SPIE's International Symposium on Remote Sensing, September 22-27, Crete, Greece, 2002

Triplet Markov Chains in hidden signal restoration

Wojciech Pieczynski, Cédric Hulard, and Thomas Veit[1]
Département CITI; Institut National des Télécommunications, Evry, France

ABSTRACT

Hidden Markov Chain (HMC) models are widely used in various signal or image restoration problems. In such models, one considers that the hidden process $X = (X_1, ..., X_n)$ we look for is a Markov chain, and the distribution $p(y|x)$ of the observed process $Y = (Y_1, ..., Y_n)$, conditional on X, is given by $p(y|x) = \prod_{i=1}^{n} p(y_i|x_i)$. The "a posteriori" distribution $p(x|y)$ of X given $Y = y$ is then a Markov chain distribution, which makes possible the use of different Bayesian restoration methods. Furthermore, all parameters can be estimated by the general "Expectation-Maximization" algorithm, which renders Bayesian restoration unsupervised.

This paper is devoted to an extension of the HMC model to a "Triplet Markov Chain" (TMC) model, in which a third auxiliary process U is introduced and the triplet (X, U, Y) is considered as a Markov chain. Then a more general model is obtained, in which X can still be restored from $Y = y$. Moreover, the model parameters can be estimated with Expectation-Maximization (EM) or Iterative Conditional Estimation (ICE), making the TMC based restoration methods unsupervised. We present a short simulation study of image segmentation, where the bi- dimensional set of pixels is transformed into a mono-dimensional set via a Hilbert-Peano scan, that shows that using TMC can improve the results obtained with HMC.

Keywords: Hidden Markov chains, Pairwise Markov chains, Triplet Markov chains, parameter estimation, Bayesian restoration, statistical signal segmentation, statistical image segmentation, theory of evidence, Dempster-Shafer fusion, EM algorithm

1. INTRODUCTION

The modeling by hidden Markov chains (HMC) is widely used in various problems [1, 3, 5, 7], among others. It consists of considering two stochastic processes $X = (X_n)_{s \in N}$ and $Y = (Y_n)_{s \in N}$, where X is a Markov chain. The unobservable realizations $X = x$ are of interest and have to be estimated from the observed $Y = y$. The distribution of (X, Y) that models the stochastic interactions between what is seen and what is wanted is then given by a Markov distribution $p(x)$ of X and the distributions $p(y|x)$ of Y conditional on $X = x$. When the latter are simple enough, the pair (X, Y) keeps the same Markovian form of distribution, and it is the same for the distribution $p(x|y)$ of X conditional on $Y = y$. The Markovianity of $p(x|y)$ is crucial because it allows one to estimate the unobservable $X = x$ from the observed $Y = y$ by some Bayesian methods, even in the case of very large number of observations. However, the simplicity of $p(y|x)$, which is frequently taken of the form $p(y|x) = \prod_{i=1}^{n} p(y_i|x_i)$, is often difficult to justify. To remedy the latter, the HMC model has been recently generalized to "Pairwise Markov Chains" (PMC [27]), in which one directly assumes the Markovianity of the pairwise model $Z = (X, Y)$; it is to say, the process $(Z_n)_{n \in N} = (X_n, Y_n)_{n \in N}$ is a Markov chain. A HMC is then a PMC, but a PMC is not necessarily a HMC and thus one obtains a strictly more general model. This larger generality allows one to take into account more complex ditributions

[1] E-mails : Wojciech.Pieczynski@int-evry.fr, http://www-citi.int-evry.fr/~pieczyns/, Cedric.Hulard@int-evry.fr , Thomas.Veit@int-evry.fr

$p(y|x)$, the latter being very useful to better model complex real situations. Of course, PMC are of interest with respect to HMC because all Bayesian processing possible in HMC context remain possible in the PMC one. Let us notice that an analogous generalization has been previously performed in the context of Markov fields, whose applications in image processing are well known. In fact, the Hidden Markov Fields (HMF [17, 19, 20]) model has been generalized to Pairwise Markov Fields (PMF) models [26], and different experiments in an unsupervised context has been presented in [32].
In this paper we will deal with "Triplet Markov Chain" (TMC) model, which has been proposed in [25] *in french*, and which generalizes the PMC model. More precisely, our aim is to present some further comments, specify some further properties, and present some first experiments showing how TMC allows one to improve the results obtained with PMC, especially in statistical image segmentation.

2. TRIPLET MARKOV CHAINS

2.1 General Properties

Let $X = (X_n)_{n\in N}$ be the searched process and $Y = (Y_n)_{n\in N}$ the observed one. The problem is to estimate X from Y. We assume that for each $n \in N$, the variables X_n and Y_n take their values in $\Omega = \{\omega_1, ..., \omega_k\}$ and R, respectively.

