

# Multiresolution Hidden Markov Chain Model and Unsupervised Image Segmentation

Laurent Fouque<sup>(2)</sup>, Alain Appriou<sup>(2)</sup> and Wojciech Pieczynski<sup>(1)</sup>

<sup>(1)</sup> INT, Departement SIM, 9 Rue Charles Fourier 91000 Evry-France

<sup>(2)</sup> ONERA, Departement DTIM, 29 avenue de la Division Leclerc 92322 Chatillon-France

laurent.fouque@sim.int-evry.fr

## Abstract

*Several approaches have been proposed in the last few years to handle the problem of multiresolution image segmentation. In a Bayesian framework, models using Markov fields have been highly effective. However the computational cost can be prohibitive. Markov tree models were therefore proposed. Although fast, these methods do not always give good results. In this article, we propose a new approach using a Markov chain built by transforming multiresolution images into one vectorial process via a Peano type scan, the Hilbert scan. We work in an unsupervised context in which parameters estimation is carried out by using a mixture distribution algorithm, the ICE algorithm. Experimental results, including classification of multiresolution synthetic images and SPOT images, are presented in this paper.*

## 1. Introduction

In recent years, multiresolution methods have been developed in different frameworks for multiresolution data fusion. Several models based on wavelet transforms have been proposed[8][3]. In a Markovian context, which is the approach we consider, multiresolution models have been used in image coding [2] and image restoration[7][6]. In this article, we deal with the problem of statistical segmentation of multiresolution images with a pyramidal structure. Intuitively, the best adapted model for the distribution of the class process is the three-dimensional Markov field model[15]. This model, which is a generalization of the classic Markov field model[11][5], requires complex and iterative computations. Other models have been suggested in order to speed up processing. Based on the Markov tree model[16][6], they have been applied successfully in various situations. They are rapidly computed but can however generate classification results in which “blocky” phenomena sometimes appear. This

is why we suggest a new approach to treat the multiresolution segmentation problem. We use the Hilbert curve to transform the two-dimensional multiresolution images into a one-dimensional vectorial signal and assign this signal a Markov chain structure. The multiresolution model we propose enables us to maintain the algorithmic simplicity of the Markov tree.

The remainder of the article is organised in six parts. The second part presents the multiresolution hidden Markov chain model. In the third part, we deal with the statistical classification problem adapted to our model. The parameters estimation will be studied in the fourth part. The fifth part presents experimental results of segmentation on synthetic and real images. In the final part, we make our conclusion.

## 2. Multiresolution Hidden Markov Chain model (MHMC)

Let  $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(d)}$  be respectively  $d$  observed images at resolution  $R^{(1)}, \dots, R^{(d)}$ ,  $R^{(1)}$  corresponding to the coarsest resolution,  $R^{(d)}$  to the finest one. These images have a pyramidal structure. We assume that  $\mathbf{Y}^{(1)}$  has  $4^N$  pixels and  $\mathbf{Y}^{(r)}$  has  $4^{N+r-1}$  pixels,  $1 \leq r \leq d$ . Classically,  $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(d)}$  are  $d$  observed processes representing a “noisy” version of  $d$  “ideal” hidden processes  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)}$ , the classes processes. Let  $\mathbf{X}^{(r)} = (X_0^{(r)}, \dots, X_{4^{N+r-1}-1}^{(r)})$ ,  $1 \leq r \leq d$ . We assume that each  $X_i^{(r)}$ ,  $0 \leq i \leq 4^{N+r-1} - 1$  takes its value in the set of classes  $\Omega = \{\omega_1, \dots, \omega_k\}$ . Using the geometrical features of the Hilbert scan, it is possible to match one “parent” pixel of the resolution  $r$  with four “children” pixels of the resolution  $r + 1$  in an optimal way, in the sense that we conserve the spatial information.

