

# Extension of Higher-Order HMC Modeling with Application to Image Segmentation

Lamia Benyoussef, Cyril Carincotte and Stéphane Derrode

## Abstract

In this work, we propose to improve the neighboring relationship ability of the Hidden Markov Chain (HMC) model, by extending the memory lengths of both the Markov chain process and the data-driven densities arising in the model. The new model is able to learn more complex noise structures, with respect to the configuration of several previous states and observations. Model parameters estimation is performed from an extension of the general Iterative Conditional Estimation (ICE) method to take into account memories, which makes the classification algorithm unsupervised. The higher-order HMC model is then evaluated in the image segmentation context. A comparative study conducted on a simulated image is carried out according to the order of the chain. Experimental results on a Synthetic Aperture Radar (SAR) image show that higher-order model can provide more homogeneous segmentations than the classical model, but to the cost of higher memory and computing time requirements.

## Index Terms

Unsupervised Image segmentation, Higher-order Hidden Markov Chain, Iterative Conditional Estimation, Maximal Posterior Mode.

## I. INTRODUCTION

Since more than twenty years and the pioneering works by Geman, Besag, Marroquin and co-authors [1], [2], [3], Hidden Markov Random Fields (HMRF) have now a wide range of applications in the field of image analysis [4], [5], [6]. For example, in remote sensing applications, HMRF have been used to segment agricultural region, to detect linear structures such as road or coastline, or to map changes between images taken at two dates [7], [8], [9].

The interest in Markovian models is mainly due to the fact that when the unobservable process  $X$  can be modeled by a finite Markov model and when the noise is not too complex, then  $X$  can be recovered from the observed process  $Y$  by using Bayesian classification techniques such as Maximum A Posteriori (MAP) or Marginal Posterior Mode (MPM). In case of unsupervised classification, Markovian parameters have to be estimated from the observed data only. Well-known iterative methods like Estimation-Maximization (EM), stochastic EM (SEM) or Iterative Conditional Estimation (ICE) [10] can be used. In case of HMRF, all estimation methods require time-expensive “trial-and-error” procedures using Monte Carlo simulations [6], [11] (Metropolis algorithm or Gibbs sampler), resulting in a considerable computational burden, although methods, such as the mean-field like approximation, try to make EM-based estimation more tractable [12].

Recently, it has been shown that the Hidden Markov Chain (HMC) model [13], applied to image segmentation through a HilbertPeano scan [14] of the image (see Fig. 1), can often compete with HMRF based methods in terms of classification accuracy [15]. The HMC estimation stage is much faster than the HMRF one since update equations of internal parameters, *i.e.* regularity parameters made of the transition matrix and the data-driven parameters assumed Gaussian here, are analytic in case of EM [16] or require simulation without iterative procedures in case of SEM and ICE. However, HMRF provide a finer and more intuitive modeling of spatial relationships between neighboring pixels, through the system of cliques representing different topologies of contextual interaction.

To improve the spatial relationship capacity of HMC models, several works try to compensate for the model limitations due to the set of simplifying assumptions. We can cite semi-Markov models, experienced with success in speech processing [17], for which the probability of there being a change in the hidden state depends on the amount of time that has elapsed since entry into the current state. Other works try to relax the conditional independence assumption, such as the factorial HMM [18], the HMM2 [19] and the double Markov chain [20] models which used several Markov chains at different levels. We must also cite the pairwise Markov chain [21] and triplet Markov chain [22] models, for which the hidden process is not necessarily Markovian. But conditionally to the observed process it remains Markovian, still allowing Bayesian classification. Another

### L. Benyoussef,

GSM group, Institut Fresnel (CNRS UMR-6133), École Centrale de Marseille, Technopôle de Château-Gombert, 38, rue Frédéric Joliot Curie, 13451 Marseille Cedex 20, France.  
phone: +33 4 91 28 28 49; fax: +33 4 91 28 88 13; E-mail: lamia.benyoussef@ec-marseille.fr

### C. Carincotte,

Multitel Asbl, Rue Pierre et Marie Curie, 7000 Mons, Belgium.  
phone: +32 65 34 27 32; fax: +32 65 34 27 29; E-mail: carincotte@multitel.be

### S. Derrode, corresponding author

GSM group, Institut Fresnel (CNRS UMR-6133), École Centrale de Marseille, Technopôle de Château-Gombert, 38, rue Frédéric Joliot Curie, 13451 Marseille Cedex 20, France.  
phone: +33 4 91 28 28 49; fax: +33 4 91 28 88 13; E-mail: stephane.derrode@ec-marseille.fr

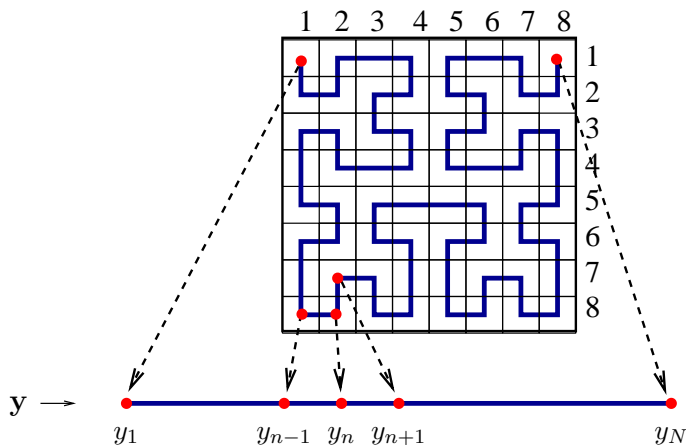


Fig. 1. Hilbert-Peano scan construction for an  $8 \times 8$  image. This scan is used to transform a 2D image into a 1D signal, and conversely.

solution consists in increasing the memory length  $R$  of the Markov chain process, *i.e.* each hidden state depends on the  $R$  previous states.  $\text{HMC}_R$  now denotes a hidden Markov process with a memory length of  $R$ ,  $\text{HMC}_1$  being the classical model. By the past, only  $\text{HMC}_2$  model has been evaluated for speech and handwritten recognition [23], [24], genomic [25] and robotic [26], with very limited improvements compared to  $\text{HMC}_1$ . This can be explained by the fact that any  $R$ -order hidden Markov chain can be transformed into an equivalent first order HMC, see for example [27].

The main novelty of this work is to increase the memory of the data-driven densities (noted  $S$ ) accordingly to the Markov chain memory length ( $S \leq R$ ). Hence, specific densities model the noise with respect to the  $R$  previous states configurations, allowing to model more complex noise structures. While the model is presented to deal with images, it is very general and of interest for all classical fields of application of HMC in including speech processing, genomic, economy, pattern recognition. . . . All algorithms are described in sufficient details to be implemented by readers that are familiar with statistical data restoration processing.