Definition 3.1
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 $T = (X, U, Y) = ((X_n, U_n, Y_n))_{n\in N}$ is a Markov chain.

Let us insist on the fact that only X and Y have physical existence and the auxiliary process U is only a tool for different calculus. So, the idea of TMC is to consider the distribution of (X, Y), which models the interactions between the observed and the searched processes, as a marginal distribution of a Markov chain $T = (X, U, Y)$. The Pairwise Markov Chain (PMC [10, 11, 27]) model, in which $Z = (X, Y)$ is assumed to be a Markov chain, is then a particular case of the TMC because the former is obtained by taking $\Omega = \Lambda$ and $U = X$. Furthermore, TMC is strictly more general than PMC, the latter being strictly more general than classical HMC, in that there exist Markov chains $T = (X, U, Y)$ such that $Z = (X, Y)$ are not Markov chains.
Let us specify how the Bayesian Maximum Posterior Mode (MPM) can be performed in the frame of TMC. Let us consider a finite number of variables $T = (T_1, ..., T_n) = ((X_1, U_1, Y_1), ..., (X_n, U_n, Y_n))$. According to the general Bayesian theory, the MPM restoration $\hat{x}^{MPM} = (\hat{x}_1^{MPM}, ..., \hat{x}_n^{MPM})$ is such that for each $1 \le i \le n$

$$p(\hat{x}_i^{MPM} | y) = \max_{\omega\in\Omega} p(x_i = \omega | y) \tag{2.1}$$

So, the problem is to calculate the posterior marginal probabilities $p(x_i|y)$. Let us consider the variables $Z_i = (X_i, Y_i)$, $V_i = (X_i, U_i)$, and Z, V the corresponding processes. Given that T is Markovian, the process (V, Y) is a PMC and so we can write the distribution of (V_i, Y) as $p(v_i, y) = \alpha^i(v_i)\beta^i(v_i)$, with $\alpha^i(v_i)$ (« Forward » probabilities) and $\beta^i(v_i)$ (« Bakward » probabilities) defined by $\alpha^i(v_i) = p(y_1, ..., y_{i-1}, y_i, v_i)$ and $\beta^i(v_i) = p(y_{i+1}, ..., y_n | v_i, y_i)$. Furthermore, $\alpha^i(v_i)$ and $\beta^i(v_i)$ are computable by the following « Forward » and « Backward » procedures [10, 27]

$$\alpha^1(v_1) = p(y_1, v_1) \text{, et } \alpha^{i+1}(v_{i+1}) = \sum_{v_i\in\Omega\times\Lambda} \alpha^i(v_i) p(y_{i+1}, v_{i+1} | y_i, v_i) \text{ for } 0 \le i \le n-1 \tag{2.2}$$

$$\beta^n(v_n) = 1 \text{, et } \beta^i(v_i) = \sum_{v_{i+1}\in\Omega\times\Lambda} \beta^{i+1}(v_{i+1}) p(y_{i+1}, v_{i+1} | y_i, v_i) \text{ for } 0 \le i \le n-1 \tag{2.3}$$

Having calculated $p(v_i, y) = \alpha^i(v_i)\beta^i(v_i)$, we then can calculate $p(x_i, y)$ by

$$p(x_i, y) = \sum_{u_i \in \Lambda} p(x_i, u_i, y) = \sum_{u_i \in \Lambda} p(v_i, y), \qquad (2.4)$$

which gives $p(x_i|y)$. Finally, the marginals $p(x_i|y)$ are computable, which makes possible the use of the Bayesian MPM restoration method given by (2.1).

Remark 2.1

When the TMC is a PMC, it is to say when $X = U$, we have $X = V$ and so $p(x_i, y) = \alpha^i(x_i)\beta^i(x_i)$ is given by (2.2), (2.3) and (2.4) is not needed. This means that we obtain the PMC « Forward » and « Backward » formulas. Furthermore, a PMC is defined by the distributions $p(z_i, z_{i+1})$ and when the latter are of the form $p(z_i, z_{i+1}) = p(x_i, x_{i+1})p(y_i|x_i)p(y_{i+1}|x_{i+1})$, the PMC is a HMC. In such a case, the PMC « Forward » and « Backward » formulas (2.2), (2.3) become the classical HMC « Forward » and « Backward » formulas [27].

The following result, extracted from [25], shows that TMC model is strictly more general that the PMC one.

