

# Signal and Image Segmentation Using Pairwise Markov Chains

Stéphane Derrode and Wojciech Pieczynski

**Abstract**—The aim of this paper is to apply the recent pairwise Markov chain model, which generalizes the hidden Markov chain one, to the unsupervised restoration of hidden data. The main novelty is an original parameter estimation method that is valid in a general setting, where the form of the possibly correlated noise is not known. Several experimental results are presented in both Gaussian and generalized mixture contexts. They show the advantages of the pairwise Markov chain model with respect to the classical hidden Markov chain one for supervised and unsupervised restorations.

**Index Terms**—Bayesian restoration, hidden data, hidden Markov chain, image segmentation, iterative conditional estimation, maximal posterior mode (MPM), maximum a posteriori, pairwise Markov chain, Pearson's system, unsupervised classification.

## I. INTRODUCTION

THE aim of this paper is to present some applications of a recent model called pairwise Markov chains (PMCs) [1], which generalizes the classical hidden Markov chain (HMC) model. The range of applications of the latter is rather vast, covering signal and image processing, economical prediction, and health sciences, among others. Specific applications of HMCs in image processing include image segmentation [2], [3], handwritten word recognition [4], document image analysis, tumor classification, vehicle detection [5], acoustic signal recognition [6], or even gesture recognition [7]. Multisensor or multiresolution images can likewise be segmented using HMCs [3], [8]. Other applications include speech recognition [9], [10], communications [11], and genome structure recognition [12].

The success of HMC models is due to the fact that when the unobservable signal process  $X$  can be modeled by a finite Markov chain and when the noise is not too complex, then  $X$  can be recovered from the observed process using different Bayesian classification techniques like maximum a posteriori (MAP) or maximal posterior mode (MPM) [13], [14]. Most remarkably, these classification methods remain valid even in very large data set cases.

Manuscript received October 25, 2002; revised September 22, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ta-Hsin Lu.

S. Derrode is with the GSM Group, Institut Fresnel (CNRS UMR-6133), École Nationale Supérieure de Physique de Marseille, Domaine Universitaire de Saint Jérôme, F-13013 Marseille Cedex 20, France (e-mail: stephane.derrode@egim-mrs.fr).

W. Pieczynski is with the GET/INT, Département CITI (CNRS UMR-5157), Institut National des Télécommunications, F-91000 Evry, France (e-mail: wojciech.pieczynski@int-evry.fr).

Digital Object Identifier 10.1109/TSP.2004.832015

The efficiency of HMMs (chains or fields) in image segmentation hinges principally on their spatial regularity constraint. This is justified by the fact that for many natural scenes, neighboring pixels are more likely to belong to the same class than pixels that are farther away from each other. However, a pixel belonging to a given class may have a rather different visual aspect when it is located near a boundary or inside a large set of pixels of the same class. According to the classical hypothesis in hidden Markov fields (HMFs) [15]–[19], this fact cannot be taken into consideration. This is the very reason for the recent pairwise Markov field (PMF) model [20], in which the pair  $(X, Y)$  is assumed to be a Markov field. For analogous reasons, the PMC model, in which  $(X, Y)$  is a Markov chain, has then been proposed [1], [21], and it has been shown that PMC is strictly more general than HMC in that  $X$  is not necessarily a Markov chain [1]. Another advantage over HMMs is that the noise correlation can be easily taken into account.

Like HMC, the PMC model allows one to perform Bayesian MAP and MPM restorations [1], and the aim of this paper is to study the advantages of PMC with respect to HMC in the context of unsupervised restoration. Beforehand, the unknown model parameters have to be estimated. When HMC is concerned, the Expectation–Maximization (EM) procedure [13], [22] is among the best known methods and generally works very well, especially when the noise is Gaussian. In this work, we will use the general iterative conditional estimation (ICE) method [23], which can be seen as an alternative to EM in situations where the latter is difficult to apply. In particular, ICE has been successfully used in several unsupervised image segmentation contexts. For instance, sonar [24] and medical [25] images, spatio-temporal video images [2], radar [26], and multiresolution [27], [28] images were segmented using different ICE-based unsupervised Bayesian methods. Otherwise, EM and ICE have been compared in the case of Gaussian noise in the HMC models [2], and it turns out that EM and ICE are equivalent, but EM remains faster. In other situations, ICE could possibly appear more flexible [19] since the likelihood is not necessarily used, and estimators other than the maximum likelihood estimator can be employed. Some relations between EM and ICE in the context of the exponential family of distributions are described in [29].

We focus here on unsupervised restoration based on the application of ICE to PMC models, in both Gaussian and “generalized” mixture estimation contexts. Our aim is to answer the following questions.

- When the data follow a PMC model, does the use of an HMC significantly degrade the restoration results, and how does ICE work in this context?

- When the data neither follows a PMC nor an HMC model, which should be used, and how does ICE behave?
- Like HMC, PMC can be used in image segmentation by converting the two-dimensional (2-D) set of pixels into a one-dimensional (1-D) sequence via a Hilbert–Peano scan. Does the use of PMC instead of HMC present any interest in such a context?

The third point is relevant since HMC and Hilbert–Peano scan-based methods [2], [3] have shown to be of interest in the image segmentation context and can even compete, in some particular situations, with the HMF-based methods in terms of classification accuracy, while being much faster [30]. Furthermore, since the HMF model provides a finer and more intuitive modeling of spatial relationships, whereas the estimation of the regularity parameters is much more robust in HMC, the two approaches can be combined in a way that conserves their respective advantages. Such a hybrid method uses HMC first and HMF only for the final estimation and classification [26]. Let us finally mention that this approach can also be used with other Markov models, e.g., Markov mesh [31] and Markov tree [32], [33].

The paper is organized as follows. The PMC model is briefly reviewed in Section II. Section III is devoted to the supervised restoration of simulated Gaussian PMCs from both MPM-based HMC and MPM-based PMC classifications. Section IV is concerned with the parameter estimation problem. In particular, we specify how the original ICE method works in the Gaussian mixture estimation case and how it can be extended to the generalized mixture estimation case. Section V contains applications of the PMC-based Bayesian MPM restoration algorithm in order to segment simulated images with non-Gaussian and correlated noise in supervised and unsupervised manners. We also show and discuss the segmentation results obtained from a real radar image. Experiments are conducted using both HMC and PMC models and for both Gaussian- and Pearson-based generalized mixture estimation cases. Finally, conclusions and further work are drawn in the Section VI.

## II. PAIRWISE MARKOV CHAINS

Let us consider two sequences of random variables  $\mathbf{X} = (X_1, \dots, X_N)$ , and  $\mathbf{Y} = (Y_1, \dots, Y_N)$ . Each  $X_n$ ,  $1 \leq n \leq N$  takes its values in a finite set  $\Omega = \{1, \dots, K\}$  of classes, and each  $Y_i$  takes its values in the set of real numbers  $\mathbb{R}$ . Considering  $Z_i = (X_i, Y_i)$ , the process  $\mathbf{Z} = (Z_1, \dots, Z_N)$  is called a “pairwise process” associated with  $\mathbf{X}$  and  $\mathbf{Y}$ . Realizations of such processes will be denoted by lowercase letters. To simplify notations, we will write  $p(x_n)$  for  $p(X_n = x_n)$ , and we will denote different distributions by the same letter  $p$ , except for distributions on  $\mathbb{R}$  and  $\mathbb{R}^2$ , where  $f$  will be used.

The process  $\mathbf{Z}$  is then called a “pairwise Markov chain” (PMC) associated with  $\mathbf{X}$  and  $\mathbf{Y}$  if its distribution can be expressed as  $p(\mathbf{z}) = p(z_1) p(z_2|z_1) \dots p(z_N|z_{N-1})$ . In this paper, we only consider the “stationary PMCs” in which  $p(z_n, z_{n+1})$  does not depend on  $n$ . The distribution of a PMC  $\mathbf{Z}$  is then given by

$$p(z_1, z_2) = p(i, j) f_{i,j}(y_1, y_2) \quad (1)$$

where  $p(i, j)$  is a probability on  $\Omega^2$ , and  $f_{i,j}(y_1, y_2) = p(y_1, y_2|i, j)$  are distribution densities on  $\mathbb{R}^2$ . Hence,  $p(z_1, z_2)$  is a probability distribution on  $\Omega^2 \times \mathbb{R}^2$ . The distribution of the Markov chain  $\mathbf{Z}$  will equivalently be determined by the initial probabilities  $p(z_1)$  given by

$$p(z_1) = \sum_{\omega \in \Omega} p(i, \omega) \int_{\mathbb{R}} f_{i,\omega}(y_1, y_2) dy_2 = \sum_{\omega \in \Omega} p(i, \omega) f_{i,\omega}(y_1) \quad (2)$$

and the transition matrix  $p(z_2|z_1)$  given by

$$p(z_2|z_1) = \frac{p(z_1, z_2)}{p(z_1)} = \frac{p(i, j) f_{i,j}(y_1, y_2)}{\sum_{\omega \in \Omega} p(i, \omega) f_{i,\omega}(y_1)}. \quad (3)$$

We may then state [1] the following.

- $\mathbf{Y}$  conditionally on  $\mathbf{X} = \mathbf{x}$  and  $\mathbf{X}$  conditionally on  $\mathbf{Y} = \mathbf{y}$  are Markov chains.
- $f_{i,j}(y_1) = f_i(y_1)$  implies that  $\mathbf{X}$  is a Markov chain, and reciprocally, the Markovianity of  $\mathbf{X}$  implies  $f_{i,j}(y_1) = f_i(y_1)$ .

Property ii) states a necessary and sufficient condition that a PMC be an HMC and shows that a PMC is strictly more general than an HMC. In particular, the classical HMC widely used is obtained by putting  $f_{i,j}(y_1, y_2) = f_i(y_1) f_j(y_2)$  in (1), which gives

$$p(z_1, z_2) = p(i, j) f_i(y_1) f_j(y_2). \quad (4)$$

In the following, we will call an HMC the classical HMC, verifying (4).

