

UNSUPERVISED SEGMENTATION OF NON STATIONARY DATA HIDDEN WITH NON STATIONARY NOISE

Mohamed El Yazid Boudaren¹, Wojciech Pieczynski² and Emmanuel Monfrini²

¹ Ecole Militaire Polytechnique, Lab. Math. Appl., PO Box 17, Algiers 16111, Algeria

² Institut Telecom, Telecom SudParis, CITI Dpt., CNRS UMR 5157, 9 rue Charles Fourier,
91000 Evry, France
boudaren@gmail.com

ABSTRACT

Classical hidden Markov chains (HMC) can be inefficient in the unsupervised segmentation of non stationary data. To overcome such involvedness, the more elaborated triplet Markov chains (TMC) resort to using an auxiliary underlying process to model the behavior switches within the hidden states process. However, so far, only this latter was considered non stationary. The aim of this paper is to extend the results of a recently proposed TMC by considering both hidden states and noise non stationary. To show the efficiency of the proposed model, we provide results of non stationary synthetic and real images restoration.

1. INTRODUCTION

The *HMC* is a pairwise stochastic process $Z = (X, Y) = (X_n, Y_n)_{n=1}^N$ in which, X is an unobservable Markov chain that is to be estimated from its observable noisy version Y . In such a model, the distribution of Z is given by $p(z) = p(z_1) \prod_{n=2}^N P(z_n | z_{n-1})$ and when the classical noise assumptions hold, we have

$$p(z_n | z_{n-1}) = p(x_n | x_{n-1}) p(y_n | x_n) \quad (1)$$

Throughout this paper, each X_n will take its values from a finite set of classes $\Omega = \{\omega_1, \dots, \omega_K\}$ and each Y_n will take its values from R .

Although this model seems to be quite simple, it permits a good restoration of X in a very wide range of situations. Hence, it has been extensively used in different fields that include signal and image processing [2, 3, 6, 9]. Its efficiency is majorly due to the different Bayesian techniques, like *MPM* and *MAP*, which allow the estimation of X from the whole Y within an acceptable computational complexity. However, in the unsupervised context and when the data are non stationary, the estimation becomes poor due to the mismatch between

the estimated stationary model and the data. To overcome this drawback, authors in [7] use an underlying discrete process $U = (U_n)_{n=1}^N$ to take into account the regime switches within X . The resulting model is a particular case of the so called triplet Markov chain (*TMC*) [8] which extends, itself, the pairwise Markov chain (*PMC*) [4].

In this paper, we intend to extend the previous model so that it takes into account the heterogeneous behavior of both $p(x)$ and $p(y|x)$. For this purpose, we propose a new *TMC* whose underlying process U models regime switches of both $p(x)$ and $p(y|x)$. Such a modeling can be of a great importance since it allows one to simultaneously model visual aspect variability within the same class, due to the lighting conditions for instance, and the heterogeneous interactions between image classes.

The remainder of this paper is organized as follows: section 2 briefly reminds the *TMC* formalism. Section 3 describes our new *TMC* and its computational developments. Section 4 is devoted to experiments conducted on synthetic and real images. Future improvements and conclusion are given in section 5.

2. TRIPLET MARKOV CHAINS

Let us consider two sequences of random variables $X = (X_1, \dots, X_N)$ and $Y = (Y_1, \dots, Y_N)$, where each X_n takes its values from a finite set of classes $\Omega = \{\omega_1, \dots, \omega_K\}$ and each Y_n takes its values from R . The problem is to recover X from Y . The model $Z = (X, Y)$ is said to be a triplet Markov chain (*TMC*) if there exists an underlying process $U = (U_1, \dots, U_N)$ with each U_n taking its values from a finite set $\Lambda = \{\lambda_1, \dots, \lambda_M\}$ such that $T = (U, X, Y)$ is a Markov chain. Let $V = (U, X)$. $T = (V, Y)$ is then a pairwise Markov chain (*PMC*). This makes the computation of the distributions $p(x_n | y)$, required to perform *MPM* restoration, affordable even when Z is not Markovian. This shows the greater generality of *TMC* over *PMC*, which is more general than *HMC*.