Proposition 2.1

Let $T = (X, U, Y)$ a TMC verifying :

(H) For every $i \in N$, $z_i = z_{i+2}$ implies $p(u_{i+1}|z_i, z_{i+1}) = p(u_{i+1}|z_{i+1}, z_{i+2})$ (recall that $Z = (X, Y)$ **)**.

Then $Z = (X, Y)$ is a PMC if and only if for each $i \in N$, $p(u_{i+1}|z_i, z_{i+1}) = p(u_{i+1}|z_{i+1})$.

Proof

Let us show that $Z = (X, Y)$ is a PMC if and only if

$$\text{For every } i \in N \text{ and } z_i, z_{i+1}, z_{i+2}, \sum_{u_{i+1} \in \Omega} \frac{p(u_{i+1}|z_i, z_{i=1})\, p(u_{i+1}|z_{i+1}, z_{i+2})}{p(u_{i+1}|z_{i+1})} = 1 \qquad (2.5)$$

In fact, Z_1, …, Z_n is a Markov chain if and only if for every $i \geq 1$ the distribution of (Z_i, Z_{i+1}, Z_{i+2}) is written

$$p(z_i, z_{i+1}, z_{i+2}) = \frac{p(z_i, z_{i+1})\, p(z_{i+1}, z_{i+2})}{p(z_{i+1})} \qquad (2.6)$$

Otherwise, the distribution of (Z_i, Z_{i+1}, Z_{i+2}) is the marginal distribution of $(T_i, T_{i+1}, T_{i+2}) = (U_i, Z_i, U_{i+1}, Z_{i+1}, U_{i+2}, Z_{i+2})$. The latter being a Markov distribution, we have

$$p(z_i, z_{i+1}, z_{i+2}) = \sum_{u_i, u_{i+1}, u_{i+2} \in \Lambda} p(t_i, t_{i+1}, t_{i+2}) = \sum_{u_i, u_{i+1}, u_{i+2} \in \Lambda} \frac{p(t_i, t_{i+1})\, p(t_{i+1}, t_{i+2})}{p(t_{i+1})} = \sum_{u_i, u_{i+1}, u_{i+2} \in \Lambda} \frac{p(z_i, u_i, z_{i+1}, u_{i+1})\, p(z_{i+1}, u_{i+1}, z_{i+2}, u_{i+2})}{p(z_{i+1}, u_{i+1})} =$$

$$\sum_{u_i, u_{i+1}, u_{i+2} \in \Lambda} \frac{p(z_i, z_{i+1})(u_i, u_{i+1}|z_i, z_{i+1})\, p(z_{i+1}, z_{i+2})\, p(u_{i+1}, u_{i+2}|z_{i+1}, z_{i+2})}{p(z_{i+1})\, p(u_{i+1}|z_{i+1})} =$$

$$\frac{p(z_i, z_{i+1})\, p(z_{i+1}, z_{i+2})}{p(z_{i+1})} \sum_{u_i, u_{i+1}, u_{i+2} \in \Lambda} \frac{p(u_i, u_{i+1}|z_i, z_{i+1})\, p(u_{i+1}, u_{i+2}|z_{i+1}, z_{i+2})}{p(u_{i+1}|z_{i+1})} =$$

$$\frac{p(z_i, z_{i+1})\, p(z_{i+1}, z_{i+2})}{p(z_{i+1})} \sum_{u_{i+1} \in \Lambda} \frac{p(u_{i+1}|z_i, z_{i+1})\, p(u_{i+1}|z_{i+1}, z_{i+2})}{p(u_{i+1}|z_{i+1})}$$

thus

$$p(z_i, z_{i+1}, z_{i+2}) = \frac{p(z_i, z_{i+1}) p(z_{i+1}, z_{i+2})}{p(z_{i+1})} \sum_{u_{i+1} \in \Lambda} \frac{p(u_{i+1} | z_i, z_{i+1}) p(u_{i+1} | z_{i+1}, z_{i+2})}{p(u_{i+1} | z_{i+1})} \tag{2.7}$$

which shows that (2.5) is equivalent to (2.6).
Let us show that (2.5) is equivalent to $p(u_{i+1} | z_i, z_{i+1}) = p(u_{i+1} | z_{i+1})$.