## III. RESTORATION OF SIMULATED GAUSSIAN PMCS

The aim of this section is to test and compare the HMC- and PMC-based MPM restoration algorithms in the context of Gaussian PMCs. A PMC is called Gaussian if all densities  $f_{i,j}(y_1, y_2)$  are Gaussian. As specified above, the distribution of  $\mathbf{Z}$  is given by the distribution  $p(z_1, z_2)$  or, equivalently, by the  $K^2$  probabilities  $p(i, j)$  and the parameters  $\mu_1^{i,j}$ ,  $\mu_2^{i,j}$ ,  $\sigma_1^{i,j}$ ,  $\sigma_2^{i,j}$ , and  $\rho^{i,j}$  of the  $K^2$  bi-dimensional densities  $f_{i,j}(y_1, y_2)$ . In this section, all these parameters are supposed to be known, and we deal with, first, the simulation of a Gaussian PMC  $\mathbf{Z}$ , i.e., the simulation of  $\mathbf{X}$  and the simulation of  $\mathbf{Y}$ , and second, the restoration of  $\mathbf{X}$  from  $\mathbf{Y}$  only, according to the HMC- and PMC-based MPM restoration methods.

### A. Simulation of a PMC

The simulations of  $\mathbf{X}$  and  $\mathbf{Y}$  are done sequentially and alternatively according to the following:

- the simulation of  $x_1$  and  $y_1$  using drawings from the following probabilities:

$$p(x_1) = \sum_{\omega \in \Omega} p(x_1, x_2 = \omega) \quad (5)$$

$$p(y_1|x_1) = \sum_{\omega \in \Omega} p(x_1, x_2 = \omega) f_{x_1, x_2 = \omega}(y_1). \quad (6)$$

*Remark:*  $f_{i,j}(y_1)$  is a Gaussian density with mean  $\mu_1^{i,j}$  and standard deviation  $\sigma_1^{i,j}$ , and the simulation of  $y_1$  is obtained by a drawing from a Gaussian mixture.

- the simulation of  $x_{n+1}$  and  $y_{n+1}$ , for  $1 \leq n < N$ , using drawings from the following probabilities:

$$p(x_{n+1}|z_n) = \frac{p(x_n, x_{n+1})f_{x_n, x_{n+1}}(y_n)}{\sum_{\omega \in \Omega} p(x_n, x_{n+1} = \omega)f_{x_n, x_{n+1} = \omega}(y_n)} \quad (7)$$

$$p(y_{n+1}|x_{n+1}, z_n) = \frac{p(z_{n+1}|z_n)}{p(x_{n+1}|z_n)} = \frac{f_{x_n, x_{n+1}}(y_n, y_{n+1})}{f_{x_n, x_{n+1}}(y_n)}. \quad (8)$$

*Remark:* It can be shown that  $p(y_{n+1}|x_{n+1}, z_n)$  is a Gaussian density with mean  $\mu_2^{i,j} + \rho^{i,j} (\sigma_2^{i,j} / \sigma_1^{i,j}) (y_n - \mu_1^{i,j})$  and standard deviation  $\sigma_2^{i,j} \sqrt{1 - (\rho^{i,j})^2}$ , and the simulation of  $y_{n+1}$  is obtained using a Gaussian drawing.

### B. Bayesian MPM Restoration of a PMC

The Bayesian MPM restoration of the hidden sequence  $\mathbf{X}$  can be calculated according to

$$[\hat{\mathbf{x}} = \hat{\mathbf{x}}_{MPM}(\mathbf{y})] \Leftrightarrow [\hat{x}_n = \arg \max_i \Phi_n(i)] \quad (9)$$

with  $\Phi_n(i) = p(x_n = i|\mathbf{y})$ . As shown in [1], the MPM restoration can be achieved thanks to the ‘‘forward probabilities’’  $F_n(i) = p(y_1, \dots, y_n, x_n = i)$  and the ‘‘backward probabilities’’  $B_n(i) = p(y_{n+1}, \dots, y_N | y_n, x_n = i)$ , for  $1 \leq i \leq K$  and  $1 \leq n \leq N$ , since

$$\Phi_n(i) = \frac{p(x_n = i, \mathbf{y})}{p(\mathbf{y})} = \frac{F_n(i) B_n(i)}{\sum_{\omega \in \Omega} F_n(\omega) B_n(\omega)}. \quad (10)$$

These probabilities are generalizations of the classical ‘‘forward’’ and ‘‘backward’’ probabilities considered in HMC. They can be computed recursively using a method analogous to the well-known Baum’s algorithm [13]. However, the original algorithm is subject to serious numerical problems since  $F_n(i) \xrightarrow{n \rightarrow N} 0$  and  $B_n(i) \xrightarrow{n \rightarrow 0} 0$  at exponential rates, causing underflow on computers. The same behavior was encountered with PMC, and we adapted the variant of the Baum’s algorithm proposed in [34] to the PMC model. Indeed, one can easily show that

$$\Phi_n(i) = \alpha_n(i) \beta_n(i) \quad (11)$$

where  $\alpha_n$  and  $\beta_n$  are the ‘‘normalized’’ forward and backward probabilities defined by

$$\alpha_n(i) = p(x_n = i | y_1, \dots, y_n) \quad (12)$$

$$\beta_n(i) = \frac{p(y_{n+1}, \dots, y_N | x_n = i, y_n)}{p(y_{n+1}, \dots, y_N | y_1, \dots, y_n)}. \quad (13)$$

These probabilities no longer suffer from underflow and can be computed using the following recursion:

$$\begin{aligned} \alpha_1(i) &= \frac{p(x_1 = i, y_1)}{\sum_{\omega \in \Omega} p(x_1 = \omega, y_1)} \\ \alpha_n(i) &= \frac{\sum_{\omega \in \Omega} \alpha_{n-1}(\omega) p(x_n = i, y_n | x_{n-1} = \omega, y_{n-1})}{\sum_{(\omega_1, \omega_2) \in \Omega^2} \alpha_{n-1}(\omega_1) p(x_n = \omega_2, y_n | x_{n-1} = \omega_1, y_{n-1})} \end{aligned} \quad (14)$$

and

$$\begin{aligned} \beta_N(i) &= 1 \\ \beta_n(i) &= \frac{\sum_{\omega \in \Omega} \beta_{n+1}(\omega) p(x_{n+1} = \omega, y_{n+1} | x_n = i, y_n)}{\sum_{(\omega_1, \omega_2) \in \Omega^2} \alpha_n(\omega_1) p(x_{n+1} = \omega_2, y_{n+1} | x_n = \omega_1, y_n)}. \end{aligned} \quad (15)$$

*Remark:* Let us now sketch a Viterbi-like algorithm for the MAP restoration in a PMC model. The Bayesian MAP restoration of the hidden sequence  $\mathbf{X}$  can be calculated according to

$$[\hat{\mathbf{x}} = \hat{\mathbf{x}}_{MAP}(\mathbf{y})] \Leftrightarrow \left[ \hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \Omega^N} p(\mathbf{x}|\mathbf{y}) = \arg \max_{\mathbf{x} \in \Omega^N} p(\mathbf{x}, \mathbf{y}) \right]. \quad (16)$$

The Viterbi algorithm [9], [14] is based on the quantity

$$D_n(i) = \max_{x_1, \dots, x_{n-1} \in \Omega} p(x_1, \dots, x_{n-1}, x_n = i, y_1, \dots, y_n)$$

which can be computed recursively using a formulation similar to the one used to compute the ‘‘forward’’ probabilities  $F_n(i)$ . To avoid underflow, we used the following solution. The complete procedure can be stated as follows (see [9] for details):

1) Recursion:

$$\begin{aligned} \delta_1(i) &= \frac{p(x_1 = i, y_1)}{\sum_{\omega \in \Omega} p(x_1 = \omega, y_1)} \\ \psi_1(i) &= 0 \end{aligned} \quad (17)$$

$$\begin{aligned} \delta_n(i) &= \frac{\max_{\omega \in \Omega} \delta_{n-1}(\omega) p(x_n = i, y_n | x_{n-1} = \omega, y_{n-1})}{\sum_{\omega_2 \in \Omega} \max_{\omega_1 \in \Omega} \delta_{n-1}(\omega_1) p(x_n = \omega_2, y_n | x_{n-1} = \omega_1, y_{n-1})} \\ \psi_n(i) &= \arg \max_{\omega \in \Omega} \delta_{n-1}(\omega) p(x_n = i, y_n | x_{n-1} = \omega, y_{n-1}). \end{aligned} \quad (18)$$

2) Path backscattering:

$$\begin{aligned} \hat{x}_N &= \arg \max_{1 \leq i \leq K} \psi_N(i) \\ \hat{x}_n &= \psi_{n+1}(\hat{x}_{n+1}). \end{aligned} \quad (19)$$

### C. Restoration Results

Let us consider the case of two classes:  $\Omega = \{1, 2\}$ . The restoration experiments are conducted according to

- 1) the simulation of the PMC  $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ , with  $N = 5000$  samples (Section III-A);
- 2) the restoration of  $\mathbf{X}$  with HMC- and PMC-based Bayesian MPM and MAP segmentation methods, which give, respectively,  $\mathbf{X}_{HMC}$  and  $\mathbf{X}_{PMC}$  (Section III-B);
- 3) the computation of the misclassification rates and the gain obtained.

The four Gaussian distributions  $f_{i,j}(y_1, y_2)$  are defined by the parameters given in Table I. We test the influence of the noise correlation on the misclassification rate using the four sets of correlation coefficients reported in Table II. We also investigate two different homogeneities of the class process:

- Series 1:  $p(1, 1) = p(2, 2) = 0.48$ ,  $p(1, 2) = p(2, 1) = 0.02$ ;
- Series 2:  $p(1, 1) = p(2, 2) = p(1, 2) = p(2, 1) = 0.25$ .

TABLE I  
PARAMETERS OF THE FOUR GAUSSIAN DENSITIES USED IN PMC AND HMC SIMULATION AND RESTORATION. THE CORRELATION COEFFICIENTS ARE GIVEN IN TABLE II

Densities	PMC parameters				HMC parameters					
	$\mu_1^{i,j}$	$\mu_2^{i,j}$	$\sigma_1^{i,j}$	$\sigma_2^{i,j}$	$\rho^{i,j}$	$\mu_1^{i,j}$	$\mu_2^{i,j}$	$\sigma_1^{i,j}$	$\sigma_2^{i,j}$	$\rho^{i,j}$
$f_{1,1}$	-5	-5	14	14	$\rho^{1,1}$	-5	-5	14	14	0
$f_{1,2}$	-3	3	7	9	$\rho^{1,2}$	-5	5	14	20	0
$f_{2,1}$	3	-3	9	7	$\rho^{2,1}$	5	-5	20	14	0
$f_{2,2}$	5	5	20	20	$\rho^{2,2}$	5	5	20	20	0

TABLE II  
CORRELATION COEFFICIENTS USED IN THE FOUR SIMULATIONS

Experiment	$\rho^{1,1}$	$\rho^{1,2}$	$\rho^{2,1}$	$\rho^{2,2}$
1	0.1	0.1	0.1	0.1
2	0.1	0.9	0.9	0.1
3	0.9	0.1	0.1	0.9
4	0.9	0.9	0.9	0.9

The misclassification rates reported in Table III are means obtained from 250 independent experiments. Clearly, the results obtained with the MAP criterion are comparable with the ones obtained from the MPM criterion.

