796 (0.33×10^{-3})
	a_{22}	0.809	0.807	0.814 (0.46×10^{-3})

TABLE IV
ESTIMATIONS OF THE TWO COMPONENT OF μ_1 AND μ_2 BY EV, BEM AND BBHMC FOR DIFFERENT DATA SCENARIOS. THE STANDARD DEVIATION IS GIVEN BETWEEN BRACKETS FOR BBHMC ESTIMATIONS

Data		EV	BEM	BBHMC
C1+MD	μ_1	0.948	0.931	0.950 (0.03×10^{-2})
		1.064	1.076	1.066 (0.14×10^{-2})
	μ_2	2.031	2.044	2.027 (0.07×10^{-2})
		3.020	3.003	3.020 (0.06×10^{-2})
C1+VD	μ_1	0.981	0.982	0.983 (0.01×10^{-2})
		1.056	1.056	1.058 (0.16×10^{-2})
	μ_2	2.049	2.046	2.043 (0.11×10^{-2})
		1.990	1.993	1.994 (0.14×10^{-2})
C1+MVD	μ_1	0.981	0.982	0.983 (0.01×10^{-2})
		1.630	1.605	1.631 (0.20×10^{-2})
	μ_2	3.049	3.030	3.041 (0.11×10^{-2})
		2.018	2.044	2.022 (0.12×10^{-2})
C2+MD	μ_1	0.982	1.034	0.984 (0.35×10^{-2})
		1.040	0.908	1.046 (0.15×10^{-2})
	μ_2	1.992	1.946	1.993 (0.22×10^{-2})
		3.046	3.173	3.042 (0.21×10^{-2})
C2+VD	μ_1	0.993	0.993	0.995 (0.12×10^{-2})
		1.044	1.027	1.051 (0.16×10^{-2})
	μ_2	1.988	2.000	1.989 (0.35×10^{-2})
		2.101	2.133	2.097 (0.44×10^{-2})
C2+MVD	μ_1	0.993,	1.007	0.995 (0.12×10^{-2})
		1.550	1.495	1.557 (0.22×10^{-2})
	μ_2	2.988	3.078	2.987 (0.35×10^{-2})
		2.087	2.171	2.083 (0.39×10^{-2})

E. Level noise and misclassification error

We consider different levels (Δ, ρ) of **MD**, **VD** and **MVD** noises. These noises are combined with **C1** and **C2**. For these combinations, the behavior of the misclassification rate, computed by MPM method, using EV is given in Table VI. The performance of BEM and BBHMC through the following three cases $(\Delta, \rho) \in \{(5,1),(0,1),(0,1.63)\}$ is presented in Fig. 1 to Fig. 6.

In each figure, we give the behavior of error rate (for each class) with respect to the iterations. In comparison with EV, we constat that BEM and BBHMC give good results for $(\Delta, \rho) = (5,1)$ (see Fig. 1 and Fig.4): the error rates for BEM and BBHMC are nearer to those given by EV after $T_1 = 1000$ iterations for BBHMC and after 50 iterations for BEM. So BEM outperforms BBHMC in this easiest case (*i.e.* lower level

TABLE V
ESTIMATIONS OF Σ_1 AND Σ_2 BY EV, BEM AND BBHMC FOR DIFFERENT DATA SCENARIOS. THE STANDARD DEVIATION IS GIVEN BETWEEN BRACKETS FOR BBHMC ESTIMATIONS