3. NON STATIONARY NOISE- TMC

In this section, we describe our new *TMC* designed for unsupervised segmentation of non stationary data hidden with non stationary noise and its corresponding restoration developments. All along this paper, this model will be called non stationary noise- *TMC* (*NSN- TMC*) in contrast to the *TMC* proposed in [7] that will be called stationary noise- *TMC* (*SN- TMC*).

3.1. Model presentation

Let $X = (X_n)_{n=1}^N$ be a non stationary unobservable process that is to be estimated. According to *TMC* formalism, one can introduce an underlying process to model the regime switches of X . Let $U^1 = (U_n^1)_{n=1}^N$ be such a process, $p(x|u^1)$ is then Markovian. Let now $U^2 = (U_n^2)_{n=1}^N$ be another hidden process and $Y = (Y_n)_{n=1}^N$ be a noisy version of X . Throughout this paper, $U = (U^1, U^2)$ will be assumed to be Markovian. More explicitly, we propose a *TMC* with transition probabilities given by $p(t_n|t_{n-1}) = p(u_n|u_{n-1})p(x_n|x_{n-1}, u_n^1)p(y_n|x_n, u_n^2)$. Furthermore, if each U_n^1 (respectively U_n^2) takes its values from a finite set $\Lambda^1 = \{\lambda_1^1, \dots, \lambda_M^1\}$ (respectively $\Lambda^2 = \{\lambda_1^2, \dots, \lambda_L^2\}$), $p(x|u^1)$ (respectively $p(y|u^2)$) is then non stationary. The new underlying process U^2 , introduced here, may model the presence of shadows in aerial images or clouds in some SPOT images, which may affect the visual aspect of the hidden classes that are to be recovered from Y .

3.2. MPM Restoration

The *MPM* Bayesian decision rule is computed according to $[\hat{x} = \hat{x}_{MPM}(y)] \Leftrightarrow [\hat{x}_n = \arg\max_i \phi_n(i)]$ with $\phi_n(i) = p(x_n = \omega_i | y)$. *MPM* is a Bayesian technique that ensures a minimal misclassification rate when N is large enough.

To make segmentation possible, we have to compute the posterior distributions $p(x_n | y)$. As $T = (U, X, Y)$ is a *TMC*, $T = (V, Y)$ is then a *PMC* with $V = (U, X)$. Subsequently, the marginal posterior distributions $p(v_n | y)$ are workable thanks to the modified forward and backward functions [4]. In fact, we have $p(v_n | y) \propto \alpha_n(v_n)\beta_n(v_n)$ where $\alpha_n(v_n) = p(y_1, \dots, y_n | v_n)$ and $\beta_n(v_n) = p(y_{n+1}, \dots, y_N | y_n, v_n)$. Both of them are computed using this iterative method derived from the very well-known Baum's algorithm [1]:

$$\alpha_{n+1}(v_{n+1}) = \sum_{v_n} \alpha_n(v_n) p(v_{n+1} | v_n) p(y_{n+1} | v_{n+1}), \quad (2)$$

$$\beta_n(v_n) = \sum_{v_{n+1}} \beta_{n+1}(v_{n+1}) p(v_{n+1} | v_n) p(y_{n+1} | v_{n+1}), \quad (3)$$

with $p(v_{n+1} | v_n) = p(u_{n+1} | u_n) p(x_{n+1} | x_n, u_{n+1}^1)$ and $p(y_n | v_n) = p(y_n | x_n, u_n^2)$ as specified in section 2.