For the first series, the class process is characterized by a few transitions between classes 1 and 2, and the gain between HMC- and PMC-based classifications is more important when the coefficients  $\rho^{1,1}$  and  $\rho^{2,2}$  are bigger (experiments 3 and 4). It can also be seen from experiments 1 and 2 that the gain improves when the correlation coefficients associated with the transitions between classes 1 and 2 increase. For the second series, the class process presents a lot of transitions, and the gain increases as the correlation coefficients  $\rho^{1,2}$  and  $\rho^{2,1}$  become stronger (experiments 2 and 4).

We have made numerous other simulations, and the PMC-based MPM and MAP restorations consistently work better than the HMC-based ones, as expected from general Bayesian theory. Let us now examine how parameters can be estimated from incomplete data.

#### IV. PMC PARAMETER ESTIMATION USING ICE

In the case of unsupervised classification, all the PMC parameters are unknown and must be estimated from the observed data  $\mathbf{Y} = \mathbf{y}$ . Therefore, the problem is to estimate the  $K^2$  probabilities  $p(i, j)$  and the  $K^2$  densities  $f_{i,j}(y_1, y_2)$  from a sample  $\mathbf{y} = (y_1, \dots, y_N)$ . We use here the ICE procedure, which is based on the conditional expectation of some estimator defined from the complete data  $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ . More precisely, one can estimate all the parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$  with ICE once an estimator  $\hat{\theta} = \hat{\theta}(\mathbf{z})$  is available or a sampling of  $\mathbf{X}$  according to  $p(\mathbf{x}|\mathbf{y})$  is possible. ICE is an iterative method that produces a sequence  $(\theta_i^q)$  as follows:

- initialize  $\theta_i^0$ ;
- put  $\theta_i^{q+1} = E_{\theta_i^q} [\hat{\theta}_i(\mathbf{X}, \mathbf{Y}) | \mathbf{Y} = \mathbf{y}]$  for all components  $\theta_i$  for which this

expectation is computable, or put  $\theta_i^{q+1} = 1/r \left( \hat{\theta}_i(\mathbf{x}^1, \mathbf{y}) + \dots + \hat{\theta}_i(\mathbf{x}^r, \mathbf{y}) \right)$ , where  $\mathbf{x}^1, \dots, \mathbf{x}^r$  are sampled according to  $p(\mathbf{x}|\mathbf{y}, \theta^q)$ , for all components  $\theta_i$  for which the expectation above is not computable.

Returning to our problem, we are going to see that the expectation above is computable in the case of components given by the  $K^2$  probabilities  $p(i, j)$ , but it is not computable in the case of the  $K^2$  densities  $f_{i,j}(y_1, y_2)$ .

*Remark:* The EM algorithm produces a sequence of parameters according to the following principle: Starting from  $\theta^0$ ,  $\theta^{q+1} = \arg \max_{\theta} E_{\theta^q} [\text{Log}(p_{\theta}(\mathbf{X}, \mathbf{Y})) | \mathbf{Y} = \mathbf{y}]$ , which is quite different from  $\theta^{q+1} = E_{\theta^q} [\hat{\theta}(\mathbf{X}, \mathbf{Y}) | \mathbf{Y} = \mathbf{y}]$  given with ICE. Therefore, ICE and EM are different, and none of them can be seen as a particular case of the other. However, let us remark on two points.

- 1) When  $\hat{\theta}(\mathbf{X}, \mathbf{Y})$  used by ICE is the ML estimator  $\hat{\theta}(\mathbf{X}, \mathbf{Y}) = \arg \max_{\theta} \text{Log}(p_{\theta}(\mathbf{X}, \mathbf{Y}))$ , and when the maximization in EM and the expectation in ICE permute (which is quite exceptional), EM and ICE are identical.
- 2) When neither the maximization in EM nor the expectation in ICE can be computed, they are approximated by some stochastic methods, and the methods chosen can be close or even the same. For example, the sampling we use is the same that the sampling used in Stochastic EM [35], which is a stochastic approximation of EM.

#### A. Estimation of the Joint Probabilities

The estimation of the joint probabilities below remains valid for the Gaussian case as well as for the generalized one. Regarding  $p(i, j)$ , we can take the following standard estimation:

$$\hat{p}(i, j)(\mathbf{z}) = \frac{\text{Card}(A^{i,j})}{N-1} \quad (20)$$

where  $A^{i,j}$  denotes the set of indices  $1 \leq n < N$  for which  $(x_n, x_{n+1}) = (i, j) \in \Omega^2$ . Applying ICE, the conditional expectation of  $\hat{p}(i, j)(\mathbf{z})$  at iteration  $q+1$  is computable and gives

$$p^{q+1}(i, j) = E[\hat{p}(i, j)(\mathbf{z}) | \mathbf{Y} = \mathbf{y}] = \frac{1}{N-1} \sum_{n=1}^{N-1} \Psi_n^q(i, j) \quad (21)$$

TABLE III

HMC- AND PMC-BASED BAYESIAN MPM AND MAP MISCLASSIFICATION RATES FOR THE PARAMETERS SPECIFIED IN TABLES I AND II, AND FOR THE TWO HOMOGENEITIES. THE GAIN COLUMNS GIVE THE BENEFIT OF USING THE PMC MODEL INSTEAD OF THE HMC ONE. THE VALUES BETWEEN PARENTHESES ARE THE VARIANCES OF THE GAIN COMPUTED ON THE 200 INDEPENDENT EXPERIMENTS

Experiment		Series 1 - misclassification rates			Series 2 - misclassification rates		
		$(\mathbf{X}, \mathbf{X}_{HMC})$	$(\mathbf{X}, \mathbf{X}_{PMC})$	Gain	$(\mathbf{X}, \mathbf{X}_{HMC})$	$(\mathbf{X}, \mathbf{X}_{PMC})$	Gain
1	MAP	17.8%	15.5%	13.2% (0.19)	36.9%	30.5%	17.2% (0.07)
	MPM	15.0%	13.7%	8.2% (0.18)	36.9%	28.8%	22.0% (0.06)
2	MAP	17.9%	13.4%	25.1% (0.45)	36.8%	12.8%	65.3% (0.04)
	MPM	15.2%	12.3%	19.1% (0.24)	36.8%	12.5%	66.1% (0.04)
3	MAP	35.9%	20.6%	42.4% (0.84)	36.9%	28.8%	21.8% (0.08)
	MPM	35.6%	17.8%	49.9% (0.32)	36.9%	27.9%	24.2% (0.08)
4	MAP	36.5%	20.9%	42.6% (0.46)	37.0%	18.1%	50.3% (0.06)
	MPM	36.2%	18.3%	49.1% (0.33)	37.0%	17.8%	51.9% (0.05)

where  $\Psi_n^q(i, j) = p^q(x_n = i, x_{n+1} = j | \mathbf{y})$  is the ‘‘joint probability’’ of two subsequent classes that can be calculated, at iteration  $q$ , from

$$\Psi_n(i, j) = \frac{\alpha_n(i)p(x_{n+1}=j, y_{n+1}|x_n=i, y_n)\beta_{n+1}(j)}{\sum_{(\omega_1, \omega_2) \in \Omega^2} \alpha_n(\omega_1)p(x_{n+1}=\omega_2, y_{n+1}|x_n=\omega_1, y_n)\beta_{n+1}(\omega_2)}. \quad (22)$$

For latter use, let us notice that, similar to the HMC case, it is possible to simulate  $\mathbf{X}$  according to its distribution conditional on  $\mathbf{Y} = \mathbf{y}$ . Indeed, the *a posteriori* distribution of  $\mathbf{X}$ , i.e.,  $p(\mathbf{x} | \mathbf{Y} = \mathbf{y})$  is that of a nonstationary Markov chain with transition matrix

$$p(x_{n+1} = j | x_n = i, \mathbf{y}) = \frac{\Psi_n(i, j)}{\Phi_n(i)} \quad (23)$$

using (11) and (22), for  $1 \leq j \leq K$  and  $1 \leq n < N$ . Since the search of the families  $f_{i,j}^{q+1}(y_1, y_2)$  differs according to ‘‘Gaussian’’ or ‘‘generalized’’ cases, we present below an application of ICE to Gaussian PMC and another one to a more general setting.

### B. Gaussian PMC

First of all, we have to estimate  $K^2$  Gaussian densities from the complete data  $\mathbf{Z} = \mathbf{z}$ . Denoting by  $\mu^{i,j}$  the mean vector and by  $\Gamma^{i,j}$  the variance-covariance matrix of  $f_{i,j}(y_1, y_2)$ , we may use the following classical estimators:

$$\hat{\mu}^{i,j}(\mathbf{z}) = \frac{1}{\text{Card}(A^{i,j})} \sum_{n=1}^{N-1} \mathbf{1}_{A^{i,j}} \begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} \quad (24)$$

$$\hat{\Gamma}^{i,j}(\mathbf{z}) = \frac{1}{\text{Card}(A^{i,j})} \cdot \sum_{n=1}^{N-1} \mathbf{1}_{A^{i,j}} \left( \begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} - \hat{\mu}^{i,j}(\mathbf{z}) \right) \left( \begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} - \hat{\mu}^{i,j}(\mathbf{z}) \right)^t. \quad (25)$$

The conditional expectations of  $\hat{\mu}^{i,j}$  and  $\hat{\Gamma}^{i,j}$  are not computable. From the ICE principle, a solution consists of

- performing  $L$  realizations  $\mathbf{x}_1, \dots, \mathbf{x}_L$  of  $\mathbf{X}$  (each  $\mathbf{x}_l, l \in \{1, \dots, L\}$  is here a sequence) according to its distribu-

tion conditional on  $\mathbf{Y} = \mathbf{y}$  (using (23)) and based on the parameter at iteration  $q$ ;

- computing an estimate of the parameters from  $\mathbf{y}$  and each  $\mathbf{x}_l$ ;
- setting the parameters at iteration  $q+1$  as the mean of the  $L$  estimates.

