

# UNSUPERVISED IMAGE SEGMENTATION BASED ON HIGH-ORDER HIDDEN MARKOV CHAINS

*S. Derrode, C. Carincotte and S. Bourennane*

Multidimensional Signal Processing Group, Institut Fresnel (CNRS-UMR 6133)  
Domaine universitaire de Saint Jérôme, 13013 Marseille Cedex 20 - FRANCE  
cyril.carincotte@fresnel.fr

## ABSTRACT

First order hidden Markov models have been used for a long time in image processing, especially in image segmentation. In this paper, we propose a technique for the unsupervised segmentation of images, based on high-order hidden Markov chains. We also show that it is possible to relax the classical hypothesis regarding the state observation probability density, which allows to take into account some particular correlated noise. Model parameter estimation is performed from an extension of the general Iterative Conditional Estimation (ICE) method that takes into account the order of the chain. A comparative study conducted on a simulated image is carried out according to the order of the chain. Experimental results on Synthetic Aperture Radar (SAR) images show that the new approach can provide a more homogeneous segmentation than the classical one, implying higher complexity algorithm and computation time.

## 1. INTRODUCTION

The aim of this paper is to compare the high-order Hidden Markov Chain model (denoted by HMC- $R$ , with  $R$  the order of the Markov chain or the memory length) with the classical HMC-1 model for the unsupervised segmentation of images.

The HMC-1 model has been used successfully in image segmentation [1], thanks to the use of a Hilbert-Peano scan that converts the 2D lattice into a 1D sequence [2]. The success of HMC models is due to the fact that when the unobservable signal process  $\mathbf{X}$  can be modelled by a finite Markov chain and when the noise is not too complex, then  $\mathbf{X}$  can be recovered from the observed process  $\mathbf{Y}$  using different Bayesian classification techniques like Maximum A Posteriori (MAP) or Maximal Posterior Mode (MPM). Recently, it has been shown that the HMC-1 model can compete with Hidden Markov Random Field (HMRF) based methods in terms of classification accuracy, while being much faster, even though the latter provides a finer and more intuitive modelling of spatial relationships [3].

High-order Markov chains, especially HMC-2, have been used in a number of applications, including speech and handwritten recognition [4, 5], genomic [6] and robotic [7]. However, to our knowledge, HMC- $R$  model has not been tested in unsupervised image segmentation. This model can be of interest since increasing the memory of the Markov process allows to model more complex spatial relationships between pixels and so more complex noise structures.

The paper is organized as follows: high-order Markov chain structure is presented in Section 2. We specify in Section 3 the straightforward extension of the HMC-1, inspired by [5] and applied for image segmentation. The unknown HMC- $R$  parameters estimation, achieved with an extension of the general ICE method [1, 3], which can be seen as an alternative to well-known Estimation-Maximization (EM) algorithm, is then briefly presented. We also present in this Section a new approach which consists in taking into account the order of the chain for the estimation of the conditional observation probability density. Comparative results on simulated and SAR images are presented in Section 4, whereas conclusions are drawn in Section 5.

## 2. HIGH-ORDER MARKOV CHAINS

To simplify notations,  $\mathbf{X}_{1 \rightarrow n}$  will denote the sequence of random variables  $\{X_1, \dots, X_n\}$  and  $\mathbf{x}$  will denote a realization of process  $\mathbf{X}$ .

$\mathbf{X} = \{X_n\}_{n \in \{1, \dots, N\}}$  is a  $R$ -order Markov chain, with length  $N$ , and with each  $X_n$  taking its value in the set of classes  $\Omega = \{1, \dots, K\}$  if and only if:

$$\begin{aligned} P(X_n = x_n \mid \mathbf{X}_{1 \rightarrow n-1} = \mathbf{x}_{1 \rightarrow n-1}) \\ = P(X_n = x_n \mid \mathbf{X}_{n-R \rightarrow n-1} = \mathbf{x}_{n-R \rightarrow n-1}). \end{aligned} \quad (1)$$

Actually, it means that each component only depends on the  $R$  immediately previous ones. Such a Markov chain is said homogeneous if Eq. (1) does not depend on the position  $n$  in the sequence. This leads to the set of state transition probabilities of high-order of the form:

$$t_{\mathbf{x}_{n-R \rightarrow n}} = P(X_n = x_n \mid \mathbf{X}_{n-R \rightarrow n-1} = \mathbf{x}_{n-R \rightarrow n-1}),$$