For z_{i+1} fixed, let us consider the scalar product defined on R^k by $\langle a, b \rangle = \sum_{j=1}^{k} \frac{a_j b_j}{p(u_i = \omega_j | z_{i+1})}$. Let $a_j^z = p(u_{i+1} = \omega_j | z_i = z, z_{i+1}) = p(u_{i+1} = \omega_j | z_{i+1}, z_{i+2} = z)$, the equality coming from the hypothesis (H). So, according to (3.3), we have a family of vectors $a^z = (a_1^z, \dots a_k^z)$ verifying $\langle a^z, a^{z'} \rangle = 1$ for every z, z'. This implies that a^z does not depend on z ; in fact, $\left\| a^z - a^{z'} \right\|^2 = \langle a^z, a^z \rangle^2 + \langle a^{z'}, a^{z'} \rangle^2 - 2\langle a^z, a^{z'} \rangle = 0$, which implies $a^z = a^{z'}$. So, a_i^z do not depend on z, which means $p(u_2 = \omega_i | z_1 = z, z_2) = p(u_2 = \omega_i | z_2, z_3 = z) = p(u_2 = \omega_i | z_2)$, which ends the proof.

Example 2.1
Let U be a Markov chain and let us assume the random variables (Z_i) independent conditionally to U, with $p(z_i | u) = p(z_i | u_i)$. The distribution of T is then given by $p(t_1) = p(u_1) p(z_1 | u_1)$ and $p(t_{i+1} | t_i) = p(u_{i+1} | u_i) p(z_{i+1} | u_{i+1})$. So T is a Markov chain and thus, accordingly to what precedes, such a model can be used to estimate X from Y. However, this model is not, in general, neither a HMC model nor a PMC model : neither X nor (X, Y) are Markov chains in general.

2.2 Evidential Context
The Dempster-Shafer theory of evidence can be seen, in some situations, as an useful extension of the probability theory [2, 6, 15, 22, 29, 30, 31, 35]. In particular, the calculus of the "posterior" distribution $p(x|y)$ needed in Bayesian processing can be seen as a particular case of the so-called " Dempster-Shafer fusion rule". The usefulness of Markov models in Bayesian signal processing being well known, the problem of generalizing them to some "evidential" contexts is of interest. Until now, very few papers consider the problem of signal processing using simultaneously hidden Markov modeling and Dempster-Shafer one. Considering the classical Hidden Markov Model (HMM), which can be HMC or HMF, $p(x)$ is a Markov distribution and the distributions $p(y|x)$ are of the simple form $p(y|x) = \prod_{i=1}^{n} p(y_i | x_i)$. Keeping $p(x)$ as a Markov probability distribution, it is then possible to generalize $p(y|x)$ to some "evidential measure" in such a way that the Dempster-Shafer fusion of $p(x)$ with the latter "evidential measure" gives a Markov distribution, which generalizes the classical posterior distribution $p(x|y)$ [4, 12, 33]. When wishing to generalize HMM to situations in which the prior Markov distribution $p(x)$ is replaced by some prior "Markov evidential measure" things become more difficult because, roughly speaking, the Dempster-Shafer fusion destroys the Markovianity (particular cases of this problem are studied in [13, 21]). Following [25], we establish below a formal link between hidden "evidential" Markov chains and TMC ones. So, in spite of the lack of Markovianity of the fused distribution, Bayesian MPM segmentation can be performed using (2.1)-(2.4).
Let us consider a finite set $\Omega = \{\omega_1, \dots, \omega_k\}$ of classes and its power set $P(\Omega) = \{A_1, \dots, A_q\}$, with $q = 2^k$. A function M from $P(\Omega)$ to $[0,1]$ is called a « mass function » if $M(\emptyset) = 0$ and $\sum_{A \in P(\Omega)} M(A) = 1$. A mass function M defines then a plausibility function Pl from $P(\Omega)$ to $[0,1]$ by $Pl(A) = \sum_{A \cap B \neq \emptyset} M(B)$, and a « credibility » function Cr from $P(\Omega)$ to $[0,1]$ by $Cr(A) = \sum_{B \subset A} M(B)$. We see how a mass function generalizes a probability function : when M is null except on a partition of Ω, then the corresponding Pl and Cr functions are equal and are a probability distribution. For example, if we have a family of possible probability distribution $(p_\theta)_{\theta \in \Theta}$, the following «upper» and «lower»

probabilities $\hat{p}(A) = \sup_{\theta \in \Theta} p_\theta(A)$, $\check{p}(A) = \inf_{\theta \in \Theta} p_\theta(A)$ are Pl and Cr functions associated with the mass function M defined by $M(A) = \sum_{B \subset A} (-1)^{Card(A-B)} Cr(B)$, the latter formula being valid once Cr is defined by M.