Data			EV	BEM	BBHMC
C1+MD	Σ_1	σ_{11}	1.990	1.981	1.998 (0.028×10^{-1})
		σ_{22}	3.362	3.409	3.341 (0.019×10^{-1})
		σ_{12}	0.279	0.316	0.279 (0.028×10^{-1})
	Σ_2	σ_{11}	1.992	1.967	2.002 (0.045×10^{-1})
		σ_{22}	3.616	3.632	3.632 (0.076×10^{-1})
		σ_{12}	-0.024	-0.079	-0.021 (0.013×10^{-1})
C1+VD	Σ_1	σ_{11}	0.248	0.245	0.253 (0.004×10^{-1})
		σ_{22}	4.194	4.147	4.166 (0.023×10^{-1})
		σ_{12}	0.111	0.112	0.112 (0.011×10^{-1})
	Σ_2	σ_{11}	4.980	5.000	5.004 (0.112×10^{-1})
		σ_{22}	15.642	15.752	15.698 (0.327×10^{-1})
		σ_{12}	0.031	0.030	0.039 (0.045×10^{-1})
C1+MVD	Σ_1	σ_{11}	0.248	0.244	0.253 (0.003×10^{-1})
		σ_{22}	6.784	6.792	6.756 (0.038×10^{-1})
		σ_{12}	0.123	0.111	0.122 (0.014×10^{-1})
	Σ_2	σ_{11}	4.980	4.990	5.006 (0.112×10^{-1})
		σ_{22}	12.237	12.200	12.282 (0.256×10^{-1})
		σ_{12}	-0.030	-0.063	-0.023 (0.039×10^{-1})
C2+MD	Σ_1	σ_{11}	2.005	2.048	2.015 (0.004×10^{-1})
		σ_{22}	3.507	3.344	3.494 (0.061×10^{-1})
		σ_{12}	0.234	0.266	0.235 (0.024×10^{-1})
	Σ_2	σ_{11}	1.981	2.038	1.992 (0.058×10^{-1})
		σ_{22}	3.463	3.091	3.481 (0.010×10^{-1})
		σ_{12}	0.041	-0.002	0.049 (0.043×10^{-1})
C2+VD	Σ_1	σ_{11}	0.250	0.239	0.255 (0.144×10^{-1})
		σ_{22}	4.309	4.677	4.291 (0.043×10^{-1})
		σ_{12}	0.092	0.094	0.094 (0.140×10^{-1})
	Σ_2	σ_{11}	4.954	5.017	4.979 (0.001×10^{-1})
		σ_{22}	14.601	14.354	14.658 (0.075×10^{-1})
		σ_{12}	0.198	0.168	0.224 (0.009×10^{-1})
C2+MVD	Σ_1	σ_{11}	0.250	0.292	0.255 (0.001×10^{-1})
		σ_{22}	7.298	7.539	7.279 (0.128×10^{-1})
		σ_{12}	0.104	0.055	0.105 (0.013×10^{-1})
	Σ_2	σ_{11}	4.954	5.004	4.980 (0.144×10^{-1})
		σ_{22}	11.583	11.501	11.632 (0.034×10^{-1})
		σ_{12}	0.145	0.033	0.168 (0.125×10^{-1})

noise). In all other cases, BBHMC works better than BEM: the error rates, after stabilization, are nearer to those given by EV for BBHMC than for BEM (see Fig. 2, Fig. 3, Fig. 3 and Fig. 6).

TABLE VI
THEORETICAL ERROR RATES FOR EACH CLASS (ER_1 AND ER_2) FOR **C1** AND **C2**. ER_i IS THE MISCLASSIFICATION ERROR OF CLASS i.

Markov chain	$\Delta=5$ $\rho=1$	$\Delta=2$ $\rho=1$	$\Delta=1$ $\rho=1$	$\Delta=0$ $\rho=1$	$\Delta=0$ $\rho=1.63$	$\Delta=5$ $\rho=1.63$
C1 ER_1	0.5	1.3	2.4	9.3	6.5	2.4
ER_2	1.8	1.8	3.3	11.8	6.9	2.09
C2 ER_1	7.3	12.18	22.6	40.9	38.8	17.8
ER_2	9.7	15.8	18.5	27.09	31.6	18.7