$\forall n \in \{R+1, \dots, N\}$ , with the state transition coefficients having the properties:

$$t_{x_{n-R} \rightarrow x_n} \geq 0, \quad \sum_{x_n=1}^K t_{x_{n-R} \rightarrow x_n} = 1.$$

All these probabilities are contained in a  $(R+1)$ -dimensions transition probabilities matrix  $\mathbf{T} = \{t_{x_{n-R} \rightarrow x_n}\}$ .

It is important to note that  $R$ -order Markov chains are also defined by  $R-1$  matrices characterizing the  $R$  first transitions in the sequence:

- $n = R : \mathbf{T}^{R-1} = \{t_{x_{1 \rightarrow R}}^{R-1}\}, \forall x_{1 \rightarrow R} \in \Omega^R,$
- $\dots,$
- $n = 3 : \mathbf{T}^2 = \{t_{x_{1 \rightarrow 3}}^2\}, \forall x_{1 \rightarrow 3} \in \Omega^3,$
- $n = 2 : \mathbf{T}^1 = \{t_{x_{1 \rightarrow 2}}^1\}, \forall x_{1 \rightarrow 2} \in \Omega^2.$

Finally, for  $n = 1$ , we get the initial state probabilities  $\pi_{x_1} = P(X_1 = x_1), \forall x_1 \in \Omega.$

### 3. HIGH-ORDER HIDDEN MARKOV CHAINS

HMC-based image segmentation methods assume that each component of the observation vector  $\mathbf{y} = \{y_1, \dots, y_N\}$  can be modelled as states of an underlying Markov chain  $\mathbf{X}$ .

In this section, we investigate models in which the underlying states sequence is an homogeneous  $R$ -order Markov chain. Similarly to the HMC-1 context, we first consider the usual two following assumptions:

- $\mathbf{H}_1$ : the random variables  $Y_1, \dots, Y_N$  are independent conditionally on  $\mathbf{X}$ .
- $\mathbf{H}_2$ : the distribution of each  $Y_n$  conditionally on  $\mathbf{X}$  is equal to its distribution conditionally on  $X_n$ .

Fig. 1 illustrates assumption  $\mathbf{H}_2$  for a HMC-2 model. The continuous lines of the process  $\mathbf{X}$  represent the order of the HMC:  $X_{n+1}$  is attached to  $X_n$  and  $X_{n-1}$ . The continuous lines connecting  $\mathbf{Y}$  with  $\mathbf{X}$  symbolize  $\mathbf{H}_2$ : each  $Y_n$  is linked with the corresponding  $X_n$ . This assumption will be relaxed in Section 3.3.

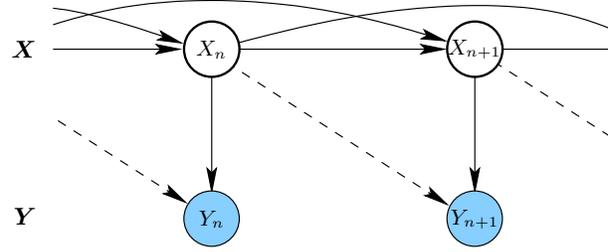
#### 3.1. HMC- $R$ model

As specified above, let  $\mathbf{X} = \mathbf{X}_{1 \rightarrow N}$  be an homogeneous  $R$ -order Markov chain, corresponding to the unknown class image. We get:

$$P(\mathbf{X} = \mathbf{x}) = \pi_{x_1} \prod_{r=1}^{R-1} t_{x_{1 \rightarrow r+1}}^r \prod_{n=R+1}^N t_{x_{n-R} \rightarrow x_n}.$$

Each state of the state space is associated with a distribution, characterizing the repartition of observations:

$$f_{x_n}(y_n) = P(Y_n = y_n | X_n = x_n). \quad (2)$$



**Fig. 1.** Independence assumptions assumed in a HMC-2 model. The dotted lines represent the new relation introduced by the more general assumption  $(\mathbf{H}_2^R)$ , see text in Section 3.3.