The remaining of the paper is organized as follows. Higher-order Markov chains and higher-order hidden Markov chains are briefly presented in Section II, together with the integration of an extended memory length for data-driven densities. Then, in order to achieve unsupervised MPM classification, an extension of the ICE procedure to those higher-order HMC models is described in Section III. Experimental results with a simulated image and a Synthetic Aperture Radar (SAR) image are presented in Section IV. Conclusion and perspectives are drawn in Section V.

## II. HIGH-ORDER HIDDEN MARKOV CHAINS

We start by recalling basic facts about higher-order Markov chains and by introducing notations. We next present higher-order hidden Markov chains and show how it is possible to relax the classical assumption on HMC to introduce a memory greater than one in the data-driven densities arising in the model.

### A. High-Order Markov Chains

To simplify notations,  $\mathbf{X}_{n:m}$ , with  $n \leq m$ , will denote the sequence of random variables  $\{X_n, \dots, X_m\}$ , and  $p(X = x)$  will be written  $p(x)$ . The process  $\mathbf{X} = \mathbf{X}_{1:N}$ , with  $X_n \in \Omega = \{1, \dots, K\}$  is a discrete and finite  $R$ -order Markov chain if and only if

$$p(x_{n+1} | \mathbf{x}_{1:n}) = p(x_{n+1} | \mathbf{x}_{n-R+1:n}). \quad (1)$$

Assuming further that the Markov chain is homogeneous, Eq. (1) does not depend on the index  $n$  and the distribution of  $\mathbf{X}$  is determined by

- the  $R$ -order transition matrix  $\mathbf{A}^R$ , valid for pixels  $n \geq R$ , whose elements are given by

$$a_{j|i_1, \dots, i_R}^R = p(X_{R+1} = j | X_1 = i_1, \dots, X_R = i_R),$$

where  $i_1, \dots, i_R$  and  $j$  belong to  $\Omega$ .

- the intermediate transitions matrices  $\mathbf{A}^n$ , for pixels  $1 \leq n < R$ , obtained by marginalizing  $\mathbf{A}^R$ . For  $n = R - 1$ , the intermediate transition matrix is defined by

$$a_{j|i_1, \dots, i_{R-1}}^{R-1} = p(X_R = j | X_1 = i_1, \dots, X_{R-1} = i_{R-1}) = \sum_{i_R=1}^K a_{j|i_1, \dots, i_R}^R.$$

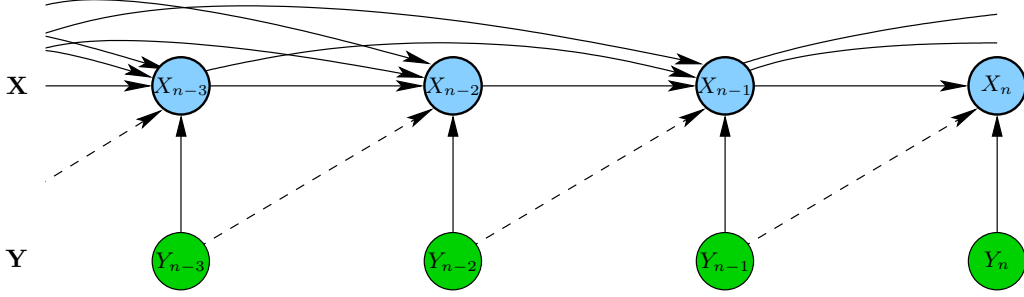


Fig. 2. Independence assumptions assumed in a  $\text{HMC}_2$  model. The dotted lines represent the new connection made by relaxing assumption **H\_2** (here  $S = 2$ ), illustrating model  $\text{HMC}_{2-2}$ .

- the distribution  $\pi$  of  $X_1$ , denoted by  $\pi_j = p(X_1 = j)$ .

The distribution of  $\mathbf{X}$  can then be written as

$$p(\mathbf{x}) = \pi_{x_1} \prod_{n=1}^{R-1} a_{x_{n+1}|x_{1:n}}^n \prod_{n=R}^{N-1} a_{x_{n+1}|x_{n-R+1:n}}^R.$$

### B. High-Order Hidden Markov Chains ( $\text{HMC}_R$ )

Using the Hilbert-Peano scan illustrated in Fig. 1, an image  $\mathbf{y} = \mathbf{y}_{1:N}$ ,  $N$  being the total number of pixels, is considered as a realization of the 1D observed process  $\mathbf{Y} = \mathbf{Y}_{1:N}$ , each  $Y_n \in \mathbb{R}$ . The segmented image  $\mathbf{x} = \mathbf{x}_{1:N}$  is considered as a realization of the  $R$ -order Markov chain  $\mathbf{X}$ .

With the following additional properties

- **H\_1** - Random variables  $Y_n$  are independent conditionally to  $\mathbf{X}$  :  $p(\mathbf{y} | \mathbf{x}) = \prod_{n=1}^N p(y_n | \mathbf{x})$ ,
- **H\_2** - The distribution of each  $Y_n$  conditionally on  $\mathbf{X}$  is equal to its distribution conditionally on  $X_n$  :  $p(y_n | \mathbf{x}) = p(y_n | x_n)$ ,

the distribution of the joint process  $(\mathbf{X}, \mathbf{Y})$  can be written as

$$p(\mathbf{x}, \mathbf{y}) = \pi_{x_1} f_{x_1}(y_1) \prod_{n=1}^{R-1} a_{x_{n+1}|x_{1:n}}^n f_{x_{n+1}}(y_{n+1}) \prod_{n=R}^{N-1} a_{x_{n+1}|x_{n-R+1:n}}^R f_{x_{n+1}}(y_{n+1}), \quad (2)$$

with  $f_k(y_n) = p(y_n | x_n = k)$  the data-driven densities, which are assumed Gaussian in this work (see [13] for generalized mixtures in  $\text{HMC}_1$ ).

Fig. 2 illustrates independence assumptions for a  $\text{HMC}_2$  model. The continuous lines connecting variables  $X_n$  represent the order of the Markov chain, while the continuous lines connecting  $Y_n$  with  $X_n$  symbolize assumption **H\_2**. This assumption is not strictly necessary and can be relaxed by considering additional connections of  $Y_n$  with  $X_{n-1}, X_{n-2}, \dots$

- **H\_bis** - The distribution of each  $Y_n$  conditionally on  $\mathbf{X}$  is equal to its distribution conditionally on  $\{X_{n-S+1}, \dots, X_n\}$ , with  $S \leq R$ .

The  $R$ -order hidden Markov model using hypothesis **H\_2 bis** will be denoted by  $\text{HMC}_{R-S}$ , with  $\text{HMC}_{1-1}$  denoting the classical model. This new assumption is schematized by the dotted lines in Fig. 2.