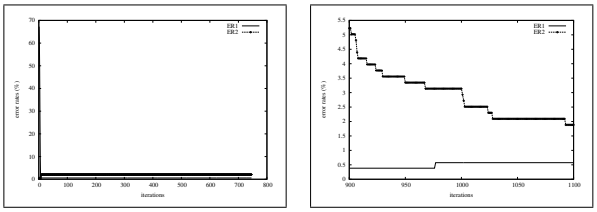

Fig. 1. Behavior of error rates for each class (ER_1 and ER_2) with respect to the iterations for chain **C1** with $(\Delta,\rho)=(5,1)$. The result of BEM (resp. BBHMC) is on the left (resp. right)

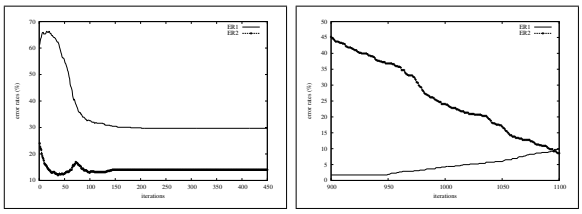

Fig. 2. Behavior of error rates for each class (ER_1 and ER_2) with respect to the iterations for chain **C1** with $(\Delta,\rho)=(0,1)$. The result of BEM (resp. BBHMC) is on the left (resp. right)

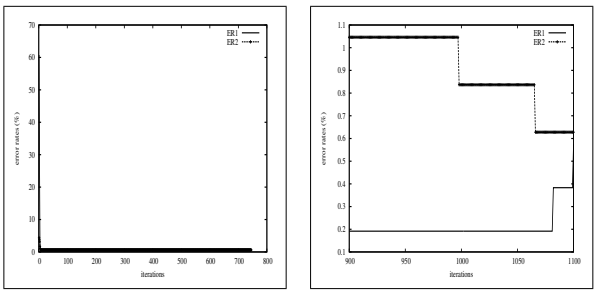

Fig. 3. Behavior of error rates for each class (ER_1 and ER_2) with respect to the iterations for chain **C1** with $(\Delta,\rho)=(5,1.63)$. The result of BEM (resp. BBHMC) is on the left (resp. right)

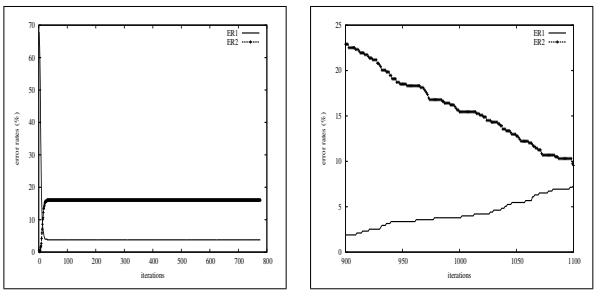

Fig. 4. Behavior of error rates for each class (ER_1 and ER_2) with respect to the iterations for chain **C2** with $(\Delta,\rho)=(5,1)$. The result of BEM (resp. BBHMC) is on the left (resp. right)

VI. UNSUPERVISED IMAGE SEGMENTATION

In the following, we compare BHMC using Hilbert-Peano scan and BHMC using a noisy supplementary information (for example: proportion of pixels in the same class as the current pixel in a neighborhood of the last) on image. We consider that a neighborhood of a pixel is composed of 3, 5 or 8 pixels around it according to the following three positions: corner, border or inside. The comparison is made through a sample of four synthetic images given in Fig. 7 (*Alphabet*, *Gibbs*,

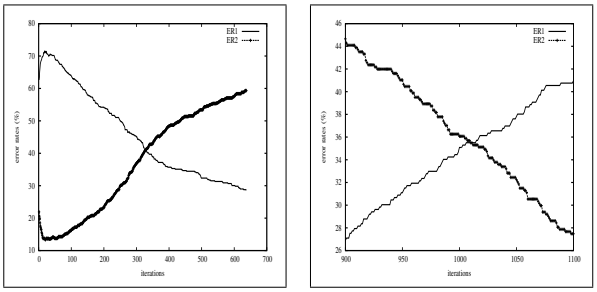

Fig. 5. Behavior of error rates for each class (ER_1 and ER_2) with respect to the iterations for chain **C2** with $(\Delta, \rho) = (0, 1)$. The result of BEM (resp. BBHMC) is on the left (resp. right)