In practice, just one realization of  $\mathbf{X} = \mathbf{x}^q$  is sampled. Thus, the next  $\hat{\mu}^{i,j}$  and  $\hat{\Gamma}^{i,j}$  are estimated using (24) and (25), with  $\mathbf{x}^q$  instead of  $\mathbf{x}$  for the computation of  $A^{i,j}$ .

### C. Generalized PMC

When the nature of the  $K^2$  densities  $f_{i,j}(y_1, y_2)$  on  $\mathbb{R}^2$  is not known, but each of them belongs to a known set  $H_{i,j}$  of possible shapes, the problem of finding them is called ‘‘generalized mixture estimation.’’ Different methods have been proposed to solve such a problem, in the HMF context [36] as well as in the HMC context [3]. Globally, one may say that when generalizing ICE means dealing with such situations, the obtained ‘‘generalized’’ ICE performs well. However, the methods in [3] and [36] are concerned with the models in which the random variables  $(Y_n)$  are independent conditionally on  $\mathbf{X}$ . Hence, one has to search  $f_i(y_1)$  in a known set  $H_i$  instead of  $f_{i,j}(y_1, y_2)$  in a known set  $H_{i,j}$ . Therefore, the method presented here is more general.

Therefore, for each  $(i, j) \in \Omega^2$ , the density  $f_{i,j}(y_1, y_2)$  belongs to  $H_{i,j} = \{F_{i,j}^1, \dots, F_{i,j}^{P_{i,j}}\}$ . Furthermore, we assume that each family  $F_{i,j}^p, 1 \leq p \leq P_{i,j}$  is parameterized by a set of parameters  $\beta_{i,j}^p$ . Once we know that the sample  $u_1, \dots, u_S$  (each  $u_s$  is in  $\mathbb{R}^2$ ) is provided according to a density in  $F_{i,j}^p$ ,  $\beta_{i,j}^p$  can be estimated by an estimator  $\hat{\beta}_{i,j}^p(u_1, \dots, u_S)$ . We can also consider a nonparametric estimate  $\hat{h}$  of a density on  $\mathbb{R}^2$ , which can be the bidimensional histogram, or any other estimator-like kernel estimators or orthogonal probability density function estimation [37]. Finally, we consider a ‘‘decision rule’’  $D$  that allows one to choose, for a given density  $h$ , one density among a given set of densities  $f^1, \dots, f^P$ . For example, for a given distance  $d$ , a decision rule  $D$  can consist of choosing  $f^k$ , which minimizes the distance to  $h$ :  $D(h) = f^k$  if  $d(h, f^k) = \inf_{1 \leq l \leq P} d(h, f^l)$ .

At iteration  $q$ , the generalized ICE we propose gives the next distribution  $f_{i,j}^{q+1}(y_1, y_2)$  by doing the following.

- Calculate  $p^{q+1}(i, j)$  given by (21).
- Simulate  $\mathbf{X} = \mathbf{x}^q$  according to its distribution conditional to  $\mathbf{Y} = \mathbf{y}$  and based on the parameters  $p^q(i, j)$  and  $f_{i,j}^q(y_1, y_2)$ , using (23).
- Calculate  $P_{i,j}$  sets of parameters  $\beta_{i,j}^1, \dots, \beta_{i,j}^{P_{i,j}}$  by applying the  $P_{i,j}$  estimators  $\hat{\beta}_{i,j}^1, \dots, \hat{\beta}_{i,j}^{P_{i,j}}$  to the sample  $(y_n, y_{n+1})_{n \in A^{i,j}}$ . These sets of parameters define  $P_{i,j}$  densities  $f_{i,j}^1 \in F_{i,j}^1, \dots, f_{i,j}^{P_{i,j}} \in F_{i,j}^{P_{i,j}}$ .
- Calculate  $\hat{h}_{i,j}$  from  $(y_n, y_{n+1})_{n \in A^{i,j}}$ , and choose  $f_{i,j}^{q+1}(y_1, y_2)$  in the set  $\{f_{i,j}^1, \dots, f_{i,j}^{P_{i,j}}\}$  as being  $D(\hat{h}_{i,j})$ .

Let us notice that when each set  $H_{i,j}$  is reduced to one element  $H_{i,j} = \{F_{i,j}\}$ , the generalized mixture becomes a classical mixture. In fact, no decision rule is needed since  $f_{i,j}^{q+1}(y_1, y_2)$  is directly given by the parameter estimated from the estimator  $\hat{\beta}_{i,j}$ . Furthermore, when all  $F_{i,j}$  are Gaussian families, we find a Gaussian ICE, as specified in the previous section. For non-Gaussian cases, bidimensional densities can be sometimes difficult to handle, and we now propose a general method to deal with 2-D distributions. Note that in the following, only families with parametric estimators will be considered, even if the method is not restricted to them.

#### D. Estimation of the 2-D Densities

In some situations, there may be  $i$  and  $j$  such that  $f_{i,j}(y_1, y_2)$  is difficult to obtain explicitly. For example, we have an idea about the marginal densities  $f_{i,j}(y_1)$  and  $f_{i,j}(y_2)$ , and we have their correlation. Of course, this does not give the density  $f_{i,j}(y_1, y_2)$ , which is needed in generalized ICE iterations. We can then use the following technique, inspired from the method successfully applied to bisensor image segmentation in [8].

Denoting by  $\sigma_1^{i,j}$  and  $\sigma_2^{i,j}$  the standard deviations of  $f_{i,j}(y_1)$  and  $f_{i,j}(y_2)$ , let us put  $Y_1^* = Y_1/\sigma_1^{i,j}$ ,  $Y_2^* = Y_2/\sigma_2^{i,j}$ , and  $\rho^{i,j} = \text{Cov}(Y_1^*, Y_2^*)$ . Considering

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} Y_1^* \\ Y_2^* \end{pmatrix} \quad (26)$$

with  $a = \frac{1}{2}\sqrt{1+\rho^{i,j}} + \frac{1}{2}\sqrt{1-\rho^{i,j}}$  and  $b = \frac{1}{2}\sqrt{1+\rho^{i,j}} - \frac{1}{2}\sqrt{1-\rho^{i,j}}$ , we have

- $\text{Var}(U_1) = \text{Var}(U_2) = 1$ ;
- $\text{Cov}(U_1, U_2) = 0$ ;
- $\text{Cov}(U_1, Y_1^*) = \text{Cov}(U_2, Y_2^*)$ .

The choice for  $(U_1, U_2)$  verifying i) and ii) is not unique. However, we also need the third condition to obtain a stationary PMC. Indeed, the distribution of  $(z_n, z_{n+1})$  should not depend on  $n$ , just like  $p(z_n)$ , so that we must be able to write  $p(z_1) = p(z_2)$ . A sufficient condition is given by  $p(z_1, z_2) = p(z_2, z_1)$  and iii) allows the conservation of this symmetry.

We then assume that  $U_1$  and  $U_2$  are independent, which gives  $g(u_1, u_2) = g_1(u_1)g_2(u_2)$ . At each iteration of the generalized ICE and for each  $(i, j) \in \Omega^2$ , we thus seek  $g_1$  and  $g_2$ . Since they are densities on  $\mathbb{R}$ , this is much simpler than to seek a density on

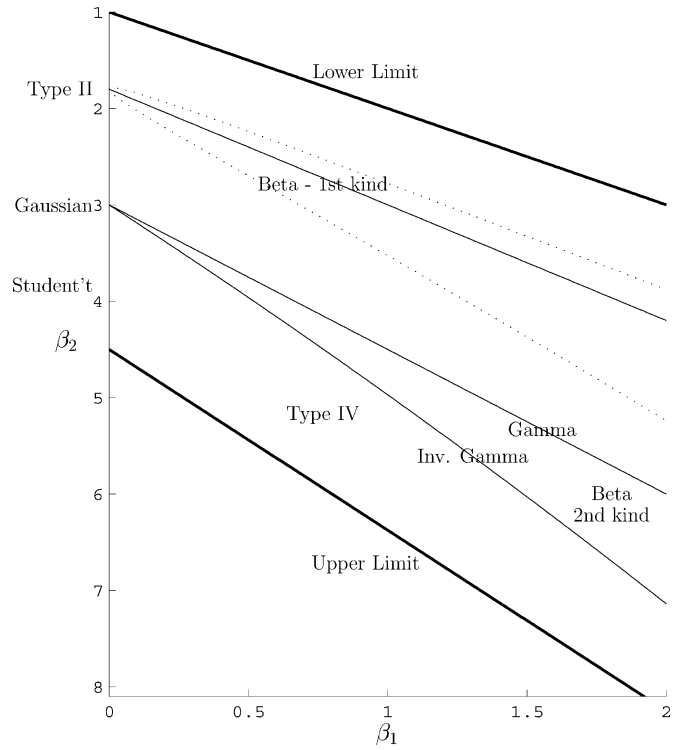


Fig. 1. Pearson's diagram in terms of skewness and kurtosis. Note that the  $\beta_2$  axis is reversed.

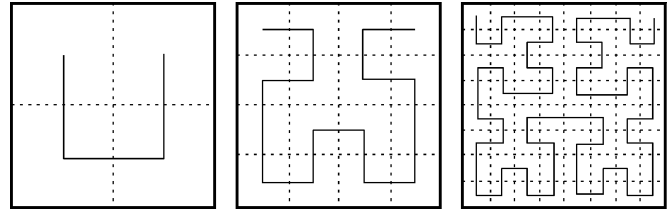


Fig. 2. Three steps in the construction of the Hilbert-Peano scan.

$\mathbb{R}^2$ . The density  $f$  of the distribution of  $(Y_1^*, Y_2^*)$  is then linked to  $g_1$  and  $g_2$  by the relation  $f_{i,j}(y_1^*, y_2^*) = |\det(A)| g_1(ay_1^* + by_2^*) g_2(by_1^* + ay_2^*)$ , and we can finally write

$$f_{i,j}^{q+1}(y_1, y_2) = \sigma_1^{i,j} \sigma_2^{i,j} |\det(A)| g_1(ay_1 + by_2) g_2(by_1 + ay_2) \quad (27)$$

where  $\det(A)$  is the determinant of the  $2 \times 2$  matrix in (26). Hence, the parameters that have to be estimated are  $\sigma_1^{i,j}$ ,  $\sigma_2^{i,j}$ ,  $\rho^{i,j}$  and the parameters needed to describe the families of  $g_1$  and  $g_2$ . In the next section, we will consider the case where the families of  $g_1$  and  $g_2$  belong to the Pearson's system of distributions.