Let us start with the case of a two-resolution process ( $d = 2$ ) with two classes ( $k = 2$ ). In the case of figure (1), the father  $X_0^{(1)}$  corresponds to the four children

$(X_0^{(2)}, X_1^{(2)}, X_2^{(2)}, X_3^{(2)})$ , and so on. We use this fractal property of the Hilbert scan to define the multiresolution hidden Markov chain model. Therefore, the two processes  $\mathbf{X}^{(1)}$  and  $\mathbf{X}^{(2)}$  will be defined by a unique random process  $\mathbf{X}$  on  $\Omega^2$  (figure (2)):

$$\mathbf{X} = \left( \left( \begin{array}{c} X_0^{(1)} \\ X_0^{(2)} \end{array} \right), \left( \begin{array}{c} X_0^{(1)} \\ X_1^{(2)} \end{array} \right), \dots, \left( \begin{array}{c} X_{4^{N+1}-1}^{(1)} \\ X_{4^{N+1}-1}^{(2)} \end{array} \right) \right)$$

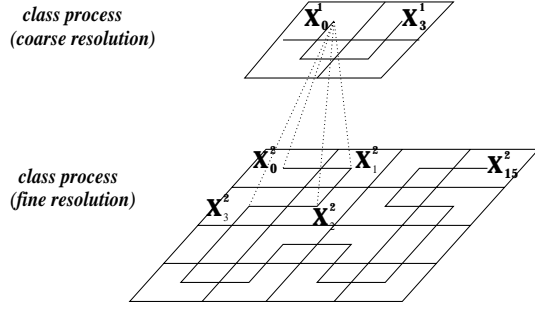


Figure 1: multiresolution model

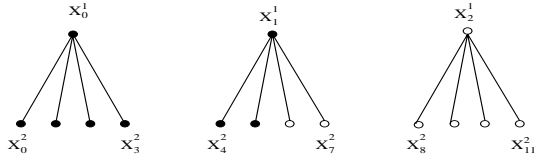


Figure 2: Realization of a multiresolution process with two resolutions and two classes (white and black)

Let  $\mathbf{x} = (x_0, \dots, x_{4^{N+1}-1})$  be a realization of  $\mathbf{X}$ . We consider the state space  $\Omega^2 = \{e_1, \dots, e_4\}$ , with  $e_1 = \begin{pmatrix} w_1 \\ w_1 \end{pmatrix}$ ,  $e_2 = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ ,  $e_3 = \begin{pmatrix} w_2 \\ w_1 \end{pmatrix}$ ,  $e_4 = \begin{pmatrix} w_2 \\ w_2 \end{pmatrix}$ . The MHMC model is defined by assigning a non-homogeneous Markov chain structure to the class process  $\mathbf{X}$ . The distribution of  $\mathbf{X}$  is then given by:

- the distribution of  $\begin{pmatrix} X_0^{(1)} \\ X_0^{(2)} \end{pmatrix}$ :

$$\pi \left( \begin{array}{c} x_0^{(1)} \\ x_0^{(2)} \end{array} \right) = P \left( \left( \begin{array}{c} X_0^{(1)} \\ X_0^{(2)} \end{array} \right) = \left( \begin{array}{c} x_0^{(1)} \\ x_0^{(2)} \end{array} \right) \right) \quad (1)$$

for  $(x_0^{(1)}, x_0^{(2)}) \in \Omega^2$ .

- The transition matrices which are given for  $1 \leq n \leq 4^{N+1} - 1$  by (we have to consider 2 cases):

- for  $n \bmod 4 = 0$

$$t_{x_{n-1}, x_n}^{(1)} = t^{(1)} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \quad (2)$$

with  $0 \leq a_{ij} \leq 1$  and  $\sum_{j=1}^4 a_{ij} = 1$ ,  $1 \leq i \leq 4$ . This matrix corresponds to the case in which “the father and its son are updated”; all transitions are possible, which is represented by an unspecified matrix  $t^{(1)}$ .