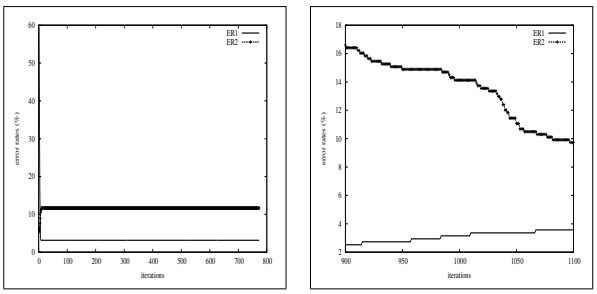

Fig. 6. Behavior of error rates for each class (ER_1 and ER_2) with respect to the iterations for chain **C2** with $(\Delta, \rho) = (5, 1.63)$. The result of BEM (resp. BBHMC) is on the left (resp. right)

Letter B and *Ring*). These standard images are used for tests in many image segmentation procedures [3], [7], [12], [23], [35]. Also, we compare performance (error rates) of BEM and BBHMC by applying them on the same synthetic images. These images are corrupted by the following noises:

MD: $\mu_1 = (112, 112)^{'}$, $\mu_2 = (115, 116)^{'}$ and $\Sigma_1 = \Sigma_2 = \begin{pmatrix} 300 & 105 \\ 105 & 360 \end{pmatrix}$.

These values correspond to $(\Delta, \rho) = (25, 1)$.

VD: $\mu_1 = \mu_2 = (112, 112)^{'}$, $\Sigma_1 = \begin{pmatrix} 300 & 325 \\ 325 & 360 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 500 & 350 \\ 350 & 900 \end{pmatrix}$.

These values correspond to $(\Delta, \rho) = (0, 0.27)$.

MVD: $\mu_1 = (112, 112)^{'}$, $\mu_2 = (115, 116)^{'}$, $\Sigma_1 = \begin{pmatrix} 300 & 325 \\ 325 & 360 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 500 & 350 \\ 350 & 900 \end{pmatrix}$. These values correspond to $(\Delta, \rho) = (25, 0.27)$.

To appreciate visually the level of the considered noises, some examples are given in Fig.8. The segmentation result of these examples by BBHMC with MPM method is presented in Fig. 9.

A. BHMC without and with Hilbert-Peano scan

The simplest idea for modeling an image by Markov chain is to consider this image "line by line" or "column by column". The problem of this method is that the past and the future of a pixel doesn't correspond always to its spatial context. Thus, when one considers the method "line by line", two neighboring pixels and belonging to the same column are near spatially and distant in sense of Markov chain. To avoid the problem of the temporal and spatial contexts, a lot of authors use the Hilbert-Peano scan [6], [16], [19]. In this method, the noisy image is firstly transformed into one dimension chain using Hilbert-Peano scan and the segmented image is reconstructed using an inverse Hilbert-Peano scan. The proposed model BHMC using noisy supplementary information overcomes the problem of adequacy between temporal and spatial context without using the Hilbert-Peano scan.

Fig. 7. *Alphabet*, *Gibbs*, *Letter B* and *Ring* (128×128) images

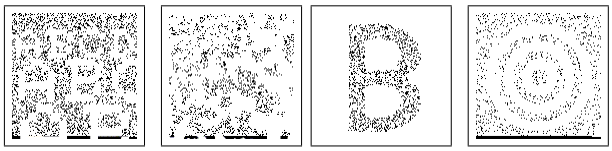

Fig. 8. *Alphabet*+**VD**, *Gibbs*+**MVD**, *Letter B*+**MD** and *Ring*+**VD**

In Tables VII and VIII, we give the error rates for each class (ER_1 and ER_2) obtained by BHMC with noisy supplementary information and BHMC with Hilbert-Peano scan, where images used are resized to 32×32. The estimation parameter problem of the hidden Markov chain is not landed here; the noise and chain parameters are estimated from EV and are used by MPM method for the segmentation.