#### V. APPLICATION TO IMAGE SEGMENTATION

In this section, we present some results regarding the application of the PMC model to the segmentation of images in both Gaussian and generalized mixture contexts. In the first set of experiments, we deal with noisy simulated images corrupted with some correlated noise, and we compare the restoration results obtained from both HMC- and PMC-based Bayesian MPM restoration algorithms. In the second set of experiments, we illustrate the segmentation results obtained on a synthetic aperture radar (SAR) image. These experiments give an idea of the

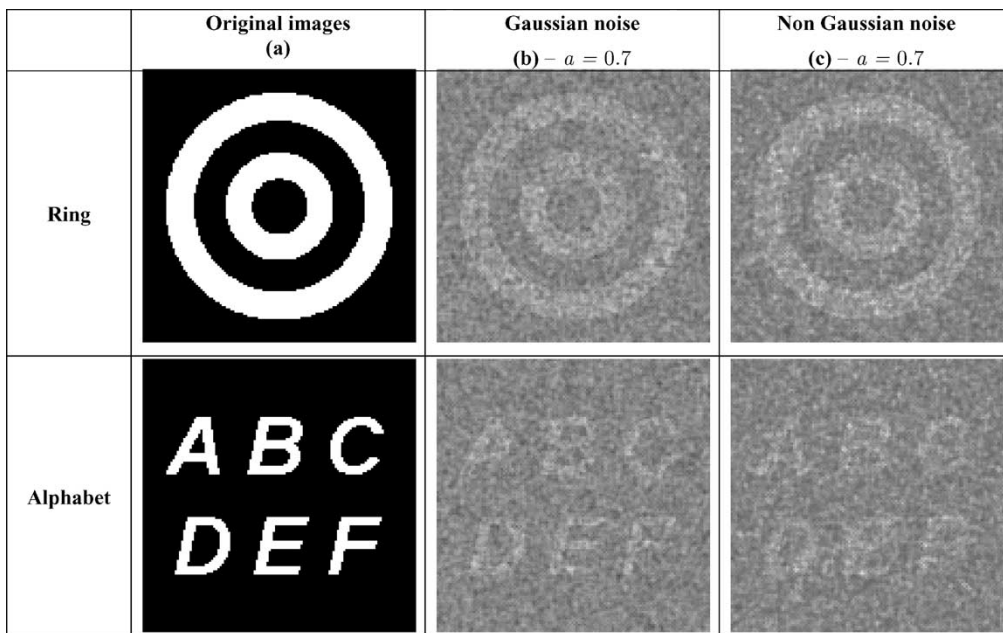


Fig. 3. Two synthetic images (column 1) and the noisy simulated images (columns 2 and 3) with correlated noise ( $a = 0.7$ ). Images size:  $128 \times 128$ . We stretch the dynamic of the noisy images to ensure a better visualization. The shape and the parameters of noise are reported in Table IV.

behavior of the ICE method for real image segmentation. Beforehand, let us first introduce the Pearson system of distributions that will be used in the subsequent sections in order to illustrate how ICE performs in a Pearson-based generalized mixture context.

#### A. Pearson System of Distributions

Since the problem of finding bidimensional distributions has been reduced to the search for 1-D densities, a number of candidate families can be used, such as Gamma and Beta. In order to enlarge the set of available shapes, one solution is to consider the Pearson system of distributions. Comprehensive introduction and detailed statements on the Pearson system are given in [38]. This system consists of mainly eight families of distributions of various types with monomodal and possibly non-symmetrical shapes (Gamma, Exponential, and Beta distributions of the first and second kinds, among others). The Pearson system has shown to be really interesting in the context of SAR image segmentation [36], [39]. Let  $\mu_2$ ,  $\mu_3$ , and  $\mu_4$  denote the centered moments of order 2, 3, and 4. All the families can be expressed in terms of the skewness ( $\beta_1 = \mu_3^2/\mu_2^3$ ) and kurtosis ( $\beta_2 = \mu_4/\mu_2^2$ ) and can be located in the so-called Pearson's diagram; see Fig. 1. In image segmentation, the interest in these families comes essentially from the large variety of possible shapes. Some of them have a bounded extent and are well suited to model the noise appearing in several image modalities, such as radar, sonar, X-ray, etc.

We thus seek both the family of distributions and the parameters that best describe its samples. All the families of parameters can be expressed in terms of the four first centered moments so that the parameters needed to be estimated are the correlation coefficient  $\rho^{i,j}$  and the first four moments for  $g_1$  and for  $g_2$ . From a realization  $\mathbf{x}$  of  $\mathbf{X}$ , sampled by using transitions given by (23), one can estimate the empirical moments

TABLE IV  
WHITE NOISE PARAMETERS USED FOR THE SEGMENTATION EXPERIMENTS

	Distribution	$\mu_1$	$\mu_2$	$\mu_3$	$\mu_4$	$\beta_1$	$\beta_2$
Gaussian	Gaussian	120.0	49.0	0.0	7203.0	0.0	3.0
	(b) Gaussian	125.0	75.0	0.0	16875.0	0.0	3.0
Non Gaussian	Gamma	120.0	49.0	343.0	10804.5	1.0	4.5
	(c) Inv. Gamma	125.0	75.0	918.6	40159.1	2.0	7.1

and compute  $(\beta_1, \beta_2)$ . Given the Pearson's diagram, it becomes possible to select the corresponding family member and recover the parameters, which precisely identify the distribution, from the estimated moments. The densities  $f_{i,j}$  are then reconstructed using (27). The generalized ICE algorithm, when applied to the Pearson system, will be termed the "Pearson-ICE" algorithm.

In this context, one difficulty arises from the computation of

$$\begin{aligned}
 f_{i,j}(y_1) &= \int_{\mathbb{R}} f_{i,j}(y_1, y_2) dy_2 \\
 &= \sigma_1^{i,j} \sigma_2^{i,j} |\det(A)| \\
 &\quad \times \int_{\mathbb{R}} g_1(ay_1 + by_2) g_2(by_1 + ay_2) dy_2 \quad (28)
 \end{aligned}$$

which happens in the expressions of  $p(z_1)$  and  $p(z_2|z_1)$  that is needed to sample  $\mathbf{X}$ . The integral cannot be solved analytically, and numerical integration must be performed.

#### B. Segmentation of Noisy Simulated Images

The aim of this subsection is twofold. First, we want to test the robustness of PMC segmentation methods with respect to HMC-based ones. In other words, we wonder what happens when the data are neither PMC nor HMC suited? When considering a noisy simulated image with correlated noise (in some way), and when transforming the bi-dimensional set of pixels into a monodimensional set via a Hilbert-Peano scan (Fig. 2,

TABLE V  
MISCLASSIFICATION RATES FOR THE DIFFERENT MPM SEGMENTATIONS. "REAL PARAMETERS" MEANS THE PARAMETERS OBTAINED FROM COMPLETE DATA. THE GAIN COLUMN REPRESENTS THE BENEFIT OF USING THE PMC MODEL INSTEAD OF THE HMC ONE. THE SEGMENTED IMAGES OBTAINED WITH  $a = 0.7$  ARE REPORTED IN FIG. 4

		Misclassification rates			
		MPM based on PMC and real parameters	MPM based on HMC and ICE	MPM based on PMC and ICE	Gain ( $\tau$ )
Ring (Gaussian)	$a = 0.2$	7.53%	13.32%	7.65%	42.57%
	$a = 0.7$	6.21%	12.46%	7.13%	42.78%
Alphabet (Gaussian)	$a = 0.2$	7.12%	17.41%	9.05%	48.02%
	$a = 0.7$	6.34%	17.78%	8.89%	50.00%
Ring (non Gaussian)	$a = 0.2$	7.46%	18.30%	7.79%	57.43%
	$a = 0.7$	6.82%	14.00%	7.57%	45.93%
Alphabet (non Gaussian)	$a = 0.2$	7.89%	31.75%	9.52%	70.02%
	$a = 0.7$	7.67%	26.18%	13.35%	49.01%

[40]), we obtain a hidden stochastic process with a very complex structure. In particular, two neighboring pixels in the bidimensional set of pixels can be displaced far from each other in the monodimensional set obtained with the Hilbert–Peano scan. For example, the pixels at the fourth and fifth columns of the first row in the “step 3” drawn in Fig. 2 are ranked, respectively, at the sixth and 59th positions in the chain. Therefore, the process is obviously neither HMC nor PMC, and the question is how the HMC- and PMC-based unsupervised MPM methods will work. Second, the same experiments might be of interest in the image segmentation context. In fact, segmentation methods based on ICE, Hilbert–Peano scan, and HMC, which were introduced in [2] in the context of uncorrelated noise, are very fast. Furthermore, they can give good results [3] and can even be comparable, in some situations, to the results obtained with HMF [30]. In other situations, when HMF methods are clearly better, the HMC methods based on the Hilbert–Peano scan can still be of interest for the initialization of the HMF methods [26]. Finally, when the noise is correlated, the question we deal with here is whether PMC have to be used instead of HMC or not.

The two synthetic images “Rings” and “Alphabet” in column 1 of Fig. 3 are corrupted with correlated noises. More specifically, the observed field is of the form

$$Y_s = \frac{1}{1 + 4a} \left( W_s + a \sum_{i=1}^4 W_{s_i} \right)$$

where  $W = (W_s)$  is a white noise,  $s_1, s_2, s_3,$  and  $s_4$  are four neighbors of pixel  $s$ , and  $x_s = \omega_1$  (white) or  $x_s = \omega_2$  (black). The obtained image is then transformed in a sequence  $t_1, \dots, t_N$ , via the Hilbert–Peano scan. Putting  $X_i = X_{t_i}$  and  $Y_i = Y_{t_i}$ , we have a realization of the stochastic process  $\mathbf{X} = (X_1, \dots, X_N)$ , and a realization of the stochastic process  $\mathbf{Y} = (Y_1, \dots, Y_N)$ . The process  $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$  has a rather complex structure, and its distribution is neither an HMC distribution nor a PMC one; however, the noisy images are segmented in an unsupervised manner by HMC- and PMC-based MPM restoration methods, with all parameters estimated with ICE, as specified in Section IV.