- for  $n \bmod 4 \neq 0$

$$t_{x_{n-1}, x_n}^{(2)} = t^{(2)} =$$

$$\begin{pmatrix} c & 1-c & 0 & 0 \\ 1-d & d & 0 & 0 \\ 0 & 0 & e & 1-e \\ 0 & 0 & 1-f & f \end{pmatrix} \quad (3)$$

$0 \leq c, d, e, f \leq 1$ .

This matrix correspond to the case in which “the father is constant, its son is updated”. Therefore, from the class  $e_1$  or  $e_2$ , the transition to the class  $e_3$  or  $e_4$  is not possible; in the same way, from  $e_3$  or  $e_4$ , the transition to  $e_1$  and  $e_2$  is not possible.

Let us consider now the general case of a  $d$  resolution process with  $k$  classes. We note  $\mathbf{X} = (X_0, \dots, X_{4^{N+d}-1})$  the process defined by :

$$X_s = \begin{pmatrix} X_{Q(s, 4^{d-1})}^{(1)} \\ \vdots \\ X_{Q(s, 4^{d-r})}^{(r)} \\ \vdots \\ X_{Q(s, 4^0)}^{(d)} \end{pmatrix} \quad (4)$$

where  $Q(s, 4^u) = \lfloor s/4^u \rfloor$  and  $0 \leq s \leq 4^{N+d}-1$ ,  $\|$  refers to the integer part operator.

This process is defined on the cartesian product  $\Omega^d = \{\omega_1, \dots, \omega_k\}^d$ . The  $d$  processes  $\mathbf{X}^{(r)} = (X_0^{(r)}, \dots, X_{4^{N+r}-1}^{(r)})$ ,  $1 \leq r \leq d$  will be therefore considered as a unique random process  $\mathbf{X}$  on  $\Omega^d$ . Let  $\mathbf{x} = (x_0, \dots, x_{4^{N+d}-1})$  be a realization of  $\mathbf{X}$ . We define the MHMC model by assigning a non-homogeneous Markov chain structure to the class process  $\mathbf{X}$ . Its distribution is given by:

- the distribution of  $X_0 = \begin{pmatrix} X_0^{(1)} \\ \vdots \\ X_0^{(d)} \end{pmatrix}$ :

$$\pi_{X_0} = P(X_0 = x_0), x_0 \in \Omega^d \quad (5)$$

- the transition matrices which are given for  $1 \leq n \leq 4^{N+d-1} - 1$  by (we have to consider  $d$  cases):

- if  $n \bmod 4 \neq 0$ :

$$t_{x_{n-1}, x_n}^{(d)} = t^{(d)} \quad (6)$$

This matrix corresponds to the case in which the  $d^{\text{th}}$  component of  $X_n$  changes and the others remain constant.

- if  $n \bmod 4^{d-r} = 0$  and  $1 \leq r \leq d-1$ :

$$t_{x_{n-1}, x_n}^{(r)} = t^{(r)} \quad (7)$$

This matrix corresponds to the case where the last  $r$  components of  $X_n$  change and the others remain constant.

In conclusion,  $t^{(r)}$ ,  $1 \leq r \leq d$ , are  $d$  given matrices to constrain each parent pixel to be constant four times with respect to its sons.

### 3. Segmentation algorithm

As we mentioned before,  $\mathbf{X} = (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(d)})$  is an unobservable process representing the image of classes. The problem consists of estimating this process only from the available observations  $\mathbf{Y} = (\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(d)})$ . We define the observations with the same structure as the class processes. Let  $\mathbf{Y} = (Y_0, \dots, Y_{4^{N+d-1}-1})$  the observed process defined by:

$$Y_s = \begin{pmatrix} Y_{Q(s, 4^{d-1})}^{(1)} \\ \vdots \\ Y_{Q(s, 4^{d-r})}^{(r)} \\ \vdots \\ Y_{Q(s, 4^0)}^{(d)} \end{pmatrix} \quad (8)$$

Let  $\mathbf{Y}^{(r)} = (Y_0^{(r)}, \dots, Y_{4^{N+r-1}-1}^{(r)})$ ,  $1 \leq r \leq d$  and  $\mathbf{y} = (y_0, \dots, y_{4^{N+d-1}-1})$  be a realization of  $\mathbf{Y}$ .