From Table VII, we constat that the two approaches doesn't give satisfactory response for class 2 in all corrupted Alphabet cases. May be this is due to the image size, witch is small. In all other cases, BHMC with noisy supplementary information gives a reduced mean error rate with respect to the BHMC using Hilbert-Peano scan. The mean reduction factor is 2.14, 1.33 and 1.71 for Gibbs, Letter B and Ring respectively.

B. BEM and BBHMC performances

Before analyzing the results, let us note that for each statistical image segmentation method, there exists a theoretical error, obtained with the true parameter values [3]. In our case this error is estimated from empirical values EV (*i.e.* obtained from real and noisy images). For a simplification reason, we just interest to the error rates (wrongly classified pixels) and we do not give the parameters estimation. Thus we present,

TABLE VII
ERROR RATES FOR EACH CLASS (ER_1 AND ER_2) GIVEN BY BHMC WITHOUT HILBERT-PEANO SCAN AND BHMC WITH HILBERT-PEANO SCAN, FOR DIFFERENT NOISES APPLIED TO *Alphabet* AND *Gibbs* IMAGES

Model	*Alphabet*			*Gibbs*		
	MD	**VD**	**MVD**	**MD**	**VD**	**MVD**
BHMC ER_1	0.00	0.00	0.09	9.02	9.02	9.02
ER_2	100	100	100	0.16	0.16	0.16
BHMC with ER_1	0.19	0.09	0.09	15.2	15.7	15.5
Hilbert-Peano ER_2	100	100	100	4.8	4.5	4.7

TABLE VIII
ERROR RATES FOR EACH CLASS (ER_1 AND ER_2) GIVEN BY BHMC WITHOUT HILBERT-PEANO SCAN AND BHMC WITH HILBERT-PEANO SCAN, FOR DIFFERENT NOISES APPLIED TO *Letter B* AND *Ring images*

Model	*Letter B*			*Ring*		
	MD	**VD**	**MVD**	**MD**	**VD**	**MVD**
BHMC ER_1	7.3	11.05	10.6	5.21	6.9	7.2
ER_2	45.3	47.3	45.8	34.7	37.3	35.4
BHMC with ER_1	7.3	6.1	6.4	9.73	9.73	9.8
Hilbert-Peano ER_2	66.1	70.8	69.8	63.3	62.1	62.1

in Table IX, the error rates for each class (ER_1 and ER_2) obtained by EV, BEM and BBHMC on corrupted images.
We constat that BEM and BBHMC, compared to EV, prove successful in image *Letter B*. This can be explained by the fact that the two pixel classes are well separated. Also we constat that BEM is better than EV and BBHMC for class 2 in all cases, but the mean error rates given by BBHMC are significantly reduced than those obtained by BEM and they are nearer to those given by EV. The mean reduction factors between BBHMC and BEM are 2.65, 1.39, 1.79 and 1.53 for *Alphabet*, *Gibbs*, *Letter B* and *Ring* respectively. All the estimations, for each data scenario, are obtained by BEM and BBHMC for a mean CPU time equal to 36.5.