Given an observed sequence  $\mathbf{y} = \mathbf{y}_{1 \rightarrow N}$ , we can compute the joint state-observation probability density by:

$$P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}) = \pi_{x_1} f_{x_1}(y_1) \prod_{r=1}^{R-1} t_{x_{1 \rightarrow r+1}}^r f_{x_{r+1}}(y_{r+1}) \prod_{n=R+1}^N t_{x_{n-R} \rightarrow x_n} f_{x_n}(y_n). \quad (3)$$

In the case of unsupervised classification, the distribution  $P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})$  is unknown and must be estimated in order to apply a Bayesian classification criterion. Therefore we have to estimate the following sets of parameters:

- The set  $\mathbf{\Gamma}$  characterizing the homogeneous  $R$ -order Markov chain, i.e. the initial probability vector  $\boldsymbol{\pi} = (\pi_{x_1})_{\forall x_1 \in \Omega}$ , the  $R-1$  intermediate transition matrices  $\mathbf{T}^1, \dots, \mathbf{T}^{R-1}$  and the  $R$ -order transition matrix  $\mathbf{T}$ .
- The set  $\mathbf{\Delta}$  characterizing the conditional observations density presented in Eq. (2), i.e. the parameters of the  $K$  distributions  $f_k$ . In the Gaussian case,  $\mathbf{\Delta}$  is composed of the means and the variances.

#### 3.2. Parameters estimation

The estimation of all the parameters in  $\Theta = \{\mathbf{\Gamma}, \mathbf{\Delta}\}$  can be achieved using the general ICE algorithm [1, 3]. The ICE procedure is based on the conditional expectation of some estimators from the complete data  $(\mathbf{x}, \mathbf{y})$ . It is an iterative method which produces a sequence of estimations  $\theta^q$  of parameter  $\theta$  as follows: (1) initialize  $\theta^0$ , (2) compute  $\theta^{q+1} = E_q[\hat{\theta}(\mathbf{X}, \mathbf{Y}) | \mathbf{Y} = \mathbf{y}]$ , where  $\hat{\theta}(\mathbf{X}, \mathbf{Y})$  is an estimator of  $\theta$ . In practice, we stop the algorithm at iteration  $Q$  if  $\theta^{Q-1} \approx \theta^Q$ . This procedure leads to two different situations:

- For parameters in  $\mathbf{\Delta}$ ,  $\theta^{q+1}$  is not tractable. However, it can be estimated by computing the empirical mean of several estimates according to  $\theta^{q+1} = \frac{1}{L} \sum_{l=1}^L \hat{\theta}(\mathbf{x}^l, \mathbf{y})$ , where  $\mathbf{x}^l$  is an *a posteriori* realization of  $\mathbf{X}$  conditionally on  $\mathbf{Y}$ . It can be shown that  $\mathbf{X} | \mathbf{Y}$  is a non homogeneous Markov

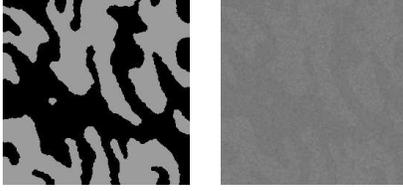


Fig. 2. Original image and noisy simulated one.

chain whose parameters can be computed with the high-order normalized Baum-Welch algorithm.

- For parameters in  $\Gamma$ , the expectation can be computed analytically, similarly to the HMC-1 case, by using the high-order normalized Baum-Welch algorithm.

### 3.3. Relaxing hypothesis $H_2$

It can be easily shown that assumption  $H_2$  is not strictly necessary and can be relaxed to some extent:

- $H_2^R$ : the distribution of each  $Y_n$  conditionally on  $\mathbf{X}$  is equal to its distribution conditionally on  $(X_n, X_{n-1}, \dots, X_{n-R+1})$  for  $\mathbf{X}$  being a  $R$ -order Markov chain,

This assumption is less limitative and is sufficient in the relations involved in the extended Baum-Welch algorithm.

Fig. 1 illustrates these two assumptions for a HMC-2 model. Continuous and dotted lines connecting  $\mathbf{Y}$  with  $\mathbf{X}$  now symbolize  $H_2^R$ : each  $Y_n$  is linked with the corresponding  $X_n$  (continuous) and the previous one  $X_{n-1}$  (dotted).

