

# CONTEXTUAL ESTIMATION OF HIDDEN MARKOV CHAINS WITH APPLICATION TO IMAGE SEGMENTATION

S. Derrode<sup>1,2\*</sup>

L. BenYoussef<sup>2</sup>

W. Pieczynski<sup>3</sup>

<sup>1</sup>GSM, Institut Fresnel (UMR 6133),  
EGIM Nord, Fac. Paul Cézanne,  
Dom. Univ. St Jérôme,  
13013 Marseille cedex 20, Fr.

stephane.derrode@egim-mrs.fr

<sup>2</sup>GRIFT, Lab. Cristal,  
Éc. Nat. des Sci. de l'Inform.,  
Camp. Univ. La Manouba,  
2010 Manouba, Tn.

by.lamia@planet.tn

<sup>3</sup>CITI Dpt (UMR 5157),  
Institut. Nat. des Télécom.,  
9, Rue Charles Fourier,  
91011 Evry Cedex, Fr.

wojciech.pieczynski@int-evry.fr

## ABSTRACT

This paper presents a contextual algorithm for the computation of Baum's forward and backward probabilities, which are intensively used in the framework of Hidden Markov Chain (HMC) models. The method differs from the original algorithm since it only takes into account a neighborhood of limited length and not all the chain for computations. Comparative experiments with respect to the neighborhood size have been conducted on both Markovian (simulations) and not Markovian (images) data, by mean of supervised and unsupervised classifications.

## 1. INTRODUCTION

In unsupervised image segmentation, the HMC model has shown to be very performing in a number of different contexts, such as in medical and satellite imaging. Before any processing, the bi-dimensional lattice of pixels is first converted into a 1D sequence of observations ( $\mathbf{y}$ ) through the Hilbert-Peano scan [1]. Then parameters estimation and Bayesian restoration techniques are applied to obtain a restored sequence of classes ( $\mathbf{x}$ ), which is finally converted back to a class image by inverting the scan [2].

For image segmentation, the HMC model structure can be considered somewhat artificial compared to the 2D structure of a Hidden Markov Random Field (HMRF), which is better adapted to model the relationship between adjacent pixels in an image. However, the computation time needed to estimate parameters is considerable and even prohibitive with this approach, even if mean field like approximations of HMRF make parameter estimation tractable without use of simulations [3]. Recently, both Markovian models have been compared in the context of satellite image segmentation and it appears that the HMC model constitutes a fast and robust alternative to the HMRF one since it can compete with in terms of estimation and classification accuracy [4].

All probabilities needed to estimate the HMC parameters in an unsupervised context can be expressed in term of the so-called Baum's *forward* and *backward* (F&B) probabilities [5]. Nevertheless, the memory requirement needed to save those probabilities can become prohibitive for large-size images or when the number of classes is high. This is for example the case with some high resolution satellite images whose size can exceed  $10000 \times 10000$  pixels, with up to 8/10 thematic classes. We propose here to study a contextual estimation of Baum's probabilities that takes into account a neighborhood of limited extent, unlike the original algorithm where the neighborhood is given by all the chain. This extension allows to understand the impact of far away data in the global HMC model. With this contextual estimation, we will be able to latter propose a bootstrap sub-sampling strategy for the HMC model, reducing both computing time and memory requirement.

The paper is organized as follows. Section 2 describes the methodological issues involved in the method, after basic facts about HMC were recalled. Section 3 and 4 present results from supervised and unsupervised classifications of noisy simulated Markov chain and image. Finally, conclusion and perspectives are drawn in section 5.

## 2. METHODOLOGY

### 2.1. Problem definition

Let  $\mathbf{X} = (X_1, \dots, X_N)$  and  $\mathbf{Y} = (Y_1, \dots, Y_N)$  be the vectors of random variables corresponding to the (hidden) state sequence and the observed sequence. Each  $X_n$  takes its values in a finite set  $\Omega = \{1, \dots, K\}$  of classes and each  $Y_n$  takes its values in the set of real numbers  $\mathbb{R}$ . Realizations of such processes will be denoted by lower-case letters  $\mathbf{x} = (x_1, \dots, x_N)$  and  $\mathbf{y} = (y_1, \dots, y_N)$ .

In this work,  $\mathbf{X}$  is supposed to be a stationary Markov chain, i.e. the probabilities of the  $\mathbf{C}$  matrix

$$c_{i,j} = P(X_n = i, X_{n+1} = j), \quad \forall i, j \in \Omega \quad (1)$$

\*ICASSP'06, Toulouse, France, May 15-19, 2006.