Such models can be useful in numerous, even very simple, situations. For example, let us imagine a remote sensing image of a part of a region in which there are forest and water. A half of the region is known and it is established that the proportion of water is 0.25; the other half is not known and nothing can be said about the proportion of the water. Furthermore, it is not known in what part of the region the image has been taken. How to model such a prior knowledge? In a strictly probabilistic framework one could say that on each pixel $p^0(W) = 0.25$ and $p^0(F) = 0.75$. In a evidential framework, it is possible to take whole the information available by taking $M^0(\{W\}) = 0.125$, $M^0(\{F\}) = 0.375$, and $M^0(\{W, F\}) = 0.5$.

When two mass functions M_1, M_2 represent two pieces of evidence, we can combine – or fuse - them using the so called « Dempster-Shafer combination rule ». The result $M = M_1 \oplus M_2$ is defined by :

$$M(A) = (M_1 \oplus M_2)(A) \propto \sum_{B_1 \cap B_2 = A} M_1(B_1) M_2(B_2) \tag{2.8}$$

We will say that a mass function M is « probabilistic » when, being null outside singletons, it defines a probability. One can then see that when either M_1 or M_2 is probabilistic, then the fusion result M is probabilistic. In particular, one may see that calculus of the posterior probability is a Dempster-Shafer fusion (DS fusion) of two probabilistic mass functions. For example, let $X = (X_1, .., X_n)$, $Y = (Y_1, .., Y_n)$ be the classical hidden Markov chain, with $p(x, y) = p(x_1) p(x_2 | x_1) ... p(x_n | x_{n-1})\ p(y_1 | x_1)(y_2 | x_2) ... p(y_n | x_n)$. The posterior distribution $p(x|y)$ of X can then be seen as the DS fusion of the probabilistic masses $M_1(x) = p(x_1) p(x_2 | x_1) ... p(x_n | x_{n-1})$ and $M_2(x) = M_2(x_1, ..., x_n) = \dfrac{p(y_1 | x_1)(y_2 | x_2) ... p(y_n | x_n)}{\sum_{x' \in \Omega^{n+1}} p(y_1 | x'_1)(y_2 | x'_2) ... p(y_n | x'_n)}$. This means that the DS fusion is more general and can possibly take into account problems in which the strictly probabilistic models are difficult to apply. For example, let us consider the problem of image segmentation into two classes "water" and "forest" mentioned above. Working "pixel by pixel" we have to decide, from the gray level y_s, whether there is "water" or "forest" on s. Let us assume that y_s is distributed according to f_W (f_F respectively) when there is water (forest, respectively) on s. In the probabilistic framework, we calculate the posterior probability

$$p(W | y_s) = \frac{p^0(W) f_W(y_s)}{p^0(W) f_W(y_s) + p^0(F) f_F(y_s)},\ p(F | y_s) = \frac{p^0(F) f_F(y_s)}{p^0(W) f_W(y_s) + p^0(F) f_F(y_s)} \tag{2.9}$$

In the evidential framework, the DS fusion (2.8) applied to M^0 defined above by $M^0(\{W\}) = 0.125$, $M^0(\{F\}) = 0.375$, $M^0(\{W, F\}) = 0.5$, and the "probabilistic" mass function M^1 defined by $M^1(\{W\}) = \dfrac{f_W(y_s)}{f_W(y_s) + f_F(y_s)}$, $M^1(\{F\}) = \dfrac{f_F(y_s)}{f_W(y_s) + f_F(y_s)}$, $M^1(\{W, F\}) = 0$ gives $M = M_1 \oplus M_2$ defined by

$$\begin{aligned} M(\{W\}) &= \frac{(M^0(\{W\}) + M^0(\{W, F\})) f_W(y_s)}{(M^0(\{W\}) + M^0(\{W, F\})) f_W(y_s) + (M^0(\{F\}) + M^0(\{W, F\})) f_F(y_s)}, \\ M(\{F\}) &= \frac{(M^0(\{F\}) + M^0(\{W, F\})) f_F(y_s)}{(M^0(\{W\}) + M^0(\{W, F\})) f_W(y_s) + (M^0(\{F\}) + M^0(\{W, F\})) f_F(y_s)} \end{aligned} \tag{2.10}$$

we see that probability (2.10) is the classical posterior probability (2.9) when the "evidential" priors M^0 become "probabilistic" priors p^0, it is to say when $M^0(\{W,F\})=0$.
The aim of what follows is to propose a workable "evidential Markovian" model for priors.

