UNSUPERVISED NON STATIONARY IMAGE SEGMENTATION USING TRIPLET MARKOV CHAINS

Pierre Lanchantin and Wojciech Pieczynski

Pierre.Lanchantin@int-evry.fr
GET/INT, Département CITI, CNRS UMR 5157, 9, rue Charles Fourier, 91000 Evry, France

ABSTRACT

This work deals with the unsupervised Bayesian hidden Markov chain restoration extended to the non stationary case. Unsupervised restoration based on "Expectation-Maximization" (EM) or "Stochastic EM" (SEM) estimates considering the "Hidden Markov Chain" (HMC) model is quite efficient when the hidden chain is stationary. However, when the latter is not stationary, the unsupervised restoration results can be poor, due to a bad match between the real and estimated models. In this paper we present a more appropriate model for non stationary HMC, via recent Triplet Markov Chains (TMC) model. Using TMC, we show that the classical restoration results can be significantly improved in the case of non stationary data. The latter improvement is performed in an unsupervised way using a SEM parameter estimation method. Some application examples to unsupervised image segmentation are also provided.

1. INTRODUCTION

The hidden Markov chain (HMC) model is widely used for various problems, including signal and image processing, economical prediction, and health sciences. In such model, $X = (X_1, ..., X_N)$ is a stochastic process modeling an unobservable – or hidden – discrete signal (each X_n takes its values in a finite set $\Omega = \{\omega_1, ..., \omega_K\}$) and $Y = (Y_1, ..., Y_N)$ is a stochastic process modeling the observations (each Y_n takes its values in the set of real numbers R). The problem is then to estimate the hidden discrete signal $X = x$ from the observed signal $Y = y$. The links between X and Y are then modeled by the following joint distribution :

$$p(x, y) = p(x_1)\, p(x_2|x_1) ...\, p(x_N|x_{N-1})\, p(y_1|x_1) ...\, p(y_N|x_N) \quad (1)$$

Such a model is called "hidden" Markov chain with "independent noise" (because the hidden process X is a Markov chain and the random variables $(Y_1, ..., Y_N)$ are independent conditionally on X). Hereafter, it will be denoted by HMC-IN. It allows one to recover the hidden data $X = x$ from the observed data $Y = y$ using different Bayesian classification techniques like Maximum A Posteriori (MAP), or Maximal Posterior Mode (MPM) [1-2]. These restoration methods use the distribution of the hidden process conditional to the observations, which is called its "posterior" distribution.

When $p(x, y)$ is based on unknown parameters $\theta \in \Theta$, the latter can be estimated from $Y = y$, considering the stationarity of the hidden process by methods like "Expectation-Maximization" (EM [3]) or "Stochastic Expectation-Maximization" (SEM [4]). Thus, when the hidden process is a stationary Markov chain, parameters are well estimated, and the restoration based on the estimated parameters gives good results. However, when this is not the case, the estimation necessarily gives wrong results, which can imply poor restoration of X. In this work, we propose to model the non stationarity by introducing an auxiliary process governing the regime switching of the hidden process. The introduction of an auxiliary process to model the non stationarity has already been treated before. In [5], the authors introduce auxiliary data to modify the transition probabilities of the Markov hidden process. In [6], the author present a model based on the superposition of two Markov chains. A non-homogeneous discrete process is described by a set of transition matrices. At each n, the choice of the active matrix is then governed by a hidden homogeneous Markov chain.
Nevertheless, our approach is quite different : we propose a model of non stationary hidden Markov chain by using the Triplet Markov Chain (TMC) model [7,8].

It is based on the three following points :
(i) When X is non stationary, we can introduce an auxiliary stochastic process U whose aim is to model the regime switching of X ;
(ii) We directly assume the markovianity and the stationarity of the pairwise process $V = (X, U)$. Then, the distributions P_X and P_U are not necessarily Markovian which allows one to deal with different kinds of stochastic processes;
(iii) The distribution of Y conditional on $V = (X, U)$ is such that the triplet process $T = (X, U, Y)$ is a Markov chain. Although the marginal distribution $P_{(X,Y)}$ of T is

not necessarily a Markov distribution, it can be used to perform Bayesian restorations, T being a particular case of Triplet Markov chain [7].
We provide different simulation studies, showing that such an introduction of an auxiliary process can improve the results obtained with the classical unsupervised Bayesian restoration. We provide a complete derivation of the Forward-Backward, MPM and SEM algorithms for the new model and give computable versions of these ones.
The paper is organized in the following way : The next section is devoted to a brief description of the Triplet Markov Chain model. The particular model used and the related processing method is described in section three and different simulation results described in section four.