#### 3.1 Assumptions on the observation model

We use the following classic assumptions:

- conditional independance of the observations  $Y_n^{(r)}$ ,  $1 \leq r \leq d$  and  $0 \leq n \leq 4^{N+d-1} - 1$ , given the class process  $\mathbf{X}$ .

- the distribution of  $Y_n^{(r)}$  given  $\mathbf{X}$  is equal to its distribution given  $X_n^{(r)}$ ,  $1 \leq r \leq d$  and  $0 \leq n \leq 4^{N+d-1} - 1$ .

Such assumptions are strong but simplify the model, which thus have some good statistical properties as the one we give in the proposition 1.

#### 3.2 Posterior distribution of $\mathbf{X}$

The Posterior distribution of  $\mathbf{X}$  will be useful for estimating parameters. Contrary to the Markov fields model, the posterior distribution of  $\mathbf{X}$  is directly tractable, without using iterative algorithms such as Gibbs sampler. General results on Markov chains are valid in the context of our model. We recall one of these results in the following proposition.

##### Proposition 1:

*The a posteriori distribution of  $\mathbf{X}$  is the distribution of an non-homogeneous Markov chain.*

In accordance with this proposition, it is easy and rapid to simulate realizations of  $\mathbf{X}$  according to its posterior distribution without having recourse to iterative Monte-Carlo methods. This constitutes a great advantage compared to Markov random field models which imply the use of algorithms such as Gibbs sampler or Metropolis algorithms [11][13]. In the context of Markov random chain models, computation of “Backward” probabilities [4] is enough to obtain such realizations.

#### 3.3 The MPM Bayesian segmentation method

In our context, we use the MPM decision rule[17] which is analytically computable and quickly computed compared to MAP solution provided by simulated annealing algorithms [1]. The MPM decision rule is characterised by:

$$\widehat{x}_n = \arg \max_{u \in \Omega^d} P(X_n = u | \mathbf{Y} = \mathbf{y}) \quad (9)$$

This rule requires the computation of “Forward” probabilities and “Backward” probabilities.

### 4. Parameter estimation algorithm

From the beginning, we have assumed that the parameters of our multiresolution hidden Markov chain model (initial distribution and transition matrices) and of the conditionnal density were known (supervised context). Parameters are usually unknown, and we need to estimate them from the observations. Different estimation algorithms are available. The iterative EM algorithm (Expectation-Maximisation)[9] maximizes the likelihood of the observations. Modifications of this algorithm have been proposed to improve

convergence of the algorithm to a global maximum of the likelihood [18]. In this paper, we use a different algorithm for mixture distribution estimation, the ICE algorithm presented in a general context in [12]. We apply this algorithm to the case of our model [10].

## 5. Application

To evaluate the pertinence of the model, we present below experimental results of segmentations obtained with synthetic and real images. First, the segmentation is carried out by using the real parameters of the model. Then we apply the model to SPOT images from Istres, an area in the south of France.

Let us consider a realization of a Markov random field with three classes. For this simulation, we chose for the two resolutions, images corrupted with a similar strong noise to show the interest of using all the information. In figure (3), image (a) is the original image corresponding to the true classes. Image (b) and (c) correspond, for the resolution  $R^{(1)}$  et  $R^{(2)}$ , to the corrupted images (noisy image with  $\mathcal{N}(0,1)$  for the first class,  $\mathcal{N}(0.5,1)$  for the second class,  $\mathcal{N}(1,1)$  for the third class). Figure (4) provides results of a segmentation obtained by using real parameters: (a) shows the result of an MPM monoresolution segmentation obtained by using information provided by (c) of figure (3) only. (b) presents the multiresolution segmentation result obtained from both images (b) and (c) of figure (3) (17.3% of pixels are misclassified in the monoresolution case, 12.9% of pixels misclassified in the multiresolution case).