Parameter  $S$  is the memory length of the class-conditional densities of observation and should be less than  $R$  (see Appendix A). For  $n \geq S$ , their expression becomes

$$f_{x_{n-S+1:n}}(y_n) = p(y_n | \mathbf{x}_{n-S+1:n}). \quad (3)$$

For  $n < S$ ,  $f_{x_{n-S+1:n}}(y_n)$  is replaced by  $f_{\underbrace{x_n, \dots, x_n}_{S \text{ terms}}}(y_n)$ .

## III. MPM CLASSIFICATION AND UNSUPERVISED PARAMETERS ESTIMATION

We start by defining the Bayesian MPM classification rule for the  $\text{HMC}_{R-S}$  model and then propose an ICE-based parameters estimation method for unsupervised segmentation.

### A. MPM classification

The restoration of hidden  $\mathbf{X}$  from observed  $\mathbf{Y}$  can be done by applying the Bayesian MPM criterion :

$$\forall n \in [R, \dots, N], \quad \hat{\mathbf{x}}_n^{\text{MPM}}(\mathbf{y}) = \mathbf{z}_R \text{ with } \mathbf{z}_{1:R} = \arg \max_{\mathbf{x}_{n-R+1:n} \in \Omega^R} \xi_n(\mathbf{x}_{n-R+1:n}), \quad (4)$$

where  $\xi_n(\mathbf{x}_{n-R+1:n}) = p(\mathbf{x}_{n-R+1:n} | \mathbf{y})$  are the marginal *a posteriori* probabilities. As for eq. (4), all next equations will be given for  $n \geq R$  ; equations for  $n < R$  can be deduced easily.

The  $\text{HMC}_{R-S}$  model allows explicit computation of the MPM solution through an adaptation of the well-known Baum's "forward and backward" algorithm [28], modified by Devijver [29] for computational reasons:

$$\begin{aligned} \alpha_n(\mathbf{x}_{n-R+1:n}) &= p(\mathbf{x}_{n-R+1:n} | \mathbf{y}_{1:n}) \cdot \\ \beta_n(\mathbf{x}_{n-R+1:n}) &= \frac{p(\mathbf{y}_{n+1:N} | \mathbf{x}_{n-R+1:n})}{p(\mathbf{y}_{n+1:N} | \mathbf{y}_{1:n})}. \end{aligned}$$

Those probabilities can be computed recursively (see Appendix A). It can be shown that marginal *a posteriori* probabilities needed to compute the MPM solution can be written<sup>1</sup>

$$\xi_n(\mathbf{x}_{n-R+1:n}) = \alpha_n(\mathbf{x}_{n-R+1:n}) \beta_n(\mathbf{x}_{n-R+1:n}), \quad (5)$$

and, for latter use, joint *a posteriori* probabilities  $\Psi_n(\mathbf{x}_{n-R+1:n+1}) = p(\mathbf{x}_{n-R+1:n}, x_{n+1} | \mathbf{y})$  as:

$$\Psi_n(\mathbf{x}_{n-R+1:n+1}) \propto \alpha_n(\mathbf{x}_{n-R+1:n}) a_{x_{n+1} | \mathbf{x}_{n-R+1:n}}^R f_{x_{n-S+2:n+1}}(y_{n+1}) \beta_{n+1}(\mathbf{x}_{n-R+2:n+1}), \quad (6)$$

with 
$$\sum_{\mathbf{x}_{n-R+1:n+1} \in \Omega^{R+1}} \Psi_n(\mathbf{x}_{n-R+1:n+1}) = 1.$$

### B. ICE-based unsupervised parameters estimation

In case of unsupervised classification, both Markov parameters  $\mathcal{M}$  and data-driven parameters  $\mathcal{D}$  must first be estimated, only from the number of classes  $K$  and  $\mathbf{y}$ , the noisy image we want to segment:

- The set  $\mathcal{M}$  is made of the  $R$ -order transition matrix  $\mathbf{A}^R$ , and the initial probability vector  $\boldsymbol{\pi}$ .
- The set  $\mathcal{D}$ , characterizing the data-driven densities in Eq. (3), is made of the parameters of the  $K^S$  densities  $f_{\mathbf{x}_{n-S+1:n}}(y_n)$ . For  $R = S = 1$ , we find again the  $K$  densities  $f_{x_n}(y_n)$  of the classical  $\text{HMC}_{1-1}$  model. In the Gaussian case we consider here,  $\mathcal{D}$  is composed of means and variances of Gaussian pdf.

The estimation of all the parameters in  $\Theta = \{\mathcal{M}, \mathcal{D}\}$  can be achieved using the general ICE procedure [10], [13], [15], which is based on the conditional expectation of some estimators from the complete data  $(\mathbf{x}, \mathbf{y})$ . ICE is an iterative method which produces a sequence of estimations  $\theta^{[q]}$  of each parameter  $\theta \in \Theta$  as follows:

- 1) Initialize  $\theta^{[0]}$  by using empirical estimators from completed data  $(\mathbf{x}^{[0]}, \mathbf{y})$ , with  $\mathbf{x}^{[0]}$  a segmentation of  $\mathbf{y}$  obtained from K-means algorithm for example. Denoting  $\hat{m}_{i_1, \dots, i_S}$  and  $\hat{\sigma}_{i_1, \dots, i_S}^2$  the empirical estimates of the mean and variance of the Gaussian density  $f_{i_1, \dots, i_S}(y)$ , we have:

$$\hat{m}_{i_1, \dots, i_S} = \frac{\sum_{n=S}^N y_n \mathbf{1}_{[\mathbf{x}_{n-S+1:n}^{[0]} = (i_1, \dots, i_S)]}}{\sum_{n=S}^N \mathbf{1}_{[\mathbf{x}_{n-S+1:n}^{[0]} = (i_1, \dots, i_S)]}}. \quad (7)$$

$$\hat{\sigma}_{i_1, \dots, i_S}^2 = \frac{\sum_{n=S}^N (y_n - \hat{m}_{i_1, \dots, i_S})^2 \mathbf{1}_{[\mathbf{x}_{n-S+1:n}^{[0]} = (i_1, \dots, i_S)]}}{\sum_{n=S}^N \mathbf{1}_{[\mathbf{x}_{n-S+1:n}^{[0]} = (i_1, \dots, i_S)]}}. \quad (8)$$

Empirical estimates  $\hat{a}_{j|i_1, \dots, i_R}^R$  of  $R$ -order transition matrix entries can be computed from

$$\hat{\pi}_j = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{[x_n^{[0]} = j]}, \quad (9)$$

$$\hat{p}(i_1, \dots, i_R, j) = \frac{1}{N-R} \sum_{n=R}^{N-1} \mathbf{1}_{[\mathbf{x}_{n-R+1:n+1}^{[0]} = (i_1, \dots, i_R, j)]}. \quad (10)$$

<sup>1</sup>Note that this is an entry-by-entry product, not a matrix one.