2. TRIPLET MARKOV CHAINS

Let $X=(X_n)_{1\le n\le N}$ and $Y=(Y_n)_{\le n\le N}$ be two stochastic processes. X is hidden (each X_n takes its values in a finite set $\Omega=\{\omega_1,...,\omega_K\}$), and Y is observed (each Y_n takes its values in the set of real numbers R). The problem is then to estimate $X=x$ from $Y=y$.

Definition 2.1
The model considered is called a Triplet Markov Chain if there exists a stochastic process $U=(U_n)_{n\in N}$, with each U_n taking its values in a set $\Lambda=\{\lambda_1,...,\lambda_M\}$, such that $T=(X,U,Y)=((X_n,U_n,Y_n))_{\le n\le N}$ is a Markov chain.

The idea of TMC is to consider the distribution of $Z=(X,Y)$, which models the interactions between the observed and the searched process, as a marginal distribution of a Markov chain $T=(X,U,Y)$. The distribution s $p(x_n,y)$, which are used in MPM restoration, are then computable. In fact, let $V=(X,U)$. As T is Markovian, the process (V,Y) is a Pairwise Markov Chain (PMC) and we can formulate and calculate the distribution of (V_n,Y) using Forward-Backward recursions [9]. This means that the distributions $p(x_n,y)=\sum_{u_n\in\Lambda}p(x_n,u_n,y)=\sum_{u_n\in\Lambda}p(v_n,y)$ also are computable. Finally, although the distribution of (X,Y) is not necessarily a Markov one, the marginal distributions $p(x_n|y)$ are computable. which in particular enables the use of the Bayesian MPM restoration method.
We said above that $Z=(X,Y)$ in not necessarily a Markov chain. More precisely, the following result specifies "locally" on what the greater generality of TMC with respect to PMC consists [8] :

Proposition 2.1

Let $T=(X,U,Y)$ be a TMC verifying :
(a) $p(t_n,t_{n+1})$ does not depend on $1\le n\le N-1$; and
(b) $p(t_n,t_{n+1})=p(t_{n+1},t_n)$.

Then the three following conditions:
(i) $Z=(X,Y)$ is a Markov chain ;
(ii)for each $2\le n\le N$, $p(u_n|z_n,z_{n-1})=p(u_n|z_n)$;
(iii)for each $1\le n\le N$, $p(u_n|z)=p(u_n|z_n)$
are equivalent.

The Pairwise Markov Chain (PMC [9,10,11]) model, in which Z is assumed to be a Markov chain, is then a particular case of the TMC obtained by taking $\Lambda=\Omega$ and $U=X$. Furthermore, according to the Proposition 2.1 above TMC is strictly more general than PMC, the latter being strictly more general than classical HMC, in that there exist Markov chains T such that Z are not Markov chains.

3. THE PARTICULAR TMC USED AND RELATED PROCESSING METHOD

3.1. TMC used model

Let X be a non stationary hidden process X. Let us introduce $U=(U_1,...,U_N)$ (each U_n takes its values in a finite set $\Lambda=\{\lambda_1,...,\lambda_M\}$) governing the regime switching of X. The pairwise process $V=(X,U)$ is assumed to be a Markov chain. Further, we assume that $Y_1,...,Y_N$ are independent conditionally on X, and $p(y_n|x_n,u_n)=p(y_n|x_n)$. Then, $T=(V,Y)$ is a particular TMC, in which V is a Markov chain, and which will be called "TMC" in the following.

3.2. MPM restoration in TMC

According to the general Bayesian theory, the MPM restoration is given by
$\hat{s}_{MPM}(y_1,...,y_N)=(\hat{x}_1,...,\hat{x}_N)$ with

$$\hat{x}_n=\arg\max_{x_n\in\Omega}p(x_n|y) \tag{2}$$

Thus, we need to calculate $p(x_n|y)$. As T is a TMC, the process (V,Y) is a PMC and we can write the distribution of (V_n,Y) as $p(v_n,y)=\alpha(v_n)\beta(v_n)$, with. $\alpha(v_n)=p(v_n,y_1,...,y_n)$ ("Forward" probabilities) and $\beta(v_n)=p(y_{n+1},...,y_N|v_n)$ ("Backward" probabilities).