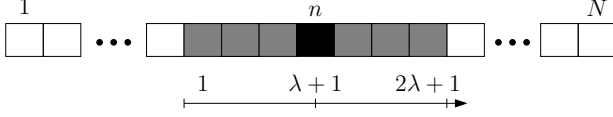


Fig. 1. Neighborhood used to estimate  $\alpha_n$  and  $\beta_n$ .

are independent of  $n$ . The distribution of  $\mathbf{X}$  is consequently determined by the initial distribution, denoted by  $\pi_i$ ,

$$\pi_i = P(X_1 = i) = \sum_{l=1}^K c_{i,l}$$

and the transition matrix  $\mathbf{A}$  whose entries are given by:

$$a_{i,j} = P(X_{n+1} = j | X_n = i) = \frac{c_{i,j}}{\pi_i}.$$

Under usual assumptions, i.e. (i) random variables  $Y_1, \dots, Y_N$  are conditionally independent with respect to  $\mathbf{X}$  and (ii) the distribution of each  $Y_n$  conditionally on  $\mathbf{X}$  is equal to its distribution conditional on  $X_n$ , we can write:

$$P(\mathbf{X}, \mathbf{Y}) = \pi_{x_1} f_{x_1}(y_1) \prod_{n=2}^N a_{x_{n-1}, x_n} f_{x_n}(y_n),$$

where  $f_{x_n}(y_n) = P(Y_n = y_n | X_n = x_n)$ .

One reason for the success of HMC models comes from the so-called Baum's F&B probabilities [5], revised by P. De Vijver in [6] to solve numerical problems that arise in their computation:

$$\begin{aligned} \alpha_n(i) &= P(X_n = i | \mathbf{Y}_{1 \rightarrow n} = \mathbf{y}_{1 \rightarrow n}), \\ \beta_n(i) &= \frac{P(\mathbf{Y}_{n+1 \rightarrow N} = \mathbf{y}_{n+1 \rightarrow N} | X_n = i)}{P(\mathbf{Y}_{n+1 \rightarrow N} = \mathbf{y}_{n+1 \rightarrow N} | \mathbf{Y}_{1 \rightarrow n} = \mathbf{y}_{1 \rightarrow n})}, \end{aligned}$$

where  $P(\mathbf{Y}_{a \rightarrow b}) = P(Y_a, \dots, Y_b)$ . The recursion algorithm is given by:

- forward initialization ( $1 \leq i \leq K$ ):

$$\alpha_1(i) = \frac{\pi_i f_i(y_1)}{\sum_{j=1}^K \pi_j f_j(y_1)} \quad (2)$$

- forward induction ( $1 \leq i \leq K, n = 2, \dots, N$ ):

$$\alpha_n(i) = \frac{f_i(y_n) \sum_{j=1}^K \alpha_{n-1}(j) a_{j,i}}{\sum_{l=1}^K f_l(y_n) \sum_{j=1}^K \alpha_{n-1}(j) a_{j,l}} \quad (3)$$

- backward initialization ( $1 \leq i \leq K$ ):

$$\beta_N(i) = 1 \quad (4)$$

- backward induction ( $1 \leq i \leq K, n = N - 1, \dots, 1$ ):

$$\beta_n(i) = \frac{\sum_{j=1}^K f_j(y_{n+1}) \beta_{n+1}(j) a_{i,j}}{\sum_{l=1}^K f_l(y_{n+1}) \sum_{j=1}^K \alpha_n(j) a_{j,l}} \quad (5)$$

## 2.2. The proposed contextual estimation

It can be observed that both F&B probabilities at index  $n$  depend on previous and next probabilities, illustrating the global behavior of Markov modeling. On the other hand, one can ask for the real numerical influence of the neighborhood away from  $n$  on the values of  $\alpha_n$  and  $\beta_n$ . Hence, it is well-known that, in a Markov chain, the correlation between two states decreases exponentially with the distance between them. To try to answer the question, we propose to consider the following algorithm which gives an estimation of F&B probabilities.