We have performed numerous simulations, and it turns out that PMC-based MPM restoration works consistently better

than the HMC-based one. To give some numerical examples, we consider two kinds of noise: Gaussian and non-Gaussian. Parameters of the white noises are reported in Table IV. For each original image, we considered two experiments corresponding to  $a = 0.2$  and  $a = 0.7$ . For  $a = 0.7$ , the obtained images are reported in columns 2 and 3 in Fig. 3.

We then performed three segmentations:

- i) MPM based on PMC and the parameters estimated from the complete data, i.e., the original and noisy images;
- ii) MPM based on HMC and the parameters estimated with ICE (Gaussian ICE for the images with Gaussian noises and Pearson-ICE for the images with non Gaussian noises);
- iii) MPM based on PMC and the parameters estimated with ICE.

From the error rates in Table V, we observe the following. The results in column 1 can be seen as a kind of reference since the complete data are used to estimate the parameters. Comparing column 1 with column 3, one can notice that both Gaussian-ICE and Pearson-ICE present a favorable behavior since most of the time the misclassification rates are similar. The only exception is the last experiment, where we get 7.67% and 13.35%. Therefore, the first conclusion of interest in real applications is that ICE presents a quite correct behavior in the context considered. The second conclusion is concerned with the comparison of columns 2 and 3. Knowing that the data fits neither an HMC nor a PMC model, the HMC-based MPM, which is simpler with well-known robustness, might have given better results than PMC-based MPM. According to the results, and the ones from other experiments we have performed, this does not occur. Each time, PMC-based MPM restorations give better results than HMC-based ones, even in the presence of rather strong noise (Fig. 4). Hence, the use of MPM based on PMC, the Hilbert–Peano scan, and ICE provides fast unsupervised segmentation methods that can strongly improve the HMC-based ones.

Let us also notice that the estimated parameters may be used for other objectives beyond MPM segmentation. Table VI reports one example of parameters estimated from the complete data and the parameters found with ICE. In addition, the estima-





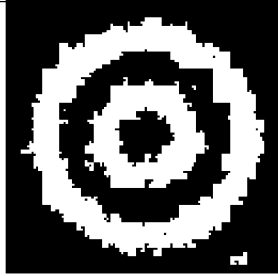


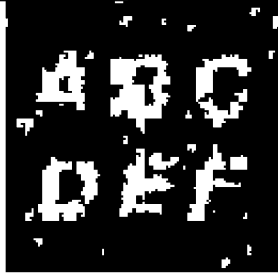
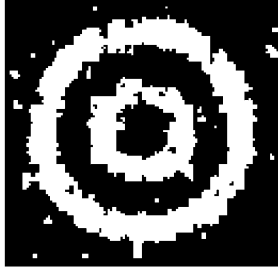
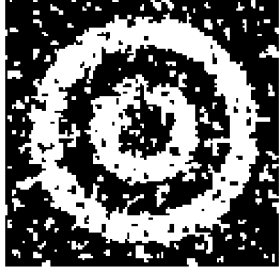

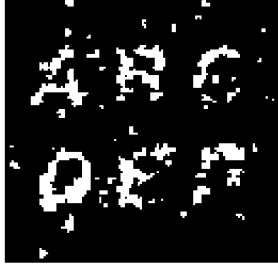


	MPM based on PMC and real parameters	MPM based on HMC and ICE	MPM based on PMC and ICE
Ring (Gaussian) $\mathcal{T} = 42.8\%$			
Alphabet (Gaussian) $\mathcal{T} = 50.0\%$			
Ring (Non Gaussian) $\mathcal{T} = 45.9\%$			
Alphabet (Non Gaussian) $\mathcal{T} = 49.0\%$			

Fig. 4. Segmentation results corresponding to the second experiment ( $a = 0.7$ ) in Table V. The gain  $\tau$  is reported here for convenience.

tors based on the moments  $\mu_1, \mu_2, \mu_3$ , and  $\mu_4$  we use inside ICE could be replaced by some other estimators [41], which could possibly improve the whole procedure.

*Remark:* The Hilbert–Peano scan belongs to the family of space filling curves (SFCs), which include the Lebesgue (or Z-order), the zig-zag, and the standard raster scans. SFCs have been used in a wide variety of image processing applications such as compression, halftoning, pattern recognition, and texture analysis (see [42] and [43] and references cited therein). The Hilbert–Peano scan furnishes a good exploitation of the 2-D locality. This property, together with the pseudo randomness of direction changes in the scan, implies that the Hilbert–Peano scan would work well (statistically) for a large family of images, especially for real images. Recently, a context-based SFC approach [42] has been proposed in order to adapt the scan to

the content in the image, using some criterion. This technique would work well for an image or a given set of images.

### C. Segmentation of a SAR Image

Let us carry on with the comparison between HMC and PMC models in the context of generalized mixture and radar image segmentation. Part of this experiment has been presented in [44]. Fig. 5 shows an extract of an ERS image of rice growing in Semarang (Java island) with mainly early rice, late rice, and other cultivations [45]. It was decided to classify the image into four classes with 100 ICE iterations and one sampling of  $X$  per iteration. We only test the Pearson-ICE case for both HMC and PMC models. The MPM classification results are reported in Fig. 6, together with the normalized histogram of each class. It is clear that the PMC-based segmentation is much more regular

TABLE VI

ESTIMATION OF THE PARAMETERS FOR THE “ALPHABET” IMAGE, GAUSSIAN NOISE, AND  $a = 0.7$ .  $N$  REPRESENTS THE NUMBER OF PAIRS  $(\omega_i, \omega_j)$  IN THE SEGMENTED CHAIN. HMC-BASED GAUSSIAN ICE PARAMETER ESTIMATION GIVES  $\mu_{\omega_1} = 119.3, \mu_{\omega_2} = 124.3, \sigma_{\omega_1} = 2.78, \sigma_{\omega_2} = 3.23$ , AND  $p(\omega_1, \omega_1) = 0.69, p(\omega_1, \omega_2) = 0.03, p(\omega_2, \omega_1) = 0.03, p(\omega_2, \omega_2) = 0.24$

	$\omega_i, \omega_j$	$p(\omega_i, \omega_j)$	$N$	$f_{\omega_i, \omega_j}(y_1, y_2)$				
				$\mu_{\omega_i}$	$\mu_{\omega_j}$	$\sigma_{\omega_i}$	$\sigma_{\omega_j}$	$\rho$
Estimates from complete data	$\omega_1, \omega_1$	0.650	10750	119.9	119.9	3.22	3.22	0.48
	$\omega_1, \omega_2$	0.012	266	120.0	124.9	3.57	3.89	0.55
	$\omega_2, \omega_1$	0.012	266	124.9	119.9	3.82	3.69	0.48
	$\omega_2, \omega_2$	0.326	5101	125.0	125.0	3.91	3.91	0.46
PMC Gaussian-ICE estimates from observed data	$\omega_1, \omega_1$	0.649	10627	120.0	119.9	3.26	3.24	0.47
	$\omega_1, \omega_2$	0.005	80	119.0	123.4	4.14	4.83	0.71
	$\omega_2, \omega_1$	0.005	80	123.5	122.9	3.32	5.27	-0.20
	$\omega_2, \omega_2$	0.342	5596	124.7	124.7	4.02	4.00	0.47

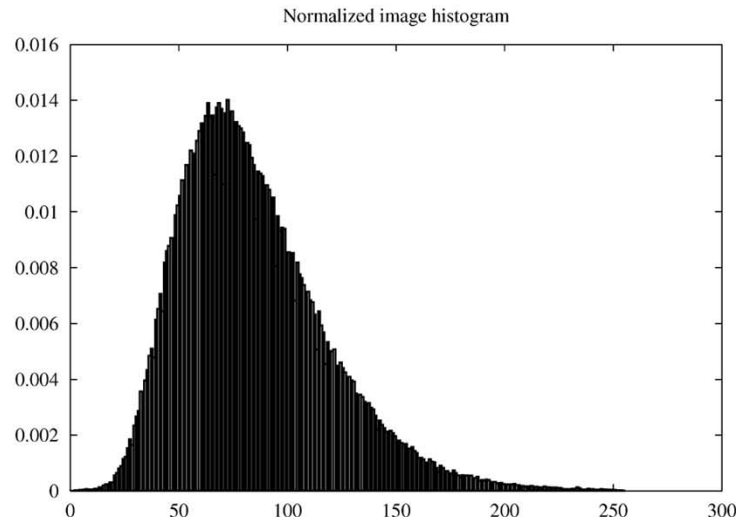
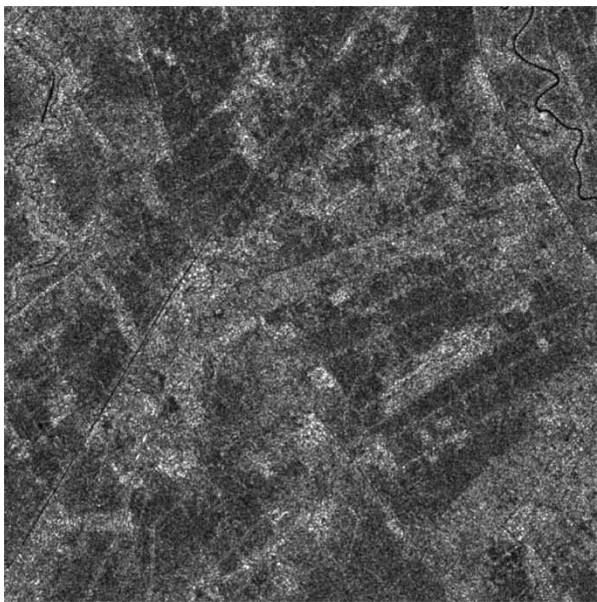


Fig. 5. ERS1 SAR image of a rice plantation in Java Island (Indonesia) and histogram. Image size:  $512 \times 512$ . Date: 1994, 10 February.

than the HMC one, and rice plots seem to be better extracted. Note also that the configurations of class histograms are quite different for the two models. In the HMC case, a class has been specialized on a few number of pixels whose intensity is located on the tail of the image histogram.

Table VII reports the correlation coefficients of the 2-D distributions  $f_{i,j}, 1 \leq i, j \leq 4$ . Most of them are far from being zero, which justifies the segmentation improvement when using a PMC model instead of an HMC one. Fig. 7 shows the 2-D density  $f_{2,3}$  obtained from the two densities representing  $g_1$  and  $g_2$ . The first one is a Type IV law with moments  $\mu_1 = 2.15, \mu_2 = 1, \mu_3 = 0.69, \mu_4 = 4.18$  ( $\beta_1 = 0.48$ , and  $\beta_2 = 4.19$ ). The second one is a Type II law with moments  $\mu_1 = 3.18, \mu_2 = 1, \mu_3 = 0, \mu_4 = 2.84$  ( $\beta_1 = 0$ , and  $\beta_2 = 2.85$ ). The correlation coefficient is given by  $\rho^{2,3} = 0.45$ .