These probabilities can then be recursively calculated for $1 \le n \le N-1$ by

$$\alpha(v_1) = p(v_1)p(y_1|x_1);$$
$$\alpha(v_{n+1}) = \sum_{v_n \in \Omega\times\Lambda} \alpha(v_n)p(v_{n+1}|v_n)p(y_{n+1}|x_{n+1}) \quad (3)$$
$$\beta(v_N) = 1;$$
$$\beta(v_n) = \sum_{v_{n+1} \in \Omega\times\Lambda} \beta(v_{n+1})p(v_{n+1}|v_n)p(y_{n+1}|x_{n+1}) \quad (4)$$

The marginal posterior distributions of the hidden state can be calculated by

$$p(v_n|y) \propto \alpha(v_n)\beta(v_n) \quad (5)$$

and the transition of the posterior Markov distribution $p(v|y)$ are calculated by

$$p(v_{n+1}|v_n, y_1,...,y_N) \propto p(v_{n+1}|v_n)p(y_{n+1}|x_{n+1})\beta(v_{n+1}) \quad (6)$$

Having calculated $p(v_n|y)$, we can then calculate $p(x_n|y) = \sum_{u_n \in \Lambda} p(v_n|y)$ and (2) to perform MPM.

3.3. Learning TMC

We assume that $p(y_n|x_n)$ are Gaussian. For K classes $\Omega = \{\omega_1,...,\omega_K\}$, we have to estimate K means μ_1, …, μ_K, and K variances σ_1^2, …, σ_K^2 of the K Gaussian densities $p(y_n|x_n = 1)$, …, $p(y_n|x_n = K)$. Further, the distribution of the stationary Markov chain $V = (U, X)$ is given by $2^{K\times M} \times (2^{K\times M} - 1)$ parameters $p_{ij} = p(v_1 = i, v_2 = j)$, which is a probability on $[\Omega\times\Lambda]^2$. The SEM method we use runs as follows :

(i) consider an initial values $\theta^0 = \left(p_{ij}^0, \mu_k^0, (\sigma_k^0)^2\right)$, for $0 \le i, j \le 2^{K\times\Lambda} - 1$, and $1 \le k \le K$;

(ii) for each $q \in N^*$:

-simulate $V = v^q$ according to $p(v|y)$ based on θ^q (which is possible due to (6));

-calculate $\theta^{q+1} = \left(p_{ij}^{q+1}, \mu_k^{q+1}, (\sigma_k^{q+1})^2\right)$ with

$$p_{ij}^{q+1} = \frac{1}{N-1}\sum_{n=1}^{N-1} 1_{[v_n = i, v_{n+1} = j]}, \ \mu_k^{q+1} = \frac{\sum_{n=1}^{N} y_n 1_{[x_n = k]}}{\sum_{n=1}^{N} 1_{[x_n = k]}},$$
$$(\sigma_k^{q+1})^2 = \frac{\sum_{n=1}^{N} (y_n - \mu_k^{q+1})^2 1_{[x_n = k]}}{\sum_{n=1}^{N} 1_{[x_n = k]}} \quad (7)$$

4. UNSUPERVISED IMAGE SEGMENTATION USING TMC

In this section, we compare the HMC-IN and the proposed TMC models in the field of image segmentation. The HMC-IN model has already been successfully applied to this problem : to do so, the bi-dimensional set of pixels is transformed into a mono-dimensional sequence through the Hilbert-Peano scan [10]. Parameters are then estimated with SEM algorithm and the MPM is applied to perform the segmentation. We present two series of experiments. In the first one, two synthetic hidden non stationary images are considered. We give different unsupervised restoration results considering the HMC and the TMC models and show that the latter is more appropriate when the hidden process presents different regimes. In the second one, we compare the two models considering a real image unsupervised segmentation.