For index  $n$ , we consider a local neighborhood delimited by range  $[n - \lambda; n + \lambda]$  (see Fig. 1), and apply the F&B recursion algorithm (Eq. (2) to (5)) on  $\tilde{\mathbf{X}}_{1 \rightarrow 2\lambda+1} = \mathbf{X}_{n-\lambda \rightarrow n+\lambda}$  and  $\tilde{\mathbf{Y}}_{1 \rightarrow 2\lambda+1} = \mathbf{Y}_{n-\lambda \rightarrow n+\lambda}$  processes. We then only keep values  $\alpha_n(i) = \tilde{\alpha}_{\lambda+1}(i)$  and  $\beta_n(i) = \tilde{\beta}_{\lambda+1}(i)$ , discarding  $\tilde{\alpha}_\ell(i)$  and  $\tilde{\beta}_\ell(i)$  for  $1 \leq i \leq K$  and  $\forall \ell \in [1, 2\lambda + 1], \ell \neq \lambda + 1$ . This algorithm is repeated for all  $n$  from 1 to  $N$ , with special care on the firsts and lasts indices. Hence, we obtain a contextual estimation of F&B probabilities and the influence of the neighborhood on the recursion can be examined through the value of  $\lambda > 0$ . As  $\lambda$  increases, we can expect to get the approximative algorithm approaches the original one.

## 3. TESTS ON SIMULATED MARKOV CHAINS

In this section, we compare the algorithm proposed above with the original one, according to  $\lambda$ . The comparison is performed by mean of the classification error rate (the number of wrong classified samples over the total number of samples), after supervised and unsupervised restorations of noisy simulated Markov chains generated according to: (i) the simulation of a  $K = 2$  classes stationary Markov chain  $\hat{\mathbf{x}}$  with  $N = 10^4$  samples, from a given matrix  $\mathbf{C}$  (Eq. (1)); (ii) the degradation of  $\hat{\mathbf{x}}$  by adding a Gaussian white noises on each class, defined by  $\mathcal{N}(\mu_1, \sigma_1)$  and  $\mathcal{N}(\mu_2, \sigma_2)$  to get  $\hat{\mathbf{y}}$ .

### 3.1. MPM restoration with true parameters

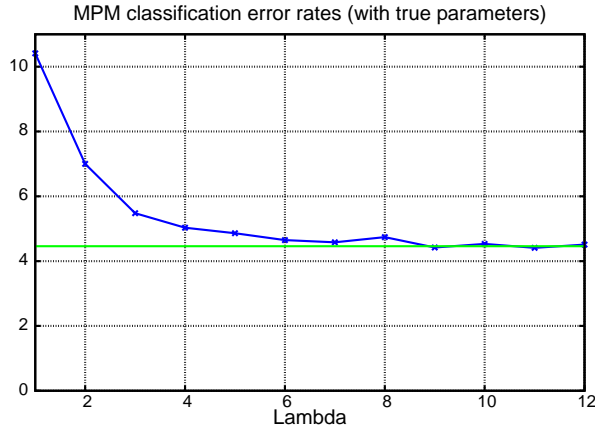
For supervised classification, we chose to restore the observations by mean of the Bayesian MPM (Maximal Posterior Marginal) decision rule since it can be directly expressed in terms of F&B probabilities:

$$\begin{aligned} [\hat{\mathbf{x}}_{\text{MPM}}(\mathbf{y}) = (\hat{x}_1, \dots, \hat{x}_N)] &\Leftrightarrow \\ \left[ \hat{x}_n = \arg \max_{x_n \in \Omega} P(X_n = x_n | \mathbf{Y}) \right] &. (6) \end{aligned}$$

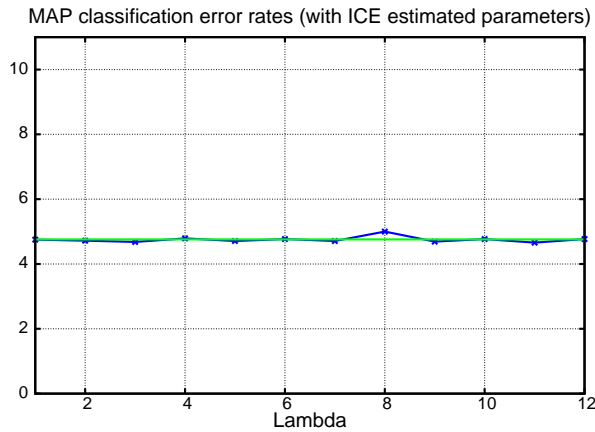
Hence, it becomes possible to measure the influence of  $\lambda$  on classification accuracy, by computing the error rate between  $\hat{\mathbf{x}}$  and  $\hat{\mathbf{x}}_{\text{MPM}}$ , and comparing it with the error rate obtained from the original algorithm.

From  $\hat{\mathbf{y}}$  and all model parameters used to generate it,

$$\mathbf{C} = \begin{pmatrix} 0.48 & 0.02 \\ 0.02 & 0.48 \end{pmatrix}, \quad \mathcal{N}(87, 14), \quad \mathcal{N}(113, 20),$$



**Fig. 2.** MPM error rates obtained with  $\lambda = 1, \dots, 12$  (mean values of 25 tests). The horizontal line shows the rate obtained with the original algorithm ( $\tau = 4.42\%$ ).