The last point concerns the computation time between the two algorithms. The program based on HMC spent about 35 min on a PC with Pentium IV 1.3-GHz processor running Linux, whereas the program based on PMC needs 2 h and 40 min. The complexity of the PMC-ICE algorithm is more

important since not only  $K = 4$  densities have to be estimated (HMC case), but  $2K^2 = 32$  1-D densities (PMC case) also have to be estimated. Another time-consuming point of the algorithm comes from the numerical integration needed to get  $f_{i,j}$ , as described in (28).

## VI. CONCLUSIONS AND PERSPECTIVES

We have described different supervised and unsupervised hidden data restoration methods using the recent “Pairwise Markov Chain” model. The main contribution was to propose an original parameter estimation method that is valid in possibly correlated and non-Gaussian noise cases. In fact, the PMC model is more general, and more complex, than the classical HMC model and allows one to take into account more complex noise structures. We have first shown that when data follow a PMC, the corresponding Bayesian restorations work better than the HMC based ones. Furthermore, the parameter estimation method proposed is quite effective since the real

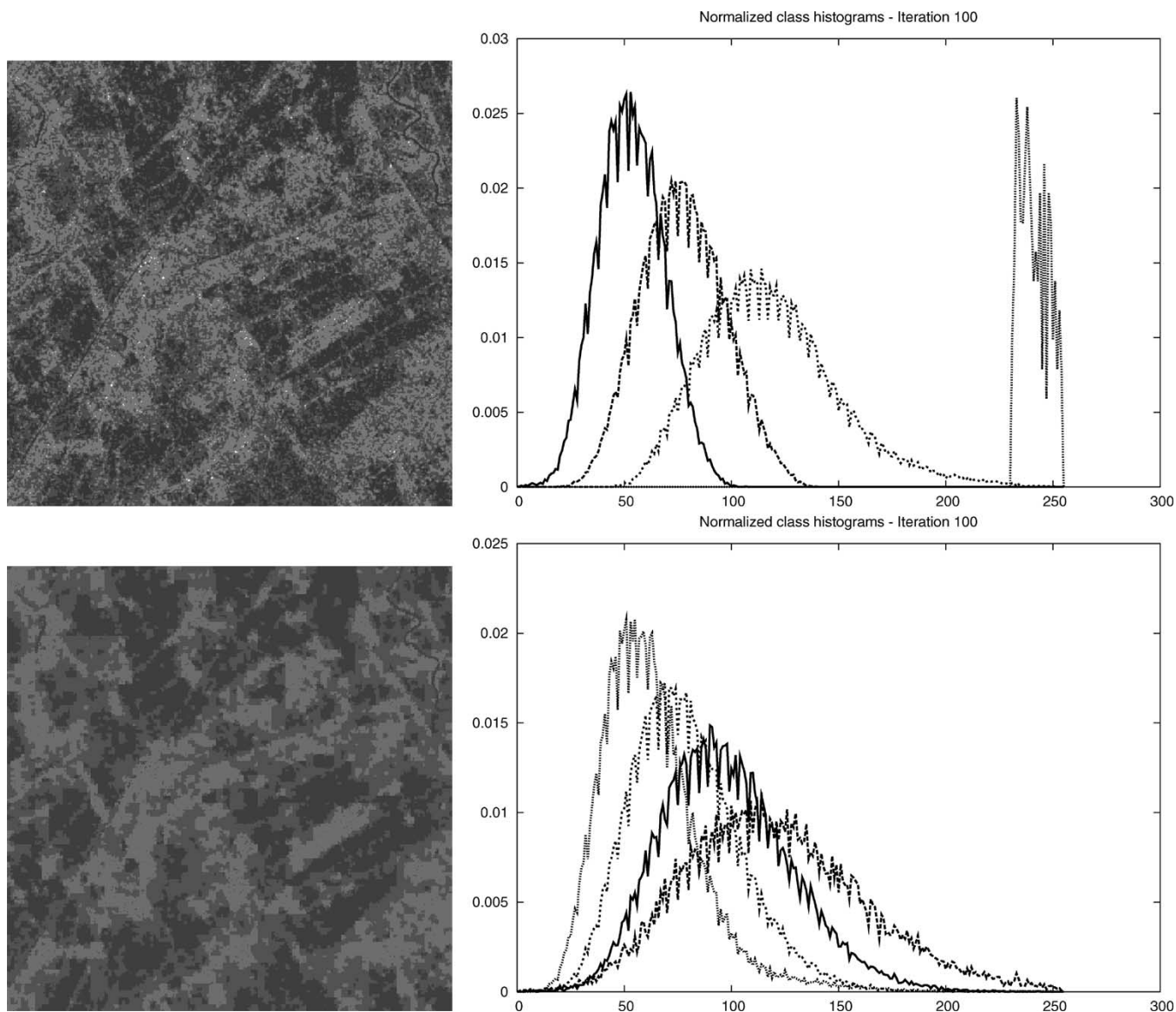


Fig. 6. Segmentation results and class histograms for both (top) ICE and HMC classification and (bottom) ICE and PMC classification. The class values (1, . . . , 4) have been changed to the mean values of the classes to facilitate a visual comparison with the original image.

TABLE VII  
CORRELATION COEFFICIENTS OBTAINED FOR THE 16 COUPLES OF DENSITIES  $(g_1, g_2)$  INVOLVED IN THE EXPERIMENT

	1	2	3	4
1	0.36	0.58	0.32	0.57
2	0.53	0.83	0.20	0.19
3	0.41	0.70	0.31	0.45
4	0.37	0.24	0.35	0.33

parameter-based restorations are close to estimated parameter-based ones. The second set of experiments was concerned with the unsupervised image segmentation application via a Hilbert–Peano scan of images. The interest of this study was then double. First, the probabilistic structure of such a stochastic process obtained from a noisy class image with correlated noise is very complex and does not correspond to a PMC nor to an HMC structure. Therefore, it was of interest

to study whether the PMC-based segmentation worked better than the HMC based one or not. We have made numerous experiments, some of which are reported in the paper, and it occurs that the PMC-based segmentation always gives better results than the HMC-based one, and the difference can even be quite striking. Therefore, in this first interpretation, using noisy images and the Hilbert–Peano scan was simply a way to produce complex data. Second, this approach is of interest in the image segmentation problem. In fact, it was shown that in some particular situations, the efficiency of such an HMC-based approach can compete with the efficiency of classical HMF-based approaches [30], the latter being much more time consuming. In other situations, when HMF-based approaches work better, one can still use the HMC-based approaches to initialize the HMF ones [26]. Finally, a particular case of the new parameter estimation method, appealing on the Pearson system of distributions, has been specified, and unsupervised segmentation of a real radar image has been

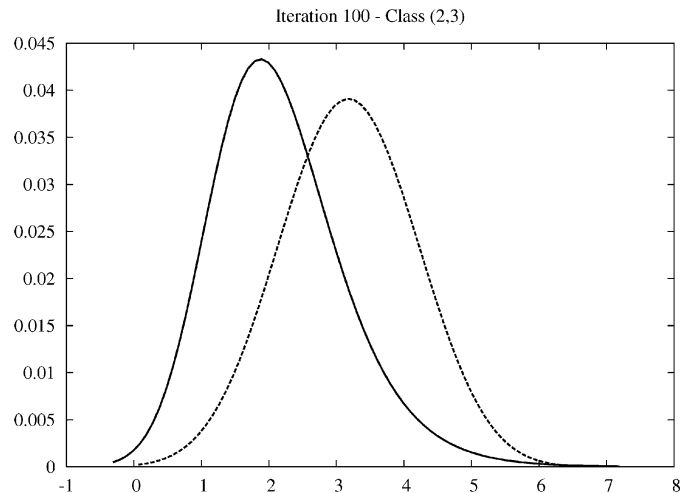
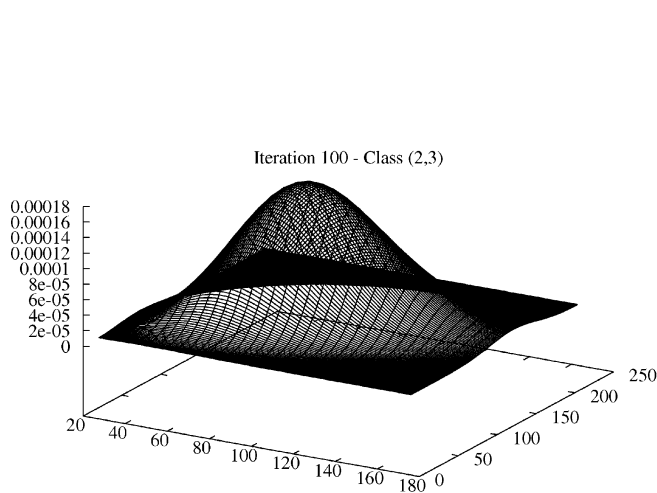


Fig. 7. Bi-dimensional density (left) obtained from the two densities (right) using (28).

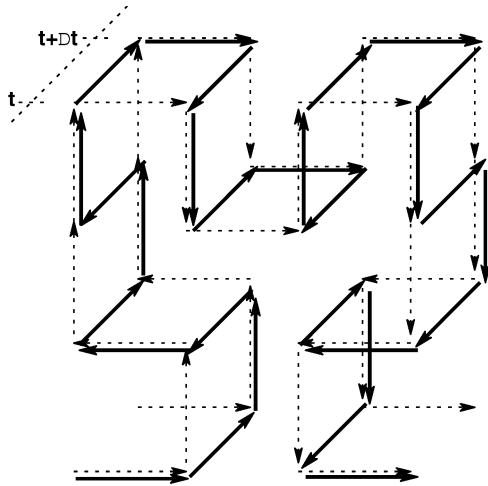


Fig. 8. Spatio-temporal Hilbert-Peano scan. Dotted arrows represent the HP scan at  $t$  and  $t + \Delta t$ , whereas solid arrows represent the spatio-temporal HP scan.

performed for both HMC and PMC contexts. PMC-based segmentation gives more regular regions, which is due to the fact that the spatial correlated speckle noise can be directly taken into account in the PMC model and not in the HMC one.