4.1. Synthetic images

We consider two HMC (X, Y) verifying (1), with $\Omega = \{\omega_1, \omega_2\}$ and $N = 128\times128$. In the two HMC the hidden processes are non-stationary in different way. We define three transition matrices : $M_1 = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}$, $M_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$ and $M_3 = \begin{bmatrix} 0.02 & 0.98 \\ 0.98 & 0.02 \end{bmatrix}$ and two auxiliary processes $U^{(1)} = (U_1^{(1)},...,U_N^{(1)})$ and $U^{(2)} = (U_1^{(2)},...,U_N^{(2)})$, each $U_n^{(1)}$ and $U_n^{(2)}$ takes its values in the finite set $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$.

$*U^{(1)}$ is define by : $(U_1^{(1)},...,U_{N/4}^{(1)}) = \lambda_1$, $(U_{N/4+1}^{(1)},...,U_{N/2}^{(1)}) = \lambda_2$, and $(U_{N/2+1}^{(1)},...,U_N^{(1)}) = \lambda_3$.

$*U^{(2)}$ is a Markov chain define as follow :

(i) the distribution of $U_1^{(2)}$ is (1/3,1/3,1/3)

(ii) the transition matrix of $U^{(2)}$ is given by

$$M_{U^{(2)}} = \begin{bmatrix} 0.99 & 0.01 & 0 \\ 0 & 0.97 & 0.3 \\ 0.3 & 0 & 0.97 \end{bmatrix}$$

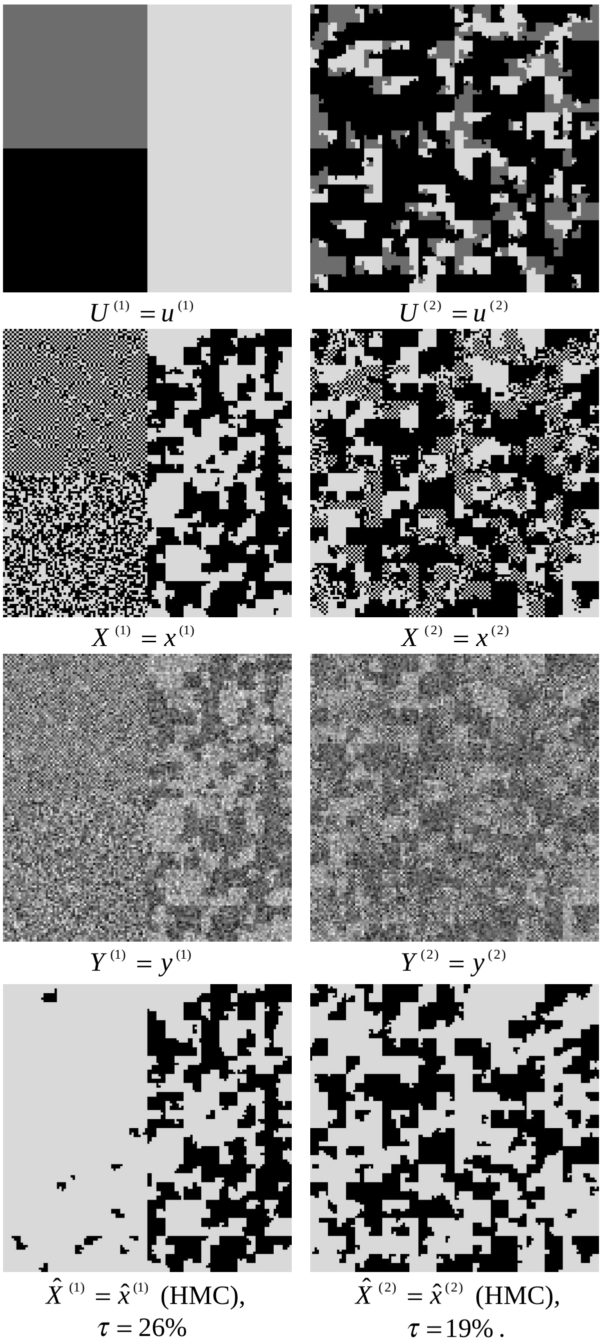
In the first experiment, the Markov chain $X^{(1)} = (X_1^{(1)},..., X_N^{(1)})$ is non stationary in the following way.