For a  $R$ -order Markov chain, the expression of the conditional probability of the observation (Eq. (2)) becomes:

$$f_{\mathbf{x}_{n-R+1 \rightarrow n}}(y_n) = P(Y_n = y_n | \mathbf{X}_{n-R+1 \rightarrow n} = \mathbf{x}_{n-R+1 \rightarrow n}). \quad (4)$$

This kind of model will be denoted HMC- $R_1(R_2)$ . For example, HMC- $R_1(1)$  is the “classical”  $R_1$ -order case, and HMC- $R_1(R_2)$  denote a segmentation with a HMC- $R_1$  and a state observation probabilities of order  $R_2$  ( $R_2 \leq R_1$ ).

## 4. EXPERIMENTAL RESULTS

Classical HMC-1 and HMC- $R$  have been comparatively assessed on two different images. Actually, in both cases, parameters initialization was done with a fuzzy C-means classifier. The ICE algorithm was stopped after fifty iterations, assuming it has converged, and the image classification was performed thanks to the Bayesian MPM criterion for the simulated image and with the MAP criterion for the SAR one.

Experimentally, we observed that the standard deviations (std) associated with non-homogeneous classes (e.g. classes “101”, “001”, ... for a HMC-3(3)) were generally under-estimated. So we decided to artificially increase these

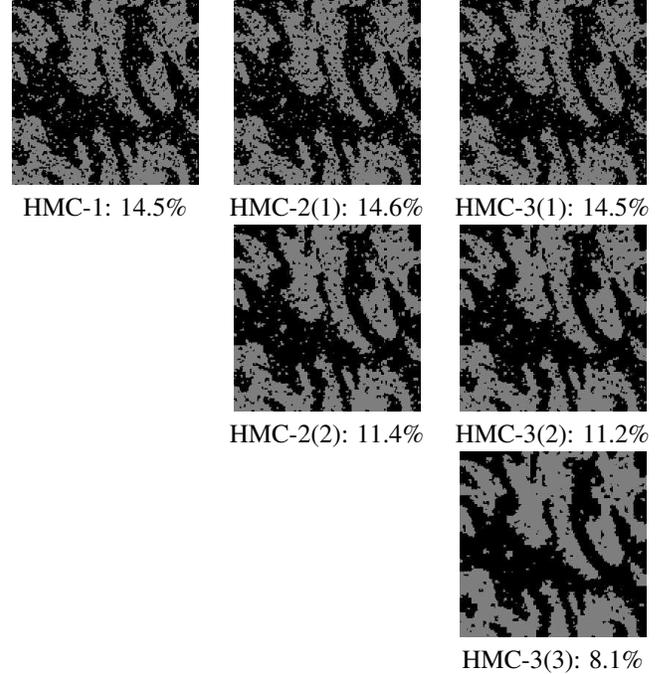


Fig. 3. Segmentation results obtained with ICE estimation and MPM classification for different memory lengths.

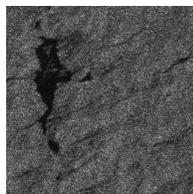
std, which allows to go through this question. However, this issue needs a deeper study.

### 4.1. Noisy Simulated image

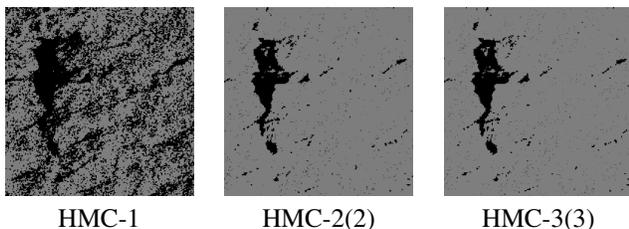
The first image is a simulated one ( $256 \times 256$ ), which represents a Gibbs field, in which the state densities are two Gaussians of near means ( $\mu_1 = 120, \mu_2 = 125$ ) and standard deviation ( $\sigma_1 = 60, \sigma_2 = 85$ ). Furthermore, the noises are correlated with the application of a smoothing filter. Original image of class and correlated noise image are presented in Fig. 2. Results of segmentation are presented in Fig. 3. The percentages give the error rates of misclassification according to the original image in Fig. 2.