- 2) Compute iteratively  $\theta^{[q+1]}$  using the mean of an estimator  $\hat{\theta}(\mathbf{X}, \mathbf{Y})$  of  $\theta$  defined with the completed data, conditionally to the observed data:

$$\theta^{[q+1]} = E \left[ \hat{\theta}(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}, \theta^{[q]} \right]. \quad (11)$$

- 3) Stop the algorithm at iteration  $Q$  if, for all parameters  $\theta$  in  $\Theta$ , we get  $\theta^{[Q-1]} \approx \theta^{[Q]}$ .

This procedure leads to two different situations:

- For parameters in  $\mathcal{M}$ , the conditional expectation can be computed analytically, similarly to the  $\text{HMC}_{1-1}$  case, by using the high-order *a posteriori* probabilities in eq. (5) and (6). Re-estimation formulae are given by

$$\begin{aligned} a_{j|i_1, \dots, i_R}^{R, [q+1]} &= \frac{\sum_{n=R}^{N-1} \psi_n^{[q]}(x_{n-R+1} = i_1, \dots, x_n = i_R, x_{n+1} = j)}{\sum_{n=R}^{N-1} \xi_n^{[q]}(x_{n-R+1} = i_1, \dots, x_n = i_R)} \\ \pi_j^{[q+1]} &= \frac{1}{N-R+1} \sum_{n=R}^N \sum_{\mathbf{x}_{n-R+2:n} \in \Omega^{R-1}} \xi_n^{[q]}(x_{n-R+1} = j, \mathbf{x}_{n-R+2:n}). \end{aligned} \quad (12)$$

Terms corresponding to  $n < R$  can be added to the sums, by using expressions of  $\xi$  and  $\psi$  for those indices, but their numerical contribution is very limited since  $N \gg R$ .

- For parameters in  $\mathcal{D}$ , the conditional expectation in (11) is not tractable. However, it can be estimated by computing, at each iteration  $q$ , the empirical mean of several estimates  $\theta^{[q+1]} = \frac{1}{\mathcal{L}} \sum_{\ell=1}^{\mathcal{L}} \hat{\theta}(\mathbf{x}_{[\ell]}^{[q]}, \mathbf{y})$  (computed using eq. (7) and (8), by replacing  $x^{[0]}$  by  $\mathbf{x}_{[\ell]}^{[q]}$ ), where  $\mathbf{x}_{[\ell]}$  is an *a posteriori* realization of  $\mathbf{X}$  conditionally on  $\mathbf{Y}$ .  $\mathbf{x}_{[\ell]}$  can be easily simulated using the fact that  $\mathbf{X} \mid \mathbf{Y}$  is a non homogeneous  $R$ -order Markov chain whose transitions matrices  $\tilde{\mathbf{A}}_n^R$ ,  $n \geq R$ , at iteration  $q+1$ , are given by (see Appendix A)

$$\begin{aligned} \tilde{a}_{x_{n+1}|x_{n-R+1:n}}^{R, [q+1]} &= p^{[q+1]}(x_{n+1} | \mathbf{x}_{n-R+1:n}, \mathbf{y}) \\ &= \frac{a_{x_{n+1}|x_{n-R+1:n}}^R f_{\mathbf{x}_{n-S+2:n+1}}(y_{n+1}) \beta_{n+1}(\mathbf{x}_{n-R+2:n+1})}{\sum_{x_{n+1} \in \Omega} a_{x_{n+1}|x_{n-R+1:n}}^R f_{\mathbf{x}_{n-S+2:n+1}}(y_{n+1}) \beta_{n+1}(\mathbf{x}_{n-R+2:n+1})}. \end{aligned} \quad (13)$$

*Remark:* Supervised learning of parameters is trivial, using eq. (7) to (9), but requires samples from homogeneous areas as well as samples exhibiting all kinds of transitions between regions. This is generally not the case in image processing since samples are made of regions of homogeneous pixels and do not include borders between classes. Hence, probabilities such as  $p(x_{n+1} = 2 \mid x_n = 1)$  and densities like  $p(y \mid x_n = 1, x_{n-1} = 2, \dots)$  can not be learned.

#### IV. EXPERIMENTAL RESULTS

This section is intended to give experimental results regarding the  $\text{HMC}_{R-S}$  model presented above. Emphasis is given to the impact of the value of  $R$  and  $S$  on the quality of unsupervised segmentation results. We first study a noisy simulated image and then an ERS-SAR<sup>2</sup> image showing an oil slick in the Mediterranean sea. Parameters initialization was performed with a K-means classifier, ICE algorithm was stopped after 35 iterations for the noisy simulated image and 60 iterations for the SAR image, and image classification was performed thanks to the Bayesian MPM criterion.

##### A. Noisy simulated image

Figure 3 shows the two classes original image, which is used latter as ground-truth, and the corresponding noisy simulated image. The noisy image was generated by adding an independent Gaussian noise to each class of the original image, defined by parameters  $(\mu_1 = 100, \sigma_1^2 = 60)$  and  $(\mu_2 = 105, \sigma_2^2 = 100)$ , and then correlating the noises by applying the  $3 \times 3$  smoothing filter given by

$$\begin{pmatrix} 0 & 0.6 & 0 \\ 0.6 & 1.0 & 0.6 \\ 0 & 0.6 & 0 \end{pmatrix}$$

Unsupervised segmentations of image in Fig. 3(b) are presented in Fig. 4, for different values of  $R$  and  $S$  (with  $S \leq R$ ). The resulting class images confirm the interest of the  $\text{HMC}_{R-S}$  model since segmentations proved to be much more accurate in

<sup>2</sup>ERS-SAR : *European Resource Sensing - Synthetic Aperture Radar*

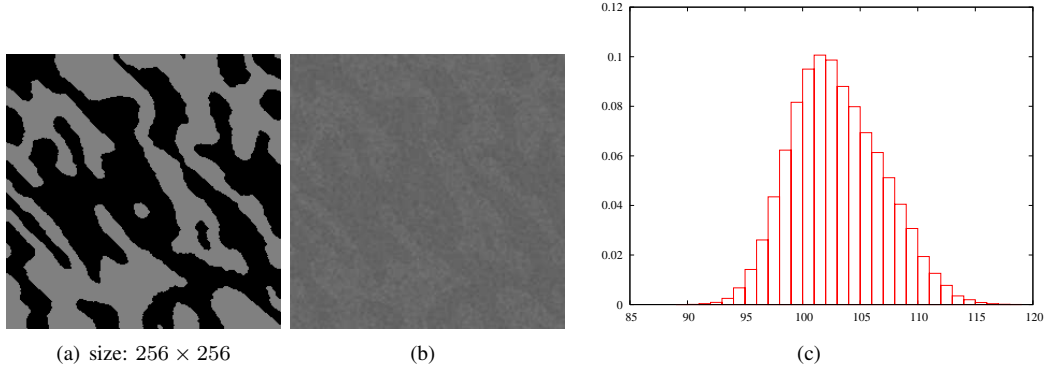


Fig. 3. Synthetic data: (a) two classes original image, (b) noisy simulated image and (c) its histogram.