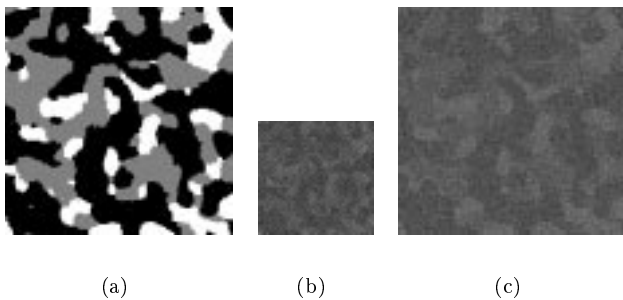


Figure 3: original scene and noisy images

We now apply our model in an unsupervised context to classify multiresolution SPOT images of an area of France, the ISTRÉ coast and its surrounding. The aim is not to provide a precise description of the scene but only show a kind of example for using such a model. Two SPOT images correspond to the same geographic zone. The scene is composed of four classes, water (pond of Berre), fields (corn and wheat), city (Istres

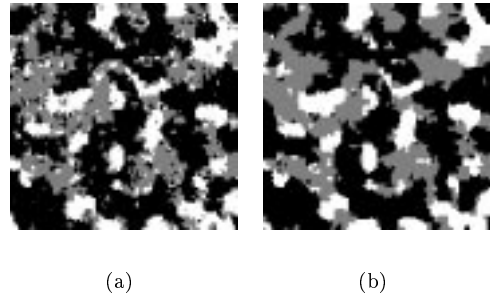


Figure 4: mono and multiresolution segmentation

on the left of pond of Berre), and vegetation (forest or other agricultural zones). The image at the resolution  $R^{(2)}$  show a kind of uncertainty in the north of the area corresponding to the water due to the depth of the water. The image at the resolution  $R^{(1)}$  provides better information which permit to detect completely the sea zone, the monoresolution segmentation obtained from the image at the resolution  $R^{(2)}$  only classifying the sea in three classes. Visually, results of the classification is improved when using information from both images.

## 6. Conclusion

In this article, we define a multiresolution Markov model which enables us to treat the problem of statistic segmentation of images defined on a pyramidal structure. We have presented the case of two images, each pixel of the coarse resolution image giving four pixels son of the fine resolution image. A solution would be to use a three-dimensional Markov model which sometimes gives very good results. However, these models require iterative algorithms which are very time-consuming in terms of computation. Much research based on the Markov trees has been conducted. These models, which assume independence of children given their parent ensure non-iterative algorithms (much rapid), but the segmentation results are not always satisfactory. In this article, we propose a new approach to define the distribution of the class process. We use Markov chain structure by transforming images via the Hilbert scan. Some first simulations indicate good performances of our model for multiresolution segmentation. However, additional research is currently being conducted to compare this model with other existing multiresolution Markov models.

## References

- [1] R. Azencott, "Simulated annealing: parallelization techniques". Wiley, 1992