Fig. 9. *Alphabet*+**VD**, *Gibbs*+**MVD**, *Letter B*+**MD** and *Ring*+**VD** segmented by BBHMC with MPM method

VII. CONCLUSION

We have proposed a bi-dimensional Markov chain model, BHMC, mathematically designed by $(X, \mathbf{Y})$ where X is the unobserved Markov chain and $\mathbf{Y} = (Y^1, Y^2)$ is the observed process. In previous works Y^1 and Y^2 represent two images when, in our case Y^1 represent noisy image and Y^2 represent noisy supplementary information on the studied image (we have taken Y^2 as proportion of pixels in the same class as the current pixel in a neighborhood of the last). For the proposed model, we have considered a bayesian algorithm BBHMC to estimate the model parameters. We have compared BBHMC to the classical Expectation-Maximization algorithm

TABLE IX
ERROR RATES FOR EACH CLASS (ER_1 AND ER_2) GIVEN BY EV, BEM AND BBHMC FOR DIFFERENT NOISY IMAGES. THE MEAN CPU TIME FOR BEM AND BBHMC IS 36.5

Data	Error rates (%)	EV	BEM	BBHMC
Alphabet	ER_1	2.66	7.87	2.77
+ **MD**	ER_2	0.00	0.00	0.00
Alphabet	ER_1	2.88	7.89	3.14
+ **VD**	ER_2	0.00	0.00	0.00
Alphabet	ER_1	2.82	7.89	3.05
+ **MVD**	ER_2	0.00	0.00	0.00
Gibbs	ER_1	8.08	12.34	8.29
+ **MD**	ER_2	0.17	0.00	1.29
Gibbs	ER_1	8.32	12.77	8.61
+ **VD**	ER_2	0.27	0.00	0.25
Gibbs	ER_1	8.27	12.75	8.57
+ **MVD**	ER_2	0.27	0.00	0.24
LettreB	ER_1	1.95	2.55	1.93
+ **MD**	ER_2	0.00	0.00	0.00
LettreB	ER_1	2.14	2.78	2.16
+ **VD**	ER_2	0.00	0.00	0.00
LettreB	ER_1	2.14	2.78	2.13
+ **MVD**	ER_2	0.00	0.00	0.01
Ring	ER_1	2.81	4.53	2.87
+ **MD**	ER_2	0.00	0.00	0.00
Ring	ER_1	3.05	4.74	3.15
+ **VD**	ER_2	0.00	0.00	0.01
Ring	ER_1	2.99	4.73	3.11
+ **MVD**	ER_2	0.00	0.00	0.01

in bi-dimensional case, BEM. The results of these two algorithms are evaluated by empirical values, EV, computed from complete data. Our propositions are favorably tested on simulation and real data. We have considered different data scenarios and we have obtained satisfactory results for our approach: the obtained parameters estimation are nearer to those given by EV and the mean factor reduction of error rates with respect to BEM is between 1.33 and 2.65 for image segmentation. Also the proposed model BHMC outperforms (see Tables VII and VIII) the same model considered with Hilbert-Peano scan.
Actually, we interest to the generalization of this work to multi-dimensional case by integrating supplementary information to each image (*i.e.* $\mathbf{Y} = (Y^1, Y^{11}, Y^2, Y^{22}, \ldots, Y^M, Y^{MM})$, where Y^i is an image and Y^{ii} is a supplementary information on Y^i, for $i = 1, \ldots, M$ with $M > 2$).

REFERENCES

[1] A. Belmaati, L. Omari, H. Elmaroufy and S. Elfakir, *Modèle de Markov caché à processus observé bidimensionnel : étude de la transition entre les tats de l'infection par le VIH*, R.E.S.P, 52: IS42-IS44, 2007.
[2] A. P. Dempster, N. M. Laird and D. B. Rubin, *Maximum likelihood from incomplete data via the em algorithm (with discussion)*, Journal of the Royal Statistical Society, B 39 :1-38, 1977.
[3] A. Peng and W. Pieczynski, *Adaptive mixture estimation and unsupervised local Bayesian image segmentation*, CVGIP: Graphical Models and Image Processing, Vol. 57, No. 5, pp. 389-399, 1995.
[4] A. E. Raftery and S. Lewis, *Implementing MCMC. In W. R. Gillks, S. Richardson Markov chain Monte Carlo in practice*, Chapman and Hall/CRC, pp. 115-130, 1996.
[5] B. Chalmond, *An iterative Gibbsian technique for reconstruction of m-ary images*, Pattern Recognition, Vol. 22, N6, pp . 747-761, 1989.
[6] B. Benmiloud, A. Peng and W. Pieczynski, *Estimation conditionnelle itérative dans les chaînes de Markov cachées et segmentation statistique*

non supervisée d'images, Actes de Quatorzième Colloque GRETSI 93, Juans-les-Pins, France, pp. 105-108, 1993.