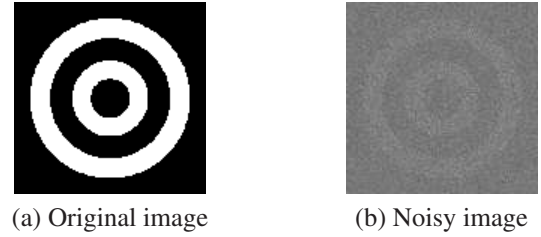
**Fig. 3.** MAP error rates, after ICE-based parameters estimation (mean values of 25 tests). The  $\tau = 4.75\%$  line corresponds to the original algorithm rate.

we computed  $\alpha_n$  and  $\beta_n$  for all  $1 \leq n \leq N$ , according to the procedure described in Sec. 2.2, for  $\lambda$  ranging from 1 to 12. We then applied Eq. (6) to get the restored sequence  $\hat{\mathbf{x}}_{\text{MPM}}$ . As shown in Fig. 2, the error rate decreases from  $\tau_1 = 10.35\%$  to  $\tau_{12} = 4.48\%$ . From  $\lambda \geq 6$ , the rates are very close to  $\tau = 4.42\%$  obtained with the original recursion algorithm, showing that the influence of the neighborhood is spatially limited.

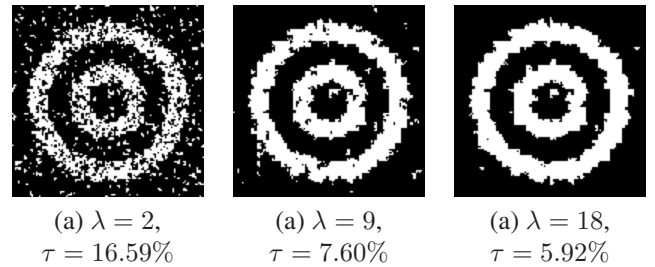
### 3.2. MAP restoration with estimated parameters

We can now ask for the influence of  $\lambda$  on the estimation of model parameters, in the context of unsupervised classification. To that goal, we decided to achieve MAP (Maximum A Posteriori) restoration

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg \max_{\mathbf{x} \in \Omega^N} P(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}),$$



**Fig. 4.** Before processing, image (b) is transformed into a 1D sequence by mean of the Hilbert-Peano scan.



**Fig. 5.** MPM segmented images for different values of  $\lambda$ .

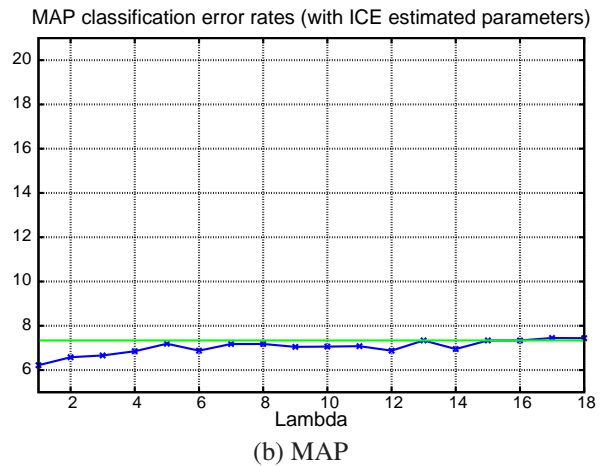
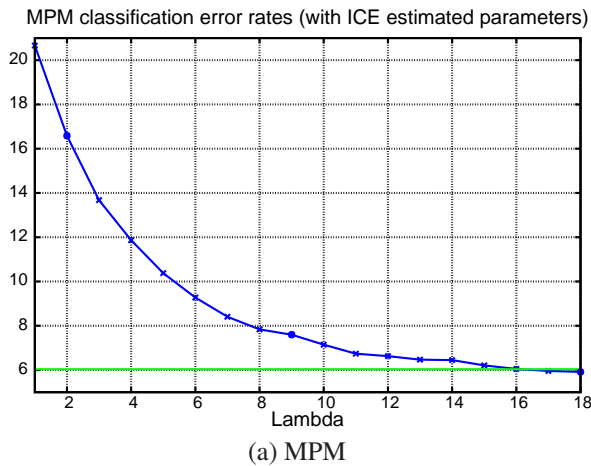
using Viterbi's algorithm. Indeed, F&B probabilities do not play a direct role in MAP restoration (unlike MPM), but only by the way parameters are estimated. Precisely, parameters estimation is done by the ICE (Iterative Conditional Estimation) procedure [7], which is an alternative to EM (Estimation-Maximisation) and SEM (Stochastic EM) algorithms.