The resulting class images confirm the interest of a HMC- $R$ , associated with high-order conditional observation probabilities. Indeed, we can notice that a HMC-2(1) or a HMC-3(1) segmentation, based on classical state-observation probabilities densities ( $H_2$ ), are equivalent with a HMC-1; whereas a HMC-2(2) and a HMC-3(3) segmentation, based on  $H_2^R$ , proved to be much more accurate in term of homogeneity.

These results confirm the well-known assumption that it is possible to transform any HMC- $R$ , based on  $H_2$ , to a mathematically equivalent first order version. Furthermore, it confirms the interest of HMC- $R$  in image segmentation, which seems to enable a more accurate segmentation for this kind of correlated noise.



**Fig. 4.** ERS SAR observation of an oil slick in the Mediterranean sea.



**Fig. 5.** Segmentation results obtained with HMC and HMC-*R* models.

#### 4.2. SAR image

Fig. 4 is an excerpt of an ERS-SAR image ( $512 \times 512$ ), acquired in October 3<sup>rd</sup> 1992, near the Egyptian coast, ©ESA. Fig. 5 shows the class images resulting from the segmentation with the classical HMC-1, and with the new HMC-2 and HMC-3 models. The difficulty of this image is due to the fact that oil on the water reduces air-sea interaction and the main observable phenomenon is the dampening of the capillary (surface) waves, which causes the major part of the noise it contains [8]. The segmentation was naturally performed with two classes: *oil slick* and *free sea*.

HMC-1 technique, which only takes into account the previous pixel to determine the pixel state, is unable to detect the noisy zone which constitute the damped waves. HMC-*R* take more in account, and reveals very performing in detecting noisy zone. In fact, the HMC-1 model, which captures only closed interactions, has a limited ability to describe noisy large scale behavior. Hence, the HMC-*R* model, which incorporate more neighboring pixels, allows one to take into account more complex noise structures.

### 5. CONCLUSION

In this work, we described a new technique based on HMC-*R* models for unsupervised image segmentation. The extension of the HMC model to HMC-*R* one is almost straightforward. However, we developed an extended version of the ICE procedure and also introduced a new high-order conditional observation probability, which allows one to take into account more complex and correlated noise. Due to the high complexity of the HMC-*R* model, implying greater num-

ber of parameters and computation time, it was important to verify the interest of the method. Experiments on simulated data and SAR images confirm this. HMC-*R* model, which is more general - and more complex - than the HMC-1 one, reveals very performing in image segmentation, especially in modelling more complex spatial relationship between pixels and so more complex noise structures.

We now plan to study the likeness between HMC-2 and the recent Pairwise Markov Chains (PMC) [9] model. A preliminary study shows that HMC-2 and PMC could produce, in particular situation, similar results. However, HMC-*R* seems to be globally more efficient in terms of quality.

### 6. REFERENCES

- [1] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor HMC and unsupervised image segmentation," *IEEE Trans. on PAMI*, vol. 19, no. 5, pp. 465–475, 1997.
- [2] W. Skarbek, "Generalized Hilbert scan in image printing," in *Theoretical Foundations of Computer Vision*. Akademik Verlag, Berlin, 1992.
- [3] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised segmentation of radar images using HMC and HMRF," *IEEE Trans. on GRS*, vol. 41, no. 3, pp. 675–686, 2003.
- [4] E. de Villiers and J. du Preez, "The advantage of using higher order HMM for segmenting acoustic files," in *12th Symp. PRASA*, South Africa, 2001.
- [5] J.F. Mari, J.P. 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.
- [6] R.J. Boys and D.A. Henderson, "A comparison of reversible jump MCMC algorithms for DNA sequence segmentation using HMMs," *Comp. Sci. and Stat.*, vol. 33, pp. 1–15, 2002.
- [7] O. Aycard, J.F. Mari, and F. Washington, "Learning to automatically detect features for mobile robots using second-order HMMs," in *IEEE IJCAI Workshop*, 2003.
- [8] F. Girard-Ardhuin, G. Mercier, and R. Garello, "Oil slick detection by SAR imagery: potential and limitation," in *Oceans 2003*, San Diego, USA, september 2003, pp. 22–26.
- [9] W. Pieczynski, "Pairwise Markov chains," *IEEE Trans. on PAMI*, vol. 25, no. 5, pp. 634–639, 2003.