(i) the distribution of $X_1^{(1)}$ is $(0.5, 0.5)$,

(ii) M_p is the transition matrix in $X_n^{(1)}$ when $U_n^{(1)} = \lambda_p$ with $1 \le p \le 3$

In the second experiment, the Markov chain $X^{(2)}$ is obtain in the same way by replacing $U^{(1)}$ by $U^{(2)}$. We will denote $v_1 = (\omega_1, \lambda_1)$, $v_2 = (\omega_2, \lambda_1)$, $v_3 = (\omega_1, \lambda_2)$, $v_4 = (\omega_2, \lambda_2)$, $v_5 = (\omega_1, \lambda_3)$, $v_6 = (\omega_2, \lambda_3)$.

The processes $U^{(1)}, U^{(2)}, X^{(1)}$, $X^{(2)}$ are then converted, via a Hilbert-Peano scan, into a bi-dimensional set as described in [12] and their realizations are presented in Figure 1.

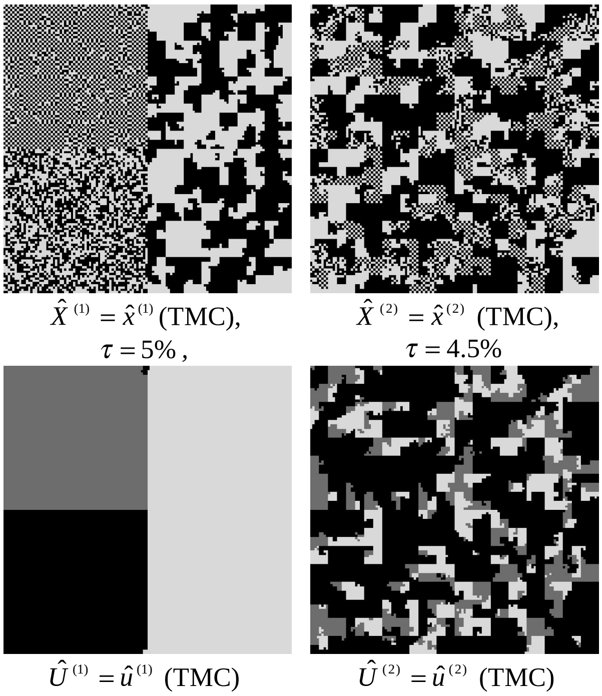


$U^{(1)} = u^{(1)}$ $U^{(2)} = u^{(2)}$

$X^{(1)} = x^{(1)}$ $X^{(2)} = x^{(2)}$

$Y^{(1)} = y^{(1)}$ $Y^{(2)} = y^{(2)}$

$\hat{X}^{(1)} = \hat{x}^{(1)}$ (HMC), $\tau = 26\%$ $\hat{X}^{(2)} = \hat{x}^{(2)}$ (HMC), $\tau = 19\%$.



$\hat{X}^{(1)} = \hat{x}^{(1)}$ (TMC), $\tau = 5\%$, $\hat{X}^{(2)} = \hat{x}^{(2)}$ (TMC), $\tau = 4.5\%$

$\hat{U}^{(1)} = \hat{u}^{(1)}$ (TMC) $\hat{U}^{(2)} = \hat{u}^{(2)}$ (TMC)

Figure 1 : *Different images and their unsupervised HMC and TMC based restorations.*

In the two experiments, a realization $X = x$ is simulated and $Y = y$ is sampled according to $p(y_1 | x_1), \ldots, p(y_N | x_N)$, where $p(y_n | x_n = 1)$ is Gaussian with mean 1 and variance 1, and $p(y_n | x_n = 2)$ is Gaussian with mean 3 and variance 1, which gives $Y^{(1)} = y^{(1)}$ and $Y^{(2)} = y^{(2)}$ in Figure 1. The realization $X = x$ is then estimated by the Bayesian MPM method from $Y = y$ in two different ways. The first restoration is obtained using the parameters estimated with the SEM algorithm and considering that (X, Y) is an HMC-IN (the hidden chain is assumed stationary), which gives $\hat{X}^{(1)} = \hat{x}^{(1)}$ (HMC), $\tau = 26\%$ and $\hat{X}^{(2)} = \hat{x}^{(2)}$ (HMC), $\tau = 19\%$ in Figure 1 (τ is the error ratio). The second restoration is obtained using the parameters estimated with the SEM algorithm assuming that $T = (X, U, Y)$ is the TMC considered above, which gives $\hat{X}^{(1)} = \hat{x}^{(1)}$ (TMC), $\tau = 5\%$, and $\hat{X}^{(2)} = \hat{x}^{(2)}$ (TMC), $\tau = 4.5\%$ in Figure 1. We also show in Figure 1 the estimates $\hat{U}^{(1)} = \hat{u}^{(1)}$ *and* $\hat{U}^{(2)} = \hat{u}^{(2)}$ of the of the auxiliary processes $U^{(1)} = u^{(1)}$ and $U^{(2)} = u^{(2)}$. We clearly see how the use of TMC can improve the results obtained with the classical HMC-IN model.