As a general conclusion, we may then affirm that the new parameter estimation method proposed is well adapted to the PMC model on the one hand and that the PMC-based unsupervised restorations improve the classical HMC-based ones on the other hand.

As perspectives for further work, we may mention the use of PMC in spatio-temporal segmentation problems, in a way similar to the method described in [2] in the case of HMC with Gaussian noise. It consists of considering a Hilbert-Peano scan defined on two successive images, as presented in Fig. 8. Of course, the same generalization could be used to segment three-dimensional data, e.g., medical images, by using the multidimensional Hilbert scanning algorithm provided in [46]. We may also mention the extension of PMC to multisensor image processing. In that model, both spatial and intersensor correlation should be taken into account at the same time. Another direc-

tion concerns the family of models called "hidden Markov trees" (HMTs), which can also be applied in image segmentation, particularly when multiresolution images are concerned [32], [33], [47]. This model, which gives fast and interesting results in image segmentation, has been recently generalized to a pairwise Markov tree model [48]. An unsupervised algorithm for parameter estimation has been proposed in [49], and first segmentation results show a benefit with respect to the HMT model for multiresolution signal processing.

## REFERENCES

- [1] W. Pieczynski, "Pairwise Markov chains," *IEEE Trans. Pattern Anal. Machine. Intell.*, vol. 25, pp. 634–639, May 2003.
- [2] B. Benmiloud and W. Pieczynski, "Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images" (in French), *Traitement du Signal*, vol. 12, no. 5, pp. 433–454, 1995.
- [3] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor HMC and unsupervised image segmentation," *IEEE Trans. Pattern Anal. Machine. Intell.*, vol. 19, pp. 465–475, May 1997.
- [4] A. El-Jacoubi, M. Gilloux, R. Sabourin, and C. Y. Suen, "An HMM-based approach for off-line unconstrained handwritten word modeling and recognition," *IEEE Trans. Pattern Anal. Machine. Intell.*, vol. 21, pp. 752–760, Aug. 1999.
- [5] K. Aas, L. Eikvil, and R. B. Huseby, "Applications of hidden Markov chains in image analysis," *Pattern Recogn.*, vol. 32, no. 4, pp. 703–713, 1999.
- [6] C. Raphael, "Automatic segmentation of acoustic musical signals using hidden Markov models," *IEEE Trans. Pattern Anal. Machine. Intell.*, vol. 21, pp. 360–370, Apr. 1999.
- [7] A. D. Wilson and A. F. Bobick, "Parametric hidden Markov models for gesture recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 884–900, Sept. 1999.
- [8] W. Pieczynski, J. Bouvrais, and C. Michel, "Estimation of generalized mixture in the case of correlated sensors," *IEEE Trans. Image Processing*, vol. 9, pp. 308–311, Feb. 2000.
- [9] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–286, Feb. 1989.
- [10] J. Dai, "Hybrid approach to speech recognition using hidden Markov models and Markov chains," *Proc. Inst. Elect. Eng.: Vision, Image Signal Process.*, vol. 141, no. 5, pp. 273–279, 1994.
- [11] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, pp. 2406–2413, July 1994.
- [12] G. A. Churchill, "Hidden Markov chains and the analysis of genome structure," *Comput. Chem.*, vol. 13, no. 2, pp. 107–115, 1992.
- [13] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statistic.*, vol. 41, pp. 164–171, 1970.

- [14] G. D. Fornay, "The viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–277, Mar. 1973.
- [15] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc.*, vol. 48, pp. 259–302, 1986.
- [16] R. Chellapa and A. Jain, *Markov Random Fields: Theory and Application*. New York: Academic, 1993.
- [17] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, June 1984.
- [18] P. A. Kelly, H. Derin, and K. D. Hartt, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1628–1641, Oct. 1988.
- [19] P. Pérez, "Markov random fields and images," *CWI Quart.*, vol. 11, no. 4, pp. 412–437, 1998.
- [20] W. Pieczynski and A.-N. Tebbache, "Pairwise Markov random fields and segmentation of textured images," *Machine Graphics Vision*, vol. 9, no. 3, pp. 705–718, 2000.
- [21] W. Pieczynski, "Pairwise Markov chains and Bayesian unsupervised fusion," in *Proc. 3rd Int. Conf. Information Fusion*, vol. 1, Paris, France, July 10–13, 2000, pp. 24–31.
- [22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 39, pp. 1–38, 1977.
- [23] W. Pieczynski, "Statistical image segmentation," *Machine Graphics Vision*, vol. 1, no. 1/2, pp. 261–268, 1992.
- [24] M. Mignotte, C. Collet, P. Pérez, and P. Bouthémy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Processing*, vol. 9, pp. 1216–1231, July 2000.
- [25] M. Mignotte, J. Meunier, J.-P. Soucy, and C. Janicki, "Comparison of deconvolution techniques using a distribution mixture parameter estimation: Application in SPECT imagery," *J. Electron. Imaging*, vol. 1, no. 1, pp. 11–25, 2002.
- [26] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised segmentation of radar images using hidden Markov chains and hidden Markov fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, pp. 675–686, Mar. 2003.
- [27] L. Fouque, A. Appriou, and W. Pieczynski, "Multiresolution hidden Markov chain model and unsupervised image segmentation," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Austin, TX, Apr. 2–4, 2000, pp. 121–125.
- [28] Z. Kato, J. Zeroubia, and M. Berthod, "Unsupervised parallel image classification using Markovian models," *Pattern Recogn.*, vol. 32, no. 4, pp. 591–604, 1999.
- [29] J.-P. Delmas, "An equivalence of the EM and ICE algorithm for exponential family," *IEEE Trans. Signal Processing*, vol. 45, pp. 2613–2615, Oct. 1997.
- [30] F. Salzenstein and W. Pieczynski, "Sur le choix de méthodes de segmentation statistique d'images" (in French), *Traitement du Signal*, vol. 15, no. 2, pp. 119–128, 1998.
- [31] J. Li, A. Najmi, and R. M. Gray, "Image classification by a two dimensional hidden Markov model," *IEEE Trans. Signal Processing*, vol. 48, pp. 517–533, Feb. 2000.
- [32] C. A. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 162–177, Mar. 1994.
- [33] H. Choi and R. G. Baraniuk, "Image segmentation using wavelet-domain classification," in *Proc. SPIE*, vol. 3816, 1999, pp. 306–320.
- [34] P. A. Devijver, "Baum's forward-backward algorithm revisited," *Pattern Recogn. Lett.*, vol. 3, pp. 369–373, 1985.
- [35] G. Celeux and G. Govaert, "A classification EM algorithm for clustering and two stochastic versions," *Computat. Statist. Data Anal.*, vol. 14, pp. 315–332, 1992.
- [36] Y. Delignon, A. Marzouki, and W. Pieczynski, "Estimation of generalized mixture and its application in image segmentation," *IEEE Trans. Image Processing*, vol. 6, pp. 1364–1375, Oct. 1997.
- [37] M. Zribi and F. Ghorbel, "An unsupervised and nonparametric Bayesian classifier," *Pattern Recogn.*, vol. 24, no. 1/3, pp. 97–112, 2003.
- [38] N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, second ed. New York: Wiley, 1994, vol. 1 and 2.
- [39] S. Derrode, G. Mercier, and J. M. L. et R. Garello, "Estimation of sea-ice SAR clutter statistics from Pearson's system of distributions," in *Proc. IEEE Geosci. Remote Sensing Symp.*, Sydney, Australia, July 9–13, 2001.
- [40] W. Skarbeck, "Generalized Hilbert scan in image printing," in *Theoretical Foundations of Computer Vision*, R. Klette and W. G. Kropetsch, Eds. Berlin, Germany: Akademik Verlag, 1992.
- [41] R. S. Parrish, "On an integrated approach to member selection and parameter estimation for Pearson distributions," *Comput. Statist. Data Anal.*, vol. 1, pp. 239–255, 1983.
- [42] R. Dafner, D. Cohen-Or, and Y. Matias, "Context-based space filling curves," *Comput. Graphics Forum*, vol. 19, no. 3, 2000.
- [43] J. Hungershofer and J. M. Wierum, "On the quality of partitions based on space-filling curves," in *Proc. Int. Conf. Comput. Sci.*, vol. LNCS 2331, Amsterdam, The Netherlands, 2002, pp. 36–45.
- [44] S. Derrode and W. Pieczynski, "SAR image segmentation using generalized pairwise Markov chains," in *Proc. SPIE Int. Symp. Remote Sensing*, Crete, Greece, Sept. 22–27, 2002.
- [45] T. Le Toan *et al.*, "Assessment of rice using ERS-1 SAR data : The SARI project (satellite assessment of rice in Indonesia)," in *Proc. RS GIS Environmental Resources Manage.*, Jakarta, Indonesia, 1995.
- [46] S. Kamata, R. O. Eason, and Y. Bandou, "A new algorithm for N-dimensional Hilbert scanning," *IEEE Trans. Image Processing*, vol. 8, pp. 964–973, July 1999.
- [47] J.-M. Laferté, P. Pérez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Trans. Image Processing*, vol. 9, pp. 390–404, Mar. 2000.
- [48] W. Pieczynski, "Arbres de Markov couple; pairwise Markov trees" (in French), *Compte Rendu de l'Académie des Sciences*, ser. I, vol. 335, pp. 79–82, 2002.
- [49] E. Monfrini and W. Pieczynski, "Segmentation non supervisée des images par les arbres de Markov couple" (in French), in *Actes du Colloque GRETSI*, Paris, France, Sept. 18–11, 2003.



**Stéphane Derrode** received the telecommunication engineering degree from Télécom Lille I, Lille, France in 1995 and the Ph.D. degree from the University of Rennes, Rennes, France, in 1999.

Since 2001, he has been with the Ecole Nationale Supérieure de Physique, Marseille, France, where he is currently an Associate Professor in the multidimensional signal processing group, Fresnel Institute (CNRS UMR-6133). His research interests include pattern recognition, image indexing, and segmentation.



**Wojciech Pieczynski** received the Doctorat d'Etat degree from the Université Pierre et Marie Curie, Paris, France, in 1986.

He is currently Professor at the Communications, Image, and Information Processing Department and the Coordinator of the Image Processing and its Applications (TAI) Third Year Major, Institut National des Télécommunications, Evry, France. His research interests include mathematical statistics, stochastic processes, and statistical image processing.