Definition 2.2
A mass function M defined on $\mathrm{P}(\Omega^n)$ will be called « evidential Markov chain » (EMC) if it is null outside $[\mathrm{P}(\Omega)]^n$ and if it can be written

$$M(A_1, A_2, ..., A_n) = M(A_1)M(A_2|A_1),...,M(A_n|A_{n-1}) \qquad (2.11)$$

Let us remark that a probabilistic EMC (which is null outside $A=(A_1, A_2, ..., A_n)$ such that all $A_1, A_2, ..., A_n$ are singletons) is a classical Markov chain.
Following [25] we can then state

Proposition 2.2
Let M^0 be an EMC on $[\mathrm{P}(\Omega)]^n$, and M^1 a probability on $\mathrm{P}(\Omega^n)$ defined from the observed process $Y=y\in R^n$ by $M^1(x_1,...,x_n) \propto p(y_1|x_1)...p(y_n|x_n)$.
Then the probability distribution $M = M^0 \oplus M^1$ is a TMC, with $\Lambda = \mathrm{P}(\Omega)$ and T Markov chain whose distribution is defined by

$$p(t_1) \propto 1_{[x_1\in u_1]} M^0(u_1) p(y_1|x_1) \text{, and } p(t_{i+1}|t_i) \propto 1_{[x_i\in u_i]} 1_{[x_{i+1}\in u_{i+1}]} M^0(u_{i+1}|u_i) p(y_{i+1}|x_{i+1}) \qquad (2.12)$$

Proof
Let $x=(x_1,...,x_n)\in\Omega^n$. The DS fusion is written (in the two first sums $x=(x_1,...,x_n)$ is fixed and $u=(u_1,...,u_n)$ varies in $[\mathrm{P}(\Omega)]^n$ in such a way that $x_1\in u_1,...,x_n\in u_n$) :

$$(M^0\oplus M^1)(x) \propto \sum_{x\in u} M^0(u_1,...,u_n)M^1(x) \propto \sum_{x\in u} M^0(u_1)M^0(u_2|u_1)...M^0(u_n|u_{n-1})M^1(x) =$$
$$\sum_{u\in \mathrm{P}(\Omega))^n} 1_{[x_1\in u_1]}...1_{[x_n\in u_n]} M^0(u_1)M^0(u_2|u_1)...M^0(u_n|u_{n-1})M^1(x) = \sum_{u\in \mathrm{P}(\Omega))^n} p(t_1)p(t_2|t_1)...p(t_n|t_{n-1}) = \sum_{u\in \mathrm{P}(\Omega))^n} p(t_1,...,t_n)$$

which ends the proof.

Remark 2.2
The TMC of the Proposition 2.2, defined by $p(t_1)\propto 1_{[x_1\in u_1]}M^0(u_1)p(y_1|x_1)$ and $p(t_{i+1}|t_i)\propto 1_{[x_i\in u_i]}1_{[x_{i+1}\in u_{i+1}]}M^0(u_{i+1}|u_i)p(y_{i+1}|x_{i+1})$, is a particular case of the TMC specified in Example 2.1 ; in fact, U is a Markov chain and $p(z_i|u_i)=p(x_i,y_i|u_i)\propto 1_{[x_i\in u_i]}M^0(u_i)p(y_i|x_i)$. As a result, we can state that (X,Y) is not a PMC. In other words, the DS fusion destroys the Markovianity but, the result being a TMC, Bayesian restorations are feasible.

3. EXPERIMENTS

We propose in this section two series of experiments. In the first series we consider data $Y=y$ simulated according to a simple TMC and they are then restored according to the correct TMC model on the one hand, and according to a classical HMC model, on the other hand. Of course, the very Bayesian theory indicates that the TMC based restoration will give better results; however, it is interesting to look at how large the difference can be. Furthermore, both TMC and HMC based restorations are made in supervised (parameters estimated from the complete data (X,U,Y)) and unsupervised (parameters estimated from the observed data Y) manner. In the latter case, the parameters can be estimated by adaptations of some general methods like Expectation-Maximization (EM) [9] or Iterative Conditional estimation (ICE) [23]. These two methods have been compared in the context of HMC with Gaussian noise and it turns out that their efficiencies are equivalent [5]. As here we use Gausian noise and given that EM is faster than ICE, we propose here the use of an original variant of EM (a theoretical comparison between EM and ICE can be seen in [8]). The second series concerns the image segmentation. In fact, transforming the pixels set into a monodimensional set using some Hilbert-Peano scan allows one to successfully use HMC, or PMC, in unsupervised image segmentation [5, 10, 11, 14] and so it is of interest to look at how TMC work in such context.