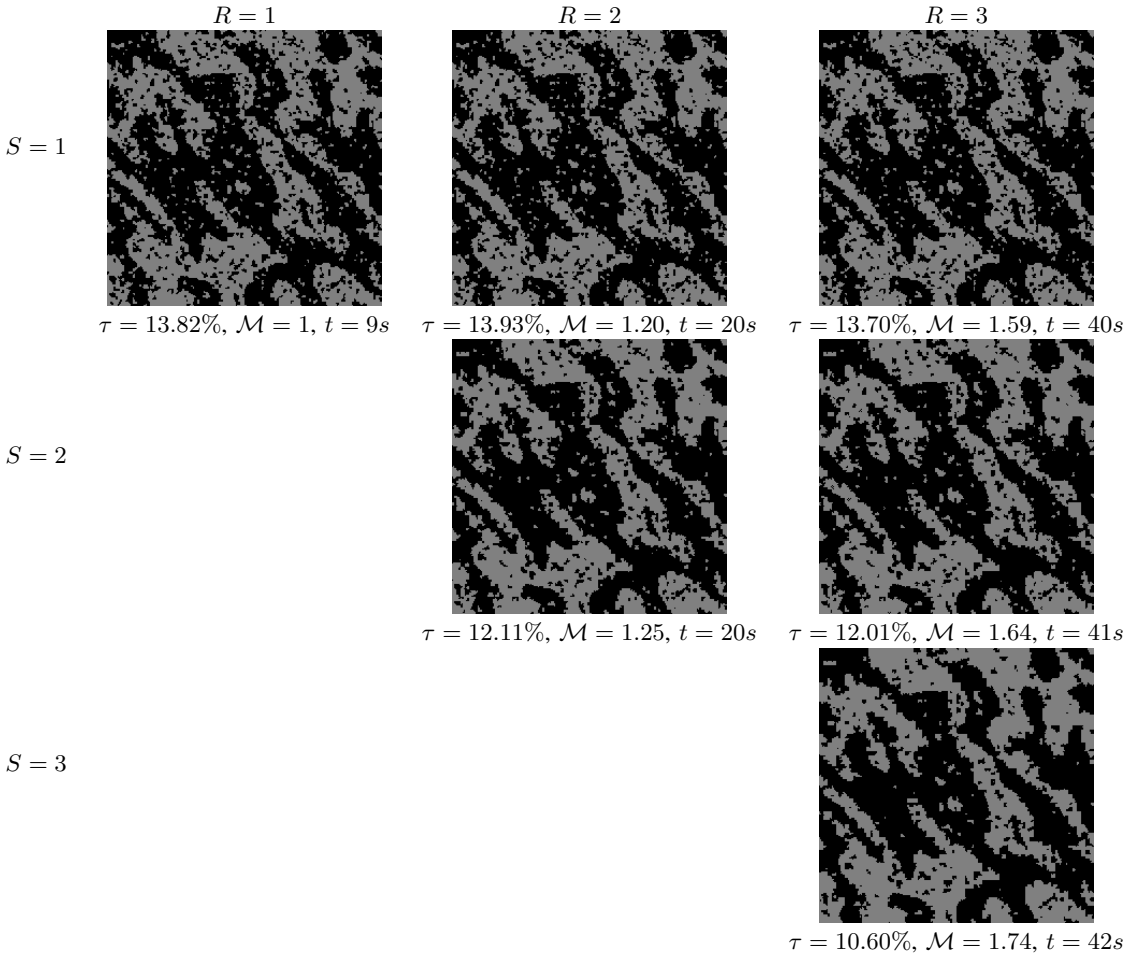


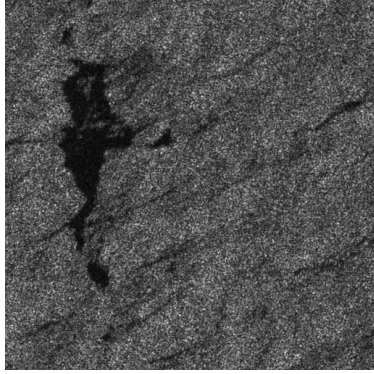
Fig. 4. Unsupervised segmentation results of the noisy image in Fig. 3.  $\tau$  is the segmentation error rate.  $\mathcal{M}$  refers to the memory required to estimate parameters with reference to the  $HMC_{1-1}$  model.  $t$  refers to the computing time (in seconds).

term of homogeneity when  $R$  and  $S$  increase. From misclassification rates ( $\tau$ ), memory requirements for parameter estimation ( $\mathcal{M}$ ) and computing times ( $t$ ), the following points can be underlined

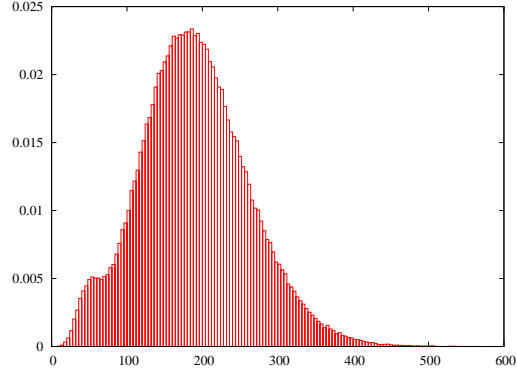
- The error rate does not improve with the increasing of  $R$  ( $S$  being constant). This confirms results published earlier with Markov chain of order 2 (*cf* [24] for an example in speech recognition). However, the error rate reduces sensibly when  $R = S$  increases, starting from  $\tau = 13.82\%$  for the classical  $HMC_{1-1}$  model to  $\tau = 10.60\%$  for the  $HMC_{3-3}$  model. The error rate still decreases, but more slowly, for higher values of  $R = S$ :  $\tau = 9.83\%$  (resp.  $\tau = 9.16\%$ ) for the  $HMC_{4-4}$  (resp.  $HMC_{5-5}$ ) model.
- The memory requirement increases exponentially with the raise of  $R$  and  $S$ . Starting from a reference of  $\mathcal{M} = 1$  for the classical  $HMC_{1-1}$  model, the  $HMC_{3-3}$  model reaches a factor of  $\mathcal{M} = 1.74$ , and even  $\mathcal{M} = 2.72$  (resp.  $\mathcal{M} = 4.68$ ) for

TABLE I  
ESTIMATED PARAMETERS FOR THE HMC<sub>2-2</sub> MODEL. LEFT: TRANSITION MATRIX  $\mathbf{A}^2$ . MIDDLE: INTERMEDIATE TRANSITION MATRIX  $\mathbf{A}^1$ . RIGHT: A PRIORI PROBABILITIES  $\boldsymbol{\pi}$ .