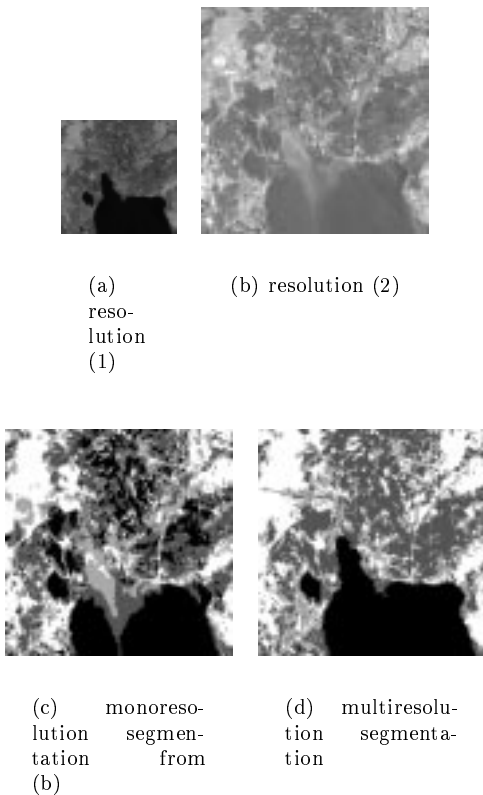


Figure 5: mono and multiresolution segmentation on SPOT images (ISTRE coast)

- [2] M. Barlaud, L. Blanc-Fraud and P. Charbonnier, "Image coding using multiresolution Markov random fields". Proc. SPIE, Image Processing Algorithms and Techniques, San Jose, CA, February 1992.
- [3] M. Basseville, A. Benveniste, K.C. Chou, S.A. Golden, R. Nikoukhah and A.S. Willsky, "Modeling and estimation of multiresolution stochastic processes". IEEE Trans. Information Theory, vol 38, n 2, pp 766-784, 1992.
- [4] B. Benmiloud and W. Pieczynski, "Estimation des paramètres dans les chaînes de Markovs cachées et segmentation d'images". Traitement du Signal, vol 12, n 5, pp 433-454, 1995.
- [5] J. Besag, "On the statistical analysis of dirty pictures". Journal of the Royal Statistical Society, B, vol 48, pp 259-302, 1986.
- [6] C. Bouman and M. Shapiro "A multiscale random field model for Bayesian image segmentation". IEEE Trans. on IP, vol 3, n 2, pp 162-177, 1994.
- [7] M. Charbonnier, L. Blanc-Fraud and P. Barlaud, "Noisy image restoration using multiresolution Markov random fields". J. of Visual Com. and Image Representation, vol 3, n 4, pp 338-346, 1992.
- [8] K.C. Chou, S.A. Golden and A. S. Willsky, "Multiresolution stochastic models, data fusion and wavelet transforms". Signal Processing, vol 34, n 3, pp 257-282, 1993.
- [9] A.P. Dempster, N. M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm". Journal of the Royal Statistical Society, B, vol 39, pp 1-38, 1977.
- [10] L. Fouque, A. Appriou and W. Pieczynski, "Modèle de chaîne de Markov Cachée multiresolution et segmentation non supervisée d'images". 12ème Congrès Francophone AFRIF-AFIA Paris 1-3 fvrier 2000.
- [11] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the bayesian restoration of Images". IEEE Trans. on PAMI, vol 6, n 6, pp 721-741, 1984.
- [12] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden markov chains and unsupervised imge segmentation". IEEE Trans. on PAMI, vol 19, n 5, pp 465-475, 1997.
- [13] X. Guyon, "Champs aléatoires sur un réseau". Masson, 1993.
- [14] S. Kamata, "Data compression for gray images using hilbert scan", System and Computer in Japan, vol 28, n 9, pp 26-34, 1997.
- [15] Z. Kato, J. Zeroubia and M. Berthod, "Unsupervised parallel image classification using Markovian models", Pattern Recognition Letters, vol 32, pp 591-604, 1999.
- [16] J.M. Laferté, F. Heitz and P. Pérez, "Modèles markoviens d'images et algorithmes d'estimations non linéaire sur le quadarbre", Traitement du Signal, Vol 15, n 3, pp 213-230, 1998.
- [17] J.L. Marroquin, S. Mittle and T. Poggio, "Probabiliste solution of ill-posed problems in computational vision". Journal of the American Statistical Association, vol 82, pp 76-89, 1987.
- [18] G.J. McLachlan and T. Krishnan, "The EM algorithm and Extensions" Wiley Series in Probability and Statistics, 1997.