Results are drawn on Fig. 3, after 50 ICE iterations and an initial estimation computed from a Kmeans-based segmentation. It appears that  $\lambda$  has nearly no influence in parameters estimation, since the error rate is nearly constant. These results have been confirmed by other experiments (not reported here) where we have modified both  $\mathbf{C}$  matrix and Gaussian noise parameters, making the mixture more difficult to estimate. Next section presents experiments with noisy simulated images, in order to determine if the same behavior can be observed when the data do not follow a Markov chain.

## 4. TESTS ON NOISY SIMULATED IMAGES

To pursue the analysis of the algorithm, we have segmented the noisy simulated image in Fig. 4(b), generated from the (a) image by adding a Gaussian white noise on each class. The graphs in Fig. 6 report the error rates obtained by using unsupervised MPM and MAP restorations, after 50 ICE iterations. We can observe the same behavior as in the case of Markov chains (section 3), but with these additional comments:

(1) In the MPM restoration case, the size of the neighborhood needed to get an error rate close to the one given by the original algorithm ( $\tau = 6.04\%$ ) is much larger than for simulated data in previous section ( $\lambda \simeq 16$  vs  $\lambda \simeq 6$ ). For illustration purposes, Fig. 5 shows three MPM-segmented im-



**Fig. 6.** MPM and MAP error rates after ICE-based parameters estimation. The rates computed from the original algorithm are respectively given by  $\tau = 6.04\%$  and  $\tau = 7.34\%$ . Several MPM based segmentations are shown in Fig. 5.

ages obtained by constructing back the image using an invert Peano scan.

(2) The error rate given by MAP restoration is always better than the rate obtained with the original algorithm ( $\tau = 7.34\%$ ). This can be understood by recalling that the Peano scan of the target image is not Markovian and the approximative algorithm, by its contextual approach, can better take into account the difference to the underlying model (Markovian assumption) than the original algorithm. Nevertheless, as expected, the rates tend to be equal when  $\lambda$  increases.

## 5. CONCLUSION

This work describes an algorithm for the contextual estimation of forward and backward probabilities. It consists in taking into account only a neighborhood of limited extent in the computation of probabilities, and not all the chain as it is done in the original algorithm. Hence, we break the global nature of the HMC model, while preserving that it is still governed by only one set of parameters (unlike contextual models).

We have conducted a series of experiments on the size of the neighborhood, from what we can make the three following conclusions: (i) The approximation of F&B probabilities has a small impact on parameters estimation (a neighborhood of only one sample seems enough in our simulations). However, the MPM criterion seems much more sensitive than the MAP one. (ii) The computation is valid for both Markovian and not Markovian data, i.e. the algorithm tends to have the same behavior than the original algorithm when  $\lambda$  increases. (iii) The optimal size of the neighborhood depends on data.

It is worth noting that such an algorithm results in an increase of computational time (which is proportional to  $\lambda$ ), and has no other interest than to analyse the role of the neighborhood in an HMC context. Nevertheless, this algorithm will allow to propose in a near futur a data-driven bootstrap sub-

sampling strategy for HMC, which will result in a reduction of computational time and memory requirement. Hence, we will be able to work on large images such as those encountered in remote sensing imagery.

## 6. REFERENCES

- [1] W. Skarbek, *Generalized Hilbert scan in image printing*, chapter Theoretical Foundations of Computer Vision, Akademie Verlag, Berlin, Germany, R. Klette and W. G. Kropfsh edition, 1992.
- [2] B. Benmiloud and W. Pieczynski, “Estimation des paramètres dans les chaînes de Markov cachées et segmentation d’images,” *Traitement du Signal*, vol. 12, no. 5, pp. 433–454, 1995, in French.
- [3] 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.
- [4] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, “Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields,” *IEEE Trans. on Geosci. and Rem. Sens.*, vol. 41, no. 3, pp. 675–686, March 2003.
- [5] 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. 164171, 1970.
- [6] P. Devijver, “Baums forward-backward algorithm revisited,” *Patt. Recog. Lett.*, vol. 3, pp. 369–373, 1985.
- [7] W. Pieczynski, “Statistical image segmentation,” *Mach. Graph. and Vis.*, vol. 1, no. 1/2, pp. 261–268, 1992.