The posterior distributions are then derived as follows

$$p(x_n | y) = \sum_{u_n} p(v_n | y) \quad (4)$$

and the estimates of U may be derived in the same way.

$$p(u_n | y) = \sum_{x_n} p(v_n | y) \quad (5)$$

3.3. Parameters estimation

In this paper, we consider Gaussian noise densities. Accordingly, the noise parameters to be estimated are $K \times L$ means $\mu = (\mu_{i,j})$ and $K \times L$ standard deviations $\sigma = (\sigma_{i,j})$ with $1 \leq i \leq K$ and $1 \leq j \leq L$. On the other hand, the distribution of U which is assumed to be *PMC* here is given by the matrix $A = [a_{\lambda,\lambda'}]_{\lambda,\lambda' \in \Lambda}$ where $a_{\lambda,\lambda'} = p(u_n = \lambda' | u_{n-1} = \lambda)$, i.e. $(M \times L)^2$ parameters. Finally, X is governed by M matrices $P = (P_m)_{m=1}^M$ with K^2 parameters per each. For parameters estimation, one can use either the Expectation-Maximization algorithm (*EM*) or its stochastic version *SEM*. In this paper, we used *EM*.

The parameters are then estimated in the following iterative manner:

- Choose an initial set of parameters $\Theta^0 = (A^0, P^0, \mu^0, \sigma^0)$.
- For each iteration, we compute $\psi_n(v_n, v_{n+1}) = p(v_n, v_{n+1} | y)$ and $\xi_n(v_n) = p(v_n | y)$ according to Θ^q .

$$\psi_n(v_n, v_{n+1}) \propto \alpha_n(v_n) A_{u_n u_{n+1}}^q P_{u_{n+1}^1, x_n, x_{n+1}}^q \beta_{n+1}(v_{n+1}) \quad (6)$$

$$\xi_n(v_n) = \sum_{v_{n+1}} \psi_n(v_n, v_{n+1}) \quad (7)$$

Then we derive Θ^{q+1} as follows

$$\mu_{i,j}^{q+1} = \frac{\sum_{n=1}^N \sum_{v_n} \xi_n(v_n) y_n \mathbf{1}_{[(x_n, u_n^2) = (\omega_i, \lambda_j^2)]}}{\sum_{n=1}^N \sum_{v_n} \xi_n(v_n) \mathbf{1}_{[(x_n, u_n^2) = (\omega_i, \lambda_j^2)]}} \quad (8)$$

$$\sigma_{i,j}^{q+1} = \frac{\sum_{n=1}^N \sum_{v_n} \xi_n(v_n) (y_n - \mu_{i,j}^{q+1})^2 \mathbf{1}_{[(x_n, u_n^2) = (\omega_i, \lambda_j^2)]}}{\sum_{n=1}^N \sum_{v_n} \xi_n(v_n) \mathbf{1}_{[(x_n, u_n^2) = (\omega_i, \lambda_j^2)]}} \quad (9)$$

$$a_{\lambda,\lambda'}^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(u_n, u_{n+1}) = (\lambda, \lambda')]} \mathbf{1}_{[u_n = \lambda]}}{\sum_{n=1}^{N-1} \sum_{v_n} \xi_n(v_n) \mathbf{1}_{[u_n = \lambda]}} \quad (10)$$

$$P_{m,\omega,\omega'}^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(u_{n+1}^1, x_n, x_{n+1}) = (\lambda_m^1, \omega, \omega')]} \mathbf{1}_{[(u_{n+1}^1, x_n) = (\lambda_m^1, \omega)]}}{\sum_{n=1}^{N-1} \sum_{v_n} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(u_{n+1}^1, x_n) = (\lambda_m^1, \omega)]}} \quad (11)$$

- We repeat the previous step until an end criterion is reached.

4. EXPERIMENTS

The aim of this section is to test the performance of the proposed *NSN- TMC* and to demonstrate the gain in classification accuracy when taking into account the heterogeneous behavior of noise densities, which are assumed to be Gaussian here. For this purpose, we

achieved some experiments on 128×128 images. In this paper, we show the results of *MPM* restoration on two images using the *EM* estimation algorithm.