$x_{n-1}$	$x_n$	$x_{n+1}$	$a_{x_{n+1} x_{n-1},x_n}^2$	$x_n$	$x_{n+1}$	$a_{x_{n+1} x_n}^1$	$x_{n+1}$	$\pi_{x_{n+1}}$
1	1	1	0.97515	1	1	0.974623	1	0.557663
1	1	2	0.0248778	1	2	0.0254035	2	0.442337
1	2	1	0.0197173	2	1	0.0319929		
1	2	2	0.980283	2	2	0.967973		
2	1	1	0.954396					
2	1	2	0.0455927					
2	2	1	0.0323991					
2	2	2	0.967566					



(a) size:  $512 \times 512$



(b)

Fig. 5. ERS-SAR image of an oil slick in the Mediterranean sea (a) and its histogram (b).

the HMC<sub>4-4</sub> (resp. HMC<sub>5-5</sub>) model. As a consequence, large size images may cause memory overflow.

- The computing time also increases exponentially when  $R$  raises. On the other hand, the data-driven memory length  $S$  has nearly no impact on model complexity. For the HMC<sub>4-4</sub> (resp. HMC<sub>5-5</sub>) model, we get  $t = 92$  sec. (resp.  $t = 109$  sec.).

As an example, Table I shows all Markovian parameters estimated for the HMC<sub>2-2</sub> model which has given the segmented image in Fig. 4. The mean values of the data-driven Gaussian densities  $f_{1,1}(y)$ ,  $f_{1,2}(y)$ ,  $f_{2,1}(y)$  and  $f_{2,2}(y)$  were estimated to 99.96, 103.03, 103.06 and 105.69 respectively.

### B. Synthetic aperture radar image

Fig. 5 is an excerpt of an ERS-SAR image, acquired in the third of October 1992, showing an oil spill to be segmented from the free sea near the Egyptian coast. This scene is a typical spill in the Mediterranean sea. The spill appears in black in this SAR image since oil on the water reduces air-sea interaction and results in the dampening of the capillary (surface) waves [30]. To follow Goodman's approach and the multiplicative nature of noise in SAR data, we use Gamma laws instead of Gaussian ones. Fig. 6 shows the maps obtained from the segmentation with the classical HMC<sub>1-1</sub> model, and with the HMC<sub>2-2</sub> and HMC<sub>3-3</sub> models.

The HMC<sub>1-1</sub> model produces a map with a high level of false alarms, integrating numerous pixels from the free sea to the spill class. The HMC<sub>2-2</sub> segmentation is much more reliable and produce an accurate cartography of the spill with almost no detection error. Hence, in this image a memory length of two is able to learn the complex noise structure associated to the waves whereas a memory length of one is not. The HMC<sub>2-2</sub> result is similar to the one obtained by segmenting a multi-scale decomposition of the image with a vectorial HMC<sub>1-1</sub> model [31]. The HMC<sub>3-3</sub> model does not improve classification with respect to the HMC<sub>2-2</sub> one, confirming that a compromise must be find between the complexity of the model and the application.

## V. CONCLUSION

In this work we described an extension of higher-order Hidden Markov Chains for unsupervised classification. While the model is evaluated for image processing, the algorithms is of interest in all classical fields of application of HMC (speech processing, genomic, economy, pattern recognition...). The main novelty was to introduce a memory  $S$  for the data-driven densities arising in the model, in parallel to the memory of the Markov process  $R$ . An extension of the general ICE procedure was proposed in order to estimate all parameters, making the overall algorithm unsupervised. Classification was performed

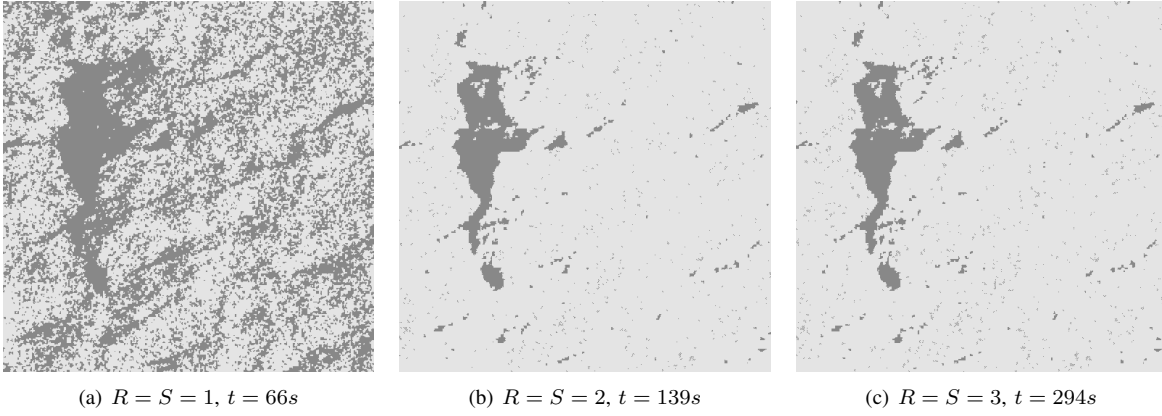


Fig. 6. Segmentation results obtained with the  $\text{HMC}_{R-S}$  model for different values of  $R = S$ .

with the Maximal Posterior Mode Bayesian criterion, while the extension of the Viterbi algorithm [16] to achieve Maximum A Posteriori classification is an interesting perspective to this work.

Experiments on simulated data and SAR images confirm the interest of the  $\text{HMC}_{R-S}$  model with respect to the  $\text{HMC}_{1-1}$  one. Indeed, the  $\text{HMC}_{R-S}$  model, which integrates more neighboring pixels, reveals very performing for modeling complex spatial relationships between pixels and, as a consequence, managing strong and possibly correlated noises. However, this is to the cost of higher memory and computing time requirements as  $R$  and  $S$  raise. In a near future, we plan to extent the Gaussian data-driven densities assumed in this work to deal with generalized mixture through the Pearson's system of distributions for example [13], [32].