3.1 TMC and HMC

Let us consider a TMC $T=(X,U,Y)$, with $\Omega=\{\omega_1,\omega_2\}$ and $\Lambda=\{\lambda_1,\lambda_2\}$. So, we have two real classes and two auxiliary ones. Let the distribution of $T=(X,U,Y)$ be defined by a Markov distribution of U, and the distributions $p(x|u)=\prod_{1\le i\le n} p(x_i|u_i)$ and $p(y|u,x)=\prod_{1\le i\le n} p(y_i|u_i,x_i)$. So, the whole distribution $p(t)=p(x,u,y)$ is given by the distribution $p(u_1,u_2)$ on Λ^2, two distributions $p(x_1|u_1=\lambda_1)$, $p(x_1|u_1=\lambda_2)$ on Ω, and four Gaussian distributions $p(y_1|(u_1,x_1)=(\omega_1,\lambda_1))$, $p(y_1|(u_1,x_1)=(\omega_1,\lambda_2))$, $p(y_1|(u_1,x_1)=(\omega_2,\lambda_1))$, and $p(y_1|(u_1,x_1)=(\omega_2,\lambda_2))$.

We performed numerous simulations and three of them are presented in Tab. 1-3. These different experiments allow us to put forth the following general conclusions:

(i) TMC based MPM always gives smaller error ration that the HMP based one; however, the difference can be significant, as in Tab. 1, or negligible, as in Tab. 3;

(ii) the good efficiency of EM in the classical Gaussian HMC remains in the Gaussian TMC studied here; in fact, the supervised and unsupervised TMC based restorations are generally close enough to each other;

(iii) the degradation of efficiency, when passing from supervised restoration to unsupervised one, is more significant when using HMC that when using TMC.

Finally, important conclusion for practical applications is that when data suit a TMC model and the parameters are not known, the use of TMC and EM based unsupervised MPM restoration can be significantly better that the HMC and EM based one.

$p(x_1=\omega_1|u_1=\lambda_1)=$ $p(x_1=\omega_2|u_1=\lambda_2)$	$p(x_1=\omega_1|u_1=\lambda_2)=$ $p(x_1=\omega_2|u_1=\lambda_1)$	$p(u_1=\lambda_1,u_2=\lambda_1)=$ $p(u_1=\lambda_2,u_2=\lambda_2)$	$p(u_1=\lambda_1,u_2=\lambda_2)=$ $p(u_1=\lambda_2,u_2=\lambda_1)$
0.7	0.3	0.49	0.01
$p(y_1|(u_1,x_1)=(\omega_1,\lambda_1))$	$p(y_1|(u_1,x_1)=(\omega_1,\lambda_2))$	$p(y_1|(u_1,x_1)=(\omega_2,\lambda_1))$	$p(y_1|(u_1,x_1)=(\omega_2,\lambda_2))$
Mean=0, Var=1	Mean=2.5, Var=1	Mean=2.5, Var=1	Mean=5, Var=1
TMC MPM	HMC MPM	TMC MPM EM	HMC MPM EM
9.7%	17.7%	12.9%	29.2%

Tab. 1 ; Parameters and error ratios of supervised (real parameters) and unsupervised (parameters estimated with EM) Bayesian MPM restorations

$p(x_1=\omega_1|u_1=\lambda_1)=$ $p(x_1=\omega_2|u_1=\lambda_2)$	$p(x_1=\omega_1|u_1=\lambda_2)=$ $p(x_1=\omega_2|u_1=\lambda_1)$	$p(u_1=\lambda_1,u_2=\lambda_1)=$ $p(u_1=\lambda_2,u_2=\lambda_2)$	$p(u_1=\lambda_1,u_2=\lambda_2)=$ $p(u_1=\lambda_2,u_2=\lambda_1)$
0.7	0.3	0.49	0.01
$p(y_1|(u_1,x_1)=(\omega_1,\lambda_1))$	$p(y_1|(u_1,x_1)=(\omega_1,\lambda_2))$	$p(y_1|(u_1,x_1)=(\omega_2,\lambda_1))$	$p(y_1|(u_1,x_1)=(\omega_2,\lambda_2))$
Mean=0, Var=1	Mean=0.4, Var=1	Mean=1.6, Var=1	Mean=2, Var=1
TMC MPM	HMC MPM	TMC MPM EM	HMC MPM EM
15.8%	17.38%	21.48%	27.63%

Tab. 2 ; Parameters and error ratios of supervised (real parameters) and unsupervised (parameters estimated with EM) Bayesian MPM restorations