Furthermore, different parameters are well estimated (Table 1 and 2), which show a good behavior of the SEM considered.

Model	μ_1	σ_1^2	μ_2	σ_2^2
$p(y^{(1)} \mid x^{(1)})$				
HMC	0.98	1.03	2.33	1.91
TMC	1.02	1.01	2.98	0.94
$p(y^{(2)} \mid x^{(2)})$				
HMC	2.4	1.82	1.03	1.04
TMC	0.99	1.02	3.01	0.98

Table 1 : *estimated densities parameters.*

(a)	v_1	v_2	v_3	v_4	v_5	v_6
v_1	**0.98**	**0.02**	0	2e-4.	0	0
v_2	**0.02**	**0.98**	0	0	0	0
v_3	0	5e-4	**0.50**	**0.49**	0	0
v_4	0	0	**0.52**	**0.47**	0	0
v_5	0	0	0	0	**0**	**1**
v_6	0	0	5e-4	0	**0.95**	**0.04**
(b)	0.26	0.24	0.13	0.12	0.12	0.13

Table 2 : *estimated (a)* $p(v_{n+1}^{(1)} \mid v_n^{(1)})$ *and (b)* $p(v_0^{(1)})$ *considering the TMC model.*

(a)	v_1	v_2	v_3	v_4	v_5	v_6
v_1	**0.97**	**0.02**	0	0	6e-3	0
v_2	**9e-3**	**0.98**	0	0	0.01	0
v_3	0	0.03	**0**	**0.94**	0.02	0
v_4	0.03	0	**0.97**	**0**	0	0
v_5	0	0	0	0	**0.46**	**0.54**
v_6	0	0	0.8	0	**0.45**	**0.46**
(b)	0.23	0.40	0.09	0.09	0.09	0.09

Table 3 : *estimated (a)* $p(v_{n+1}^{(2)} \mid v_n^{(2)})$ *and (b)* $p(v_0^{(2)})$ *considering the TMC model.*

4.2. Real Image

Our example here is a 256x256 size satellite image of Tokyo, which is presented in Figure 2. We restore the image considering different number of real and auxiliary states. Comparisons between models are carried out using the Bayes Information Criterion (BIC) [13]. This Criterion is defined as $BIC = -2LL + p\log(N)$ where LL is the log-likelihood of the model, p its number of independent parameters and N the number of data. We do not take into account parameters estimated to be zeros according to the convention established in [14].

The best model is the one with the lowest BIC. Results are presented in Table 4.

Figure 2 : *Image "Tokyo"*

On Figure 3, we present the segmentation result by MPM and SEM parameters estimation based on the stationary HMC model with 4 classes.

On Figure 4, we present the segmentation result by MPM and SEM parameters estimation based on the TMC model with 4 classes and 2 states auxiliary process.

Number of real states	Number of auxiliary states	LL	Number of parameters	BIC
2	1	-338750	7	677580
2	2	-337460	13	675060
2	3	-337360	23	**674980**
3	1	-334310	14	668780
3	2	-332070	27	664440
3	3	-331760	55	**664130**
4	1	-333020	21	666280
4	2	-331210	43	**662610**

Table 4 : *Comparisons between models.*

According to the BIC values, we can say that for each number of states studied the TMC model with the largest number of auxiliary states is more appropriate. Furthermore, we can see by comparing segmentation results on Figure 3 and Figure 4 that the segmentation result considering the TMC model present finer features.