[7] B. Benmiloud and W. Pieczynski *Estimation des paramtres dans les chaînes de Markov cachées et segmentation d'images*, Traitement du Signal, Vol. 12, No. 5, pp. 433-454, 1995.

[8] B. Braathen, W. Pieczynski and P. Masson, *Global and local methods of unsupervised Bayesian segmentation of images*, Machine Graphics and Vision, Vol. 2, N1, pp.39-52, 1993.

[9] C. P. Robert *Le choix bayésien*, Springer, 2006.

[10] F. Heitz and P. Bouthemy *Estimation et segmentation du mouvement : approche Bayésienne et modélisation Markovienne des occlusions*, Actes de la 7ème Conf. AFCET/RFIA, Paris, pp. 1359-1368, 1989.

[11] F. Desbouvries and W. Pieczynski, *Modèles de Markov Triplet et filtrage de Kalman-Triplet Markov Models and Kalman filtering*, Comptes Rendus de l'Académie des Sciences-Mathématique, Série I, Vol. 336, No. 8, pp. 667-670, 2003.

[12] F. Salzenstein and W. Pieczynski, *Sur le choix de méthode de segmentation statistique d'images*, Traitement du Signal, Vol. 15, No. 2, pp. 120-127, 1998.

[13] J. Besag, *On the statistical analysis of dirty pictures*, Journal of the Royal Statistical Society, Series B, 48, pp . 259-302, 1986.

[14] H. Caillol, A. Hillion and W. Pieczynski, *Fuzzy random fields and unsupervised image segmentation*, IEEE Transaction on geoscience and remote sensing, Vol. 31, No 4, pp. 801-810, 1993.

[15] K. Yao, *A representation theorem and its applications to spherically invariant random processes*, IEEETrans.Inform.Theory,vol.IT-19, no.5,pp.600-608, 1973.

[16] K. Abend, T. J . Harley and L. N. Kanal, *Classification of binary random patterns*, IEEE Transactions on Information Theory, Vol. IT-11, N4, 1965.

[17] L. Younes, *Parametric inference for imperfectly observed Gibbsian fields*, Probability Theory and Related Fields, 82, pp . 625-645, 1989.

[18] L. E. Baum, T. Petrie, G. Soules and N. Weiss, *A maximization technique occuring in the statstical analysis of probabilistic functions of Markov chains* , Ann, Math. Statistic, 41, pp. 164-171, 1970.

[19] M. Emsalem, H. Caillol, P. Obvié, G. Carnat and W. Pieczynski, *Fast unsupervised statistical image segmentation*, IEEE International Geoscience and Remote Sensing Symposium (IGARSS 92), Houston, Texas, 1992.

[20] M. Vieira, *Estimation bayésienne par des méthodes MCMC-Application à la surveillance des moteurs asynchrones*, Thèse, Université de Nice-Sophia-Antiplois, 1999.

[21] M. Rangaswamy, D. Weiner and A. Ozturk, *Non-Gaussian random vector identification using spherically invariant random process*, IEEE Trans. on Aerospace and Electronic Systems , 29(1) :111-124, 1993.

[22] N. Brunel, J. Lapuyade-Lahorgue and W. Pieczynski, *Modeling and unsupervised classification of multivariate hidden Markov chains with copulas*, IEEE Trans. on Automatic Control, Vol. 55, No. 2, pp. 338-349, 2010.

[23] N. Giordana and W. Pieczynski, *Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 5, pp. 465-475, 1997.