## APPENDIX

### $\text{HMC}_{R-S}$ FORWARD RECURSIONS

$\text{HMC}_{R-S}$  forward probabilities can be computed recursively. For  $R < n \leq N$  and denoting

$$S_n = p(y_n | \mathbf{y}_{1:n-1}) = \sum_{\mathbf{x}_{n-R+1}, \dots, \mathbf{x}_n \in \Omega} p(\mathbf{x}_{n-R+1:n}, y_n | \mathbf{y}_{1:n-1}),$$

we get

$$\begin{aligned} \alpha_n(\mathbf{x}_{n-R+1:n}) &= p(\mathbf{x}_{n-R+1:n} | \mathbf{y}_{1:n}) = \frac{1}{S_n} p(\mathbf{x}_{n-R+1:n}, y_n | \mathbf{y}_{1:n-1}) \\ &= \frac{1}{S_n} \sum_{\mathbf{x}_{n-R} \in \Omega} p(\mathbf{x}_{n-R:n-1}, x_n, y_n | \mathbf{y}_{1:n-1}) \\ &= \frac{1}{S_n} \sum_{\mathbf{x}_{n-R} \in \Omega} p(x_n, y_n | \mathbf{x}_{n-R:n-1}, \mathbf{y}_{1:n-1}) p(\mathbf{x}_{n-R:n-1} | \mathbf{y}_{1:n-1}) \\ &= \frac{1}{S_n} \sum_{\mathbf{x}_{n-R} \in \Omega} p(y_n | \mathbf{x}_{n-R:n}, \mathbf{y}_{1:n-1}) p(x_n | \mathbf{x}_{n-R:n-1}, \mathbf{y}_{1:n-1}) \alpha_{n-1}(\mathbf{x}_{n-R:n-1}) \\ &= \frac{1}{S_n} p(y_n | \mathbf{x}_{n-R+1:n}) \sum_{\mathbf{x}_{n-R} \in \Omega} \alpha_{n-1}(\mathbf{x}_{n-R:n-1}) a_{x_n | \mathbf{x}_{n-R:n-1}}^R. \end{aligned}$$

Using assumption **H\_2 bis**,  $p(y_n | \mathbf{x}_{n-R+1:n})$  is written  $f_{\mathbf{x}_{n-S+1:n}}(y_n)$ , with  $S \leq R$ . For  $R = S = 1$ , we find the desired expression.

Recursions for  $n \leq R$  can be easily deduced :

- When  $R > 1$ , for  $1 < n \leq R$ :  $\alpha_n(\mathbf{x}_{1:n}) \propto f_{\mathbf{x}_{1:n}}(y_n) \alpha_{n-1}(\mathbf{x}_{1:n-1}) a_{x_n | \mathbf{x}_{1:n-1}}^{n-1}$ .
- For  $n = 1$ :  $\alpha_n(x_1) \propto f_{\mathbf{x}_{1:n}}(y_1) \pi_{x_1}$ .

When  $n < S$ ,  $f_{\mathbf{x}_{1:n}}(y_n)$  is replaced by  $\underbrace{f_{x_n, \dots, x_n}}_{S \text{ terms}}(y_n)$ .



## HMC<sub>R-S</sub> BACKWARD RECURSIONS

HMC<sub>R-S</sub> backward probabilities can be computed recursively. Using  $\beta_n(\mathbf{x}_{n-R+1:n}) = 1$  for  $n = N$ , we get for  $R \leq n < N$

$$\begin{aligned}
 \beta_n(\mathbf{x}_{n-R+1:n}) &= \frac{p(\mathbf{y}_{n+1:N} \mid \mathbf{x}_{n-R+1:n})}{p(\mathbf{y}_{n+1:N} \mid \mathbf{y}_{1:n})} \\
 &= \frac{1}{S_{n+1}} \frac{\sum_{x_{n+1} \in \Omega} p(\mathbf{y}_{n+1:N}, x_{n+1} \mid \mathbf{x}_{n-R+1:n})}{p(\mathbf{y}_{n+2:N} \mid \mathbf{y}_{1:n+1})} \\
 &= \frac{1}{S_{n+1}} \frac{\sum_{x_{n+1} \in \Omega} p(\mathbf{y}_{n+1:N} \mid \mathbf{x}_{n-R+1:n+1}) p(x_{n+1} \mid \mathbf{x}_{n-R+1:n})}{p(\mathbf{y}_{n+2:N} \mid \mathbf{y}_{1:n+1})} \\
 &= \frac{1}{S_{n+1}} \frac{\sum_{x_{n+1} \in \Omega} p(y_{n+1} \mid \mathbf{y}_{n+2:N}, \mathbf{x}_{n-R+1:n+1}) p(\mathbf{y}_{n+2:N} \mid \mathbf{x}_{n-R+1:n+1}) a_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R}{p(\mathbf{y}_{n+2:N} \mid \mathbf{y}_{1:n+1})} \\
 &= \frac{1}{S_{n+1}} \sum_{x_{n+1} \in \Omega} p(y_{n+1} \mid \mathbf{x}_{n-R+2:n+1}) \beta_{n+1}(\mathbf{x}_{n-R+2:n+1}) a_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R.
 \end{aligned}$$

Using assumption **H\_2 bis**,  $p(y_{n+1} \mid \mathbf{x}_{n-R+2:n+1})$  is written  $f_{\mathbf{x}_{n-S+2:n+1}}(y_{n+1})$ , with  $S \leq R$ . For  $R = S = 1$ , we find the desired expression.

Recursions for  $n < R$  can be easily deduced :

- When  $R > 1$ , for  $1 \leq n < R$ :  $\beta_n(\mathbf{x}_{1:n}) \propto \sum_{x_{n+1} \in \Omega} p(y_{n+1} \mid \mathbf{x}_{2:n+1}) \beta_{n+1}(\mathbf{x}_{2:n+1}) a_{x_{n+1} \mid \mathbf{x}_{1:n}}^n$ .

When  $n+1 < S$ ,  $f_{\mathbf{x}_{1:n+1}}(y_{n+1})$  is replaced by  $\underbrace{f_{x_{n+1}, \dots, x_{n+1}}(y_{n+1})}_{S \text{ terms}}$ .

## HMC<sub>R-S</sub> A POSTERIORI TRANSITION MATRICES

Entries  $\tilde{a}_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R$  of the *a posteriori* transition matrices  $\tilde{A}_n^R$  with  $n \geq R$  can be computed according to:

$$\begin{aligned}
 \tilde{a}_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R &= p(x_{n+1} \mid \mathbf{x}_{n-R+1:n}, \mathbf{y}) = p(x_{n+1} \mid \mathbf{x}_{1:n-R}, \mathbf{x}_{n-R+1:n}, \mathbf{y}) \\
 &= \frac{p(\mathbf{x}_{1:n-R}, \mathbf{x}_{n-R+1:n}, x_{n+1} \mid \mathbf{y})}{p(\mathbf{x}_{1:n-R}, \mathbf{x}_{n-R+1:n} \mid \mathbf{y})} \\
 &= \frac{\sum_{x_{n+2}, \dots, x_N \in \Omega} p(\mathbf{x} \mid \mathbf{y})}{\sum_{x_{n+1}, \dots, x_N \in \Omega} p(\mathbf{x} \mid \mathbf{y})} = \frac{\sum_{x_{n+2}, \dots, x_N \in \Omega} p(\mathbf{x}, \mathbf{y})}{\sum_{x_{n+1}, \dots, x_N \in \Omega} p(\mathbf{x}, \mathbf{y})}.
 \end{aligned}$$