4.1. Experimental images

Let us consider the case where $\Omega = \{\omega_1, \omega_2\}$. Experiments are conducted on the following image sets, where images are converted to and from one-dimensional sequences using the Hilbert- Peano scan [5]:

- Experiment 1: We generate a 128×128 synthetic image in the following way: $U = (U^1, U^2)$ is Markovian with $\Lambda^1 = \{\lambda_1^1, \lambda_2^1, \lambda_3^1\}$ and $\Lambda^2 = \{\lambda_1^2, \lambda_2^2\}$. The first realization u_1 is drawn randomly (uniformly) from $\Lambda^1 \times \Lambda^2$ and the following realizations are generated according to transition probabilities given by the matrix

$$A_{\lambda, \lambda'} = \frac{1}{10^3} \begin{bmatrix} 955 & 3 & 1 & 0 & 1 & 0 \\ 3 & 955 & 0 & 1 & 0 & 1 \\ 1 & 0 & 955 & 3 & 1 & 0 \\ 0 & 1 & 3 & 955 & 0 & 1 \\ 1 & 0 & 1 & 0 & 955 & 3 \\ 0 & 1 & 0 & 1 & 3 & 955 \end{bmatrix}$$

$p(x|u^1)$ is then non stationary. The first realization x_1 is randomly drawn from Ω and the following ones are generated according to U^1 where

$$P_{\lambda_1^1} = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}, P_{\lambda_2^1} = \begin{bmatrix} 0.77 & 0.23 \\ 0.23 & 0.77 \end{bmatrix} \text{ and } P_{\lambda_3^1} = \begin{bmatrix} 0.55 & 0.45 \\ 0.45 & 0.55 \end{bmatrix}.$$

The resulting class- image is then noised according to X and U^2 as follows: $p(y_n|x_n = \omega_i, u_n^2 = \lambda_j^2) = f_{i,j}(y_n)$

where $f_{i,j}$ is a Gaussian density with $\mu_{i,j}$ as mean and $\sigma_{i,j}$ as standard deviation. In this experiment, we used the following parameters $\mu_{i,j} = \begin{bmatrix} -1.25 & -0.75 \\ 0.75 & 1.25 \end{bmatrix}, \sigma_{i,j} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

Segmentation was also performed using real parameters of the proposed NSN- *TMC* to compare its results with those obtained using a SN- *TMC* considering densities stationary using the following average parameters

$$\mu_{i,j} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \sigma_{i,j} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

- Experiment 2: We considered the Nazca bird non stationary image (Fig. 2) that we noised using the same densities parameters of the first experimentation according to X and U^2 . This latter was generated randomly using the following transition matrix:

$$A_{u_1^2, u_2^2} = \begin{bmatrix} 0.998 & 0.002 \\ 0.002 & 0.998 \end{bmatrix}$$

The first realization u_1^2 was randomly drawn from Λ^2 .

Notice that authentic U is unknown here and is to be estimated.