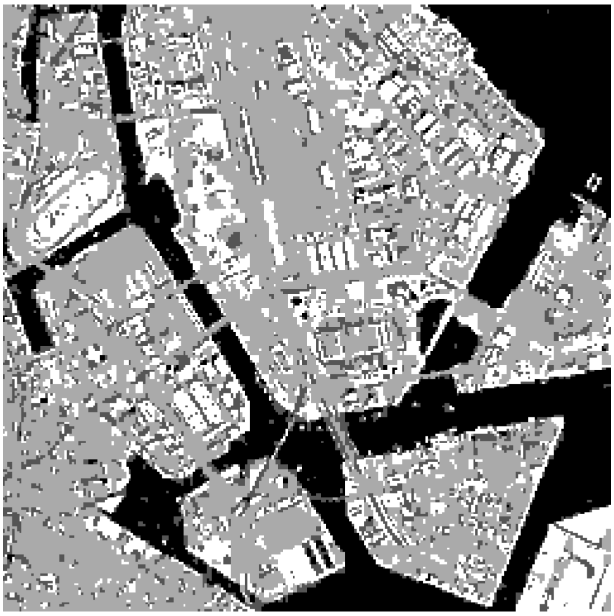


Figure 3 : *MPM segmentation after SEM parameters estimation based on the stationary HMC mode with 4 real states. BIC=666280.*

Figure 4 : *MPM segmentation after SEM parameters estimation based on the* TMC *model, with 4 real states and 2 auxiliary states. BIC=662610.*

5. CONCLUSIONS AND PERSPECTIVES

In this paper, we have provided an unsupervised method for restoring non stationary hidden Markov chains, with potential applications to unsupervised image segmentation. The main contribution was to tackle the lack of stationarity using the Triplet Markov Chain model. More precisely, the non stationary prior distribution of the hidden Markov chain was modeled by an auxiliary process governing the switching of the transition matrix. Bayesian segmentation techniques are then rendered applicable and different experiments show its interest in simulated and real images.

6. REFERENCES

1. L. E. Baum, T. Petrie, G. Soules, N. Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Statistic.*, 41, pp. 164-171, 1970.
2. G. D. Fornay, The Viterbi algorithm, *Proceedings of the IEEE*, Vol. 61, No. 3, pp. 268-277, 1973.
3. G. J. McLachlan and T. Krishnan, *EM Algorithm and Extensions*, Wiley, 1997.
4. G. Celeux and J. Diebolt, The SEM algorithm: a probabilistic teacher derived from the EM algorithm for the mixture problem, *Comput. Statist. Quarter*. 2, pp. 73-82, 1985.
5. J. P. Hugues, P. Guttorp and S. P. Charles, A non-homogeneous hidden Markov model for precipitation occurrence, *Appl.stat*, vol. 48, pp.15-30, 1999.
6. A. Berchtold, The double chain Markov model, *Commun. Statist. Theory Methods*, 28, 2569-2589, 1999.
7. W. Pieczynski, Chaînes de Markov Triplet – Triplet Markov Chains, *Comptes Rendus de l'Académie des Sciences – Mathématiques*, Paris, Ser. I 335, pp. 275-278, 2002.
8. W. Pieczynski, Triplet Markov Chains and Theory of Evidence, Submitted to *International Journal of Approximate Reasoning*, September 2003.
9. W. Pieczynski, Pairwise Markov chains, *IEEE Trans. on PAMI*, Vol. 25, No 5, pp.634-639,2003.
10. S. Derrode and W. Pieczynski, SAR image segmentation using generalized Pairwise Markov Chains, *SPIE's International Symposium on Remote Sensing*, September 22-27, Crete, Greece, 2002.
11. S. Derrode and W. Pieczynski, Signal and image restoration using Pairwise Markov Chains, *IEEE Trans. on Signal Processing*, scheduled for September 2004.
12. N. Giordana and W. Pieczynski, Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation, *IEEE Trans. on PAMI*, Vol. 19, No. 5, pp. 465-475, 1997.
13. R. W. Katz, On some criteria for estimating the order of a Markov chain, *Technometrics*, 23, 243-249, 1981.
14. Y. M. Bishop, S.E. Fienberg, P. W Holland, *Discrete Multivariate Analysis*, MIT Press, Cambridge, 1975.