By using the joint law given in eq. (2) and introducing the **H\_2 bis** assumption, we get

$$\tilde{a}_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R = \frac{a_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R f_{\mathbf{x}_{n-S+2:n+1}}(y_{n+1}) \sum_{x_{n+2}, \dots, x_N \in \Omega} a_{x_{n+2} \mid \mathbf{x}_{n-R+2:n+1}}^R f_{\mathbf{x}_{n-S+3:n+2}}(y_{n+2}) \cdots a_{x_N \mid \mathbf{x}_{N-R-N-1}}^R f_{\mathbf{x}_{N-S+1:N}}(y_N)}{\sum_{x_{n+1}, \dots, x_N \in \Omega} a_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R f_{\mathbf{x}_{n-S+2:n+1}}(y_{n+1}) \cdots a_{x_N \mid \mathbf{x}_{N-R-N-1}}^R f_{\mathbf{x}_{N-S+1:N}}(y_N)}.$$

By using the backward recursions above, we finally get

$$\tilde{a}_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R = \frac{a_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R f_{\mathbf{x}_{n-S+2:n+1}}(y_{n+1}) \beta_{n+1}(\mathbf{x}_{n-R+2:n+1})}{\sum_{x_{n+1} \in \Omega} a_{x_{n+1} \mid \mathbf{x}_{n-R+1:n}}^R f_{\mathbf{x}_{n-S+2:n+1}}(y_{n+1}) \beta_{n+1}(\mathbf{x}_{n-R+2:n+1})}.$$

For  $1 \leq n < R$ , entries  $\tilde{a}_{x_{n+1} \mid \mathbf{x}_{1:n}}^n$  can be deduced in a similar way.

## REFERENCES

- [1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [2] J. Besag, "On the statistical analysis of dirty pictures," *J. of the Royal Statistical Society*, vol. 48, pp. 259–302, 1986.
- [3] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of the Acoustical Society of America*, vol. 82, pp. 76–89, 1987.
- [4] H. Bunke, *Hidden Markov Models: Applications in Computer Vision*, ser. Machine Perception and Artificial Intelligence, H. Bunke and T. Caelli, Eds. World Scientific, June 2001, vol. 45.

- [5] B. Chalmond, *Modeling and Inverse Problems in Image Analysis*, ser. Applied Mathematical Sciences. Springer, 2003, vol. 155.
- [6] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*, 2nd ed., ser. Applications of Mathematics - Stochastic Modeling and Applied Probability. Springer, April 2003, vol. 27.
- [7] X. Descombes, M. Moctezuma, H. Maître, and J.-P. Rudant, "Coastline detection by a Markovian segmentation on SAR images," *Signal Processing*, vol. 55, no. 1, pp. 123–132, 1996.
- [8] P. Smits and S. Dellepiane, "Synthetic aperture radar image segmentation by a detail preserving Markov random field approach," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 844–857, 1997.
- [9] T. Kasetkasem and P. Varshney, "An image change detection algorithm based on Markov random field models," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, no. 8, pp. 1815–1823, 2002.
- [10] W. Pieczynski, "Statistical image segmentation," *Machine Graphics & Vision*, vol. 1, pp. 261–268, 1992.
- [11] L. Younes, "Estimation and annealing for Gibbsian fields," *Annales de l'Institut Henri Poincaré*, vol. 24, no. 2, pp. 269–294, 1988.
- [12] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, no. 1, pp. 131–144, 2003.
- [13] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor Hidden Markov Chain and unsupervised image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 465–475, 1997.
- [14] W. Skarbek, "Generalized Hilbert scan in image printing," in *Theoretical Foundations of Computer Vision*. Akademik Verlag, Berlin, 1992.
- [15] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised segmentation of radar images using Hidden Markov Chain and Hidden Markov Random Field," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 675–686, 2003.
- [16] L. R. Rabiner, "A tutorial on HMMs and selected applications in speech recognition," *Proc. of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [17] S. E. Levinson, "Continuously variable duration hidden Markov models for automatic speech recognition," *Comput. Speech Lang.*, vol. 1, no. 1, pp. 29–45, 1986.
- [18] Z. G. et M. I. Jordan, "Factorial hidden Markov models," *Machine Learning*, vol. 29, p. 245273, 1997.
- [19] K. Weber, S. Ikbal, S. Bengio, and H. Bourlard, "Robust speech recognition and feature extraction using HMM2," *Computer Speech & Language*, vol. 17, no. 2-3, pp. 195–211, 2003.
- [20] A. Berchtold, "The double chain Markov model," *Commun. Statist. Theory Methods*, vol. 28, pp. 2569–2589, 1999.
- [21] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise Markov chains," *IEEE Trans. on Signal Processing*, vol. 52, no. 9, pp. 2477–2489, 2004.
- [22] W. Pieczynski, "Multisensor triplet Markov chains and theory of evidence," *Int. J. of Approximate Reasoning*, vol. 45, no. 1, pp. 1–16, 2007.
- [23] E. de Villiers and J. du Preez, "The advantage of using higher-order HMM for segmenting acoustic files," in *Twelfth Annual Symposium of the Pattern Recognition Association of South Africa*, Franschhoek, South Africa, November 2001.
- [24] J. Mari, J. Haton, and A. Kriouille, "Automatic word recognition based on second-order HMMs," *IEEE Trans. on Speech and Audio Processing*, vol. 5, no. 1, pp. 22–25, January 1997.
- [25] R. Boys and D. Henderson, "A comparison of reversible jump MCMC algorithms for DNA sequence segmentation using HMMs," *Comp. Sci. and Stat.*, vol. 33, pp. 1–15, 2002.
- [26] O. Aycard, J.-F. Mari, and F. Washington, "Learning to automatically detect features for mobile robots using second-order HMMs," in *Proc. of the Workshop on Reasoning with Uncertainty in Robotics*, Acapulco, Mexico, July 2003.
- [27] F. Jelinek, *Statistical Methods for Speech Recognition*, ser. Language, Speech and Communication. MIT Press, Cambridge, MA, USA, 1997.
- [28] L. 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.
- [29] P. Devijver, "Baum's Forward-Backward algorithm revisited," *Pattern Recognition Letters*, vol. 3, pp. 369–373, 1985.
- [30] F. Girard-Ardhuin, G. Mercier, and R. Garelo, "Oil slick detection by SAR imagery: potential and limitation," in *Marine Technology and Ocean Science Conf. (OCEANS'03)*, San Diego, USA, 22-26 September 2003, pp. 22–26.
- [31] S. Derrode and G. Mercier, "Multiscale oil slick segmentation from SAR image using a vector HMC model," *Pattern Recognition*, vol. 40, no. 3, pp. 1135–1147, March 2007.
- [32] S. Derrode, G. Mercier, J.-M. LeCaillec, and R. Garelo, "Estimation of sea-ice SAR clutter statistics from Pearson system of distributions," in *IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS'01)*, Sydney, Australia, 9-13 July 2001.