4.2. Experimental results

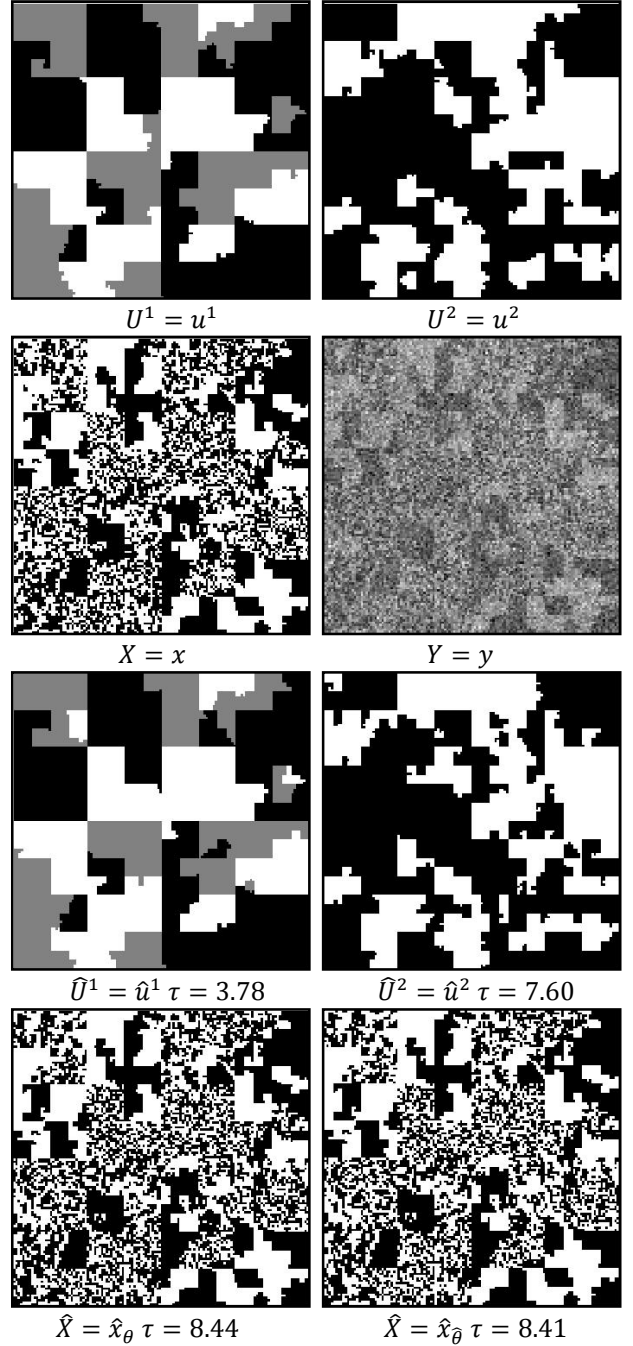


Figure 1. *MPM* restoration of a synthetic non stationary image.

The restoration results of the first experiment using *EM* estimation algorithm are presented in Fig.1. The misclassification rates of *MPM* restoration based on real and estimated parameters of the proposed NSN-*TMC* and the SN-*TMC* are given in tab. 1

For the second experiment, the restoration results are provided in Fig. 2. The corresponding misclassification rates are given in tab. 1.