[24] N. Brunel, W. Pieczynski and S. Derrode, *Copulas in vectorial hidden Markov chains for multicomponent image segmentation*, in InProc.Int.Conf.Acoust.,Speech Signal Processing(ICASSP'05), pp.717-720, 2005.

[25] O. Cappé, E. Moulines and T. Rydén, *Inference hidden Markov models*, Springer, 2005.

[26] P. A. Devijver and M. Dekesel, *Champs aléatoires de Pickard et modélisation d'images digitales*, Traitement du Signal, Vol. 5, N5, pp. 131-150, 1988.

[27] P. Lalande, *Détection du mouvement dans une séquence d'images selon une approche markovienne; Application à la robotique sous-marine*, Thèse, Université de Rennes I, 1990.

[28] P. Bouthemy and P. Lalande, *Motion detection in an image sequence usin Gibbs distribution*, Actes d'ICASSP'89, Glasgow, pp. 1651-1654, 1989.

[29] P. Masson and W. Pieczynski, *SEM Algorithm and unsupervised statistical segmentation and satellite images*, IEEE Transaction on geoscience and remote sensing, Vol. 31, No 3, pp. 618-633, 1993.

[30] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Transactions on PAMI, Vol. PAMI-6, N6, pp. 721-741, 1984.

[31] S. Derrode, G. Mercier and W. Pieczynski, *Unsupervised multicomponent image segmentation combining a vectorial HMC model and ICA*, International Conference on Image Processing (ICIP 2003), Barcelona, Spain, September 14-17, 2003.

[32] S. Rafi, M. Castella and W. Pieczynski, *Pairwise Markov model applied to unsupervised image separation*, IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA), 16-18 February, Innsbruck, Austria, 2011.

[33] T. J.BarnardandF.Khan, *Statistical normalization of spherically invariant non-Gaussian clutter*, IEEEJ.Oceanic Eng,vol.29,no.2,pp.303-309, 2004.

[34] Y. Delignon and W. Pieczynski, *Modeling non-Rayleigh speckle distribution in SAR images*, IEEETrans.Geosci.Remote Sensing, vol.40,no.6,pp.1430-1435, 2002.

[35] W. Pieczynski, *Champs de Markov cachés et estimation conditionnelle itérative*, Traitement du Signal, Vol. 11, No. 2, pp. 141-153, 1994.

[36] W. Pieczynski, *Modèles de Markov en traitement d'images* , Traitement du Signal, Vol. 20, No. 3, pp. 255-278, 2003.

[37] W. Pieczynski, *EM and ICE in hidden and triplet Markov models*, Stochastic Modeling Techniques and Data Analysis international conference (SMTDA '10), Chania, Greece, 2010.

[38] W. R. Gillks, A. Thomas and D. J. Spiegelhalter, *Software for the Gibbs Sampler*, Comp. Sci. Statist, 24: 439-448, 1992.

[39] W. R. Gillks and S. Richardson, *Markov chain Monte Carlo in practice*, Chapman and Hall/CRC, 1996.

[40] X. Guyon, *Champs aléatoires sur un réseau*, Collection Techniques Stochastiques, Masson, Paris, 1993.

Abdelali Joumad received the DESA degree in statistique et stochastique from the Cadi Ayyad University, Marrakesh, Morocco in December 2005. His research is supported by Laboratoire de Ibn-al Banna de Mathématiques et Appliquations (LIBMA), UCAM, Marrakech. His principals research interests concern Bayesian inference, hidden Markov chain, MCMC methods and image processing.

Abdelaziz Nasroallah received the Doctorat d'Université degree in applied mathematics from the Université Rennes 1, France, in 1991, and the Doctorat d'Etat degree in mathematical statistics from the Université Cadi Ayyad, Marrakesh, Morocco, in 1998. He is currently a professor at Faculty of Sciences Semlalia, Cadi Ayyad University. His research interests include mathematical statistics, MCMC simulations, Estimations of computer systems performances and discriminative analysis.