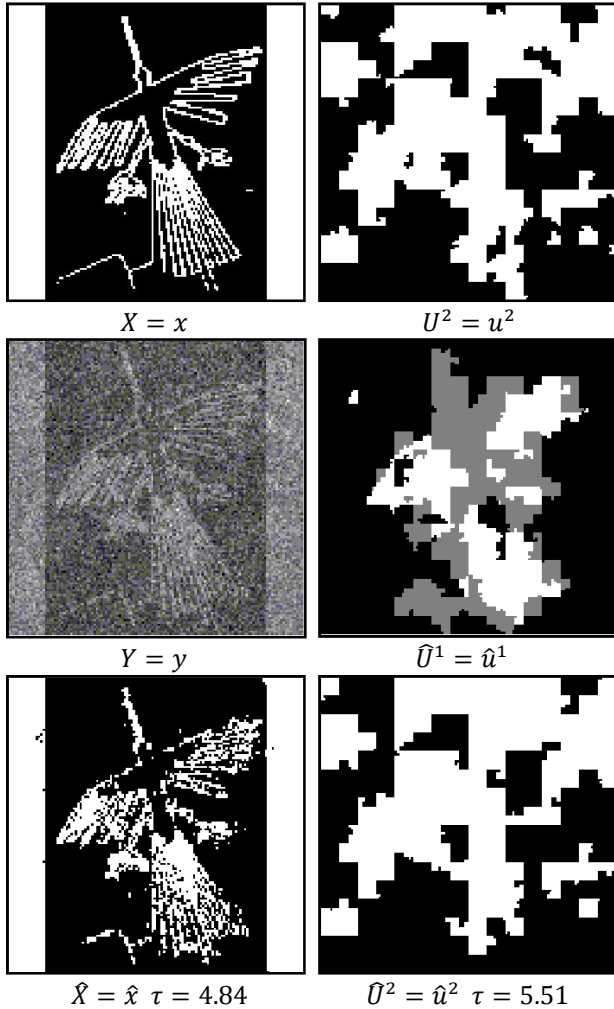


Figure 2. MPM restoration of Nazca bird non stationary image.

TABLE I.
MISCLASSIFICATION RATES OF MPM RESTORATION

Experiment	Experiment 1				Experiment 2	
	NSN TMC- θ	SN TMC- $E(\theta)$	NSN TMC- $\hat{\theta}$	SN TMC- $\hat{\theta}$	NSN TMC- $\hat{\theta}$	SN TMC- $\hat{\theta}$
τ_X (%)	8.44	9.20	8.41	9.23	4.84	5.36
τ_{U^1} (%)	2.86	3.70	3.78	4.33	-	-
τ_{U^2} (%)	5.76	-	7.60	-	5.51	-

As stated by Bayesian theory, taking noise non stationary aspect into account yields better results as illustrated in tab. 1. This demonstrates again the ability of triplet Markov chains to model some kinds of data that classical models cannot handle.

The interest of the last experiment is crucial since we deal with a non generated hidden process. As shown in Fig.2, our model allow to satisfactorily restore both processes U^1 and U^2 . The first represents the three stationary parts of the hidden process: black labels are assigned to the image background, grey labels to the bird body and white labels to the bird wings and primaries. On the other hand, the second process presents the two noise stationary parts where black labels are assigned to the dark part of the image and white labels to the bright one.

5. CONCLUSION

In this paper, we proposed a new triplet Markov chain to model switching data hidden with non stationary independent noise. We have chosen unsupervised image segmentation via Hibert-Peano as illustrative application field and showed how the new model can take such situations into account. As perspective for further work, we propose to extend the present model in three directions. First, we intend to consider the case where U^2 is continuous which would have more intuitive meaning when, for instance, this latter represents light. Second, we may consider a more general TMC when $p(z|u)$ follows a PMC. Third, we suggest considering TMC for situations where we dispose of more than one image taken in different conditions (different day-times for instance) or provided by different sensors with different sensibility degrees towards external conditions like light ones (modeled by U^2).

REFERENCES

- [1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A Maximization Technique Occurring in the Statistical analysis of the Probabilistic Functions of Markov Chains", *Ann. Math. Stat.*, Vol. 41, pp. 164-171, 1970.
- [2] M. Y. Boudaren, and A. Belaïd, "A New Scheme for Land Cover Classification in Aerial Images", *Electronic Engineering and Computing Technology, LNEE*, Chapter 40, Volume 60, pp. 471-482, Springer 2010.
- [3] O. Capé, E. Moulines, T. Ryden, *Inference in Hidden Markov Models*, Springer, Series in Statistics, 2005.
- [4] 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.
- [5] R. Fjortoft, 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 Geoscience and Remote Sensing*, Vol. 41, No. 3, pp. 675- 686, 2003.
- [6] T. Koski, *Hidden Markov Models for Bioinformatics*, Kluwer Academic Publishers, 2001.
- [7] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, "Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise", *Signal Processing*, Vol. 91, No. 2, pp. 163-175, Feb. 2011.
- [8] W. Pieczynski, "Multisensor triplet Markov chains and theory of evidence", *International Journal of Approximate Reasoning*, Vol. 45, No. 1, pp. 1-16, 2007.
- [9] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. of IEEE*, vol. 77, no. 2, pp. 257-286, 1989.