$p(x_1=\omega_1 \vert u_1=\lambda_1)=$ $p(x_1=\omega_2 \vert u_1=\lambda_2)$	$p(x_1=\omega_1 \vert u_1=\lambda_2)=$ $p(x_1=\omega_2 \vert u_1=\lambda_1)$	$p(u_1=\lambda_1,u_2=\lambda_1)=$ $p(u_1=\lambda_2,u_2=\lambda_2)$	$p(u_1=\lambda_1,u_2=\lambda_2)=$ $p(u_1=\lambda_2,u_2=\lambda_1)$
0.8	0.2	0.49	0.01
$p(y_1 \vert (u_1,x_1)=(\omega_1,\lambda_1))$	$p(y_1 \vert (u_1,x_1)=(\omega_1,\lambda_2))$	$p(y_1 \vert (u_1,x_1)=(\omega_2,\lambda_1))$	$p(y_1 \vert (u_1,x_1)=(\omega_2,\lambda_2))$
Mean=0, Var=1	Mean=0.8, Var=1	Mean=1.2, Var=1	Mean=2, Var=1
TMC MPM	HMC MPM	TMC MPM EM	HMC MPM EM
18.0%	18.6%	18.4%	20.0%

Tab. 3 ; Parameters and error ratios of supervised (real parameters) and unsupervised (parameters estimated with EM) Bayesian MPM restorations

3.2 Unsupervised image segmentation using TMC

The classical HMC have already been used in unsupervised image segmentation, in supervised or unsupervised manner, on monosensor or multisensor images [5, 14], and it is the same for the recent PMC model [10, 12]. So, we propose here an analogous work, using TMC instead of HMC and PMC. To do so, the bidimensional set of pixels has to be transformed into a monodimensional sequence, which is made through the Hilbert-Peano scan presented in Fig. 1. We will consider a TMC $T=(X,U,Y)$, with two real classes $\Omega=\{\omega_1,\omega_2\}$ and two auxiliary classes $\Lambda=\{\lambda_1,\lambda_2\}$. Let the distribution of $T=(X,U,Y)$ be defined by a Markov distribution of $V=(X,U)$, and $p(y \vert u,x)=\prod_{1\le i\le n} p(y_i \vert u_i,x_i)$. We notice that such a model is slightly more general than the model used in the previous section. So, the whole distribution $p(t)=p(x,u,y)$ here is given by the distribution $p(u_1,x_1,u_2,x_2)$ on $(\Lambda\times\Omega)^2$ and four Gaussian distributions $p(y_1 \vert (u_1,x_1)=(\omega_1,\lambda_1))$, $p(y_1 \vert (u_1,x_1)=(\omega_1,\lambda_2))$, $p(y_1 \vert (u_1,x_1)=(\omega_2,\lambda_1))$, and $p(y_1 \vert (u_1,x_1)=(\omega_2,\lambda_2))$.

We present below two supervised segmentation results. The four class images (first images in Fig. 1 and 2) are hand-written, which also gives the real two class images (second images in Fig. 1 and 2). The four class images are corrupted with Gaussian noises whose parameters are specified in Tab. 6. The distribution $p(u_1,x_1,u_2,x_2)$ on $(\Lambda\times\Omega)^2$ is estimated, from the chains obtained from the four classes images via Hilbert-Peano scan, by the classical empirical frequencies. The estimates obtained are specified in Tab. 4 and 5.

We notice that TMC based MPM restoration works better than the HMC based one; however, its advantage is less striking than in the previous subsection.

Remark 3.1

Let us notice that here the auxiliary classes can possibly have the real following signification. If we assume that in a remote sensed image the real classes $\Omega=\{\omega_1,\omega_2\}$ are " vegetation" and "urban area", there can be some subclasses (a vegetation class can contain "trees" and "bushes", urban area can contain "roofs" and "streets"). So, we could consider that the real class ω_1 is an union of two subclasses $v_1=(\omega_1,\lambda_1)$, $v_2=(\omega_1,\lambda_2)$ .and the real class ω_2 is an union of two subclasses $v_3=(\omega_2,\lambda_1)$, $v_4=(\omega_2,\lambda_2)$.

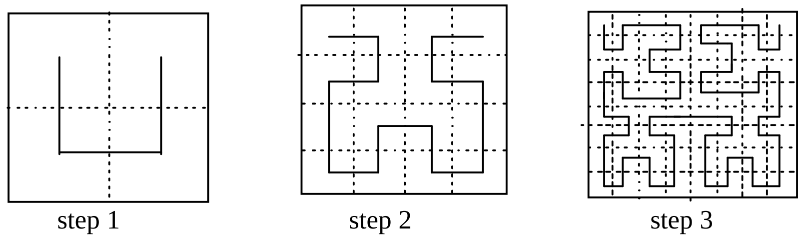


Fig. 1 Construction of the Hilbert-Peano scan

Case 1				
	v_1	v_2	v_3	v_4
v_1	0.201	0.004	0.001	0.003
v_2	0.005	0.459	0.002	0.003
v_3	0.001	0.002	0.140	0.001
c	0.003	0.003	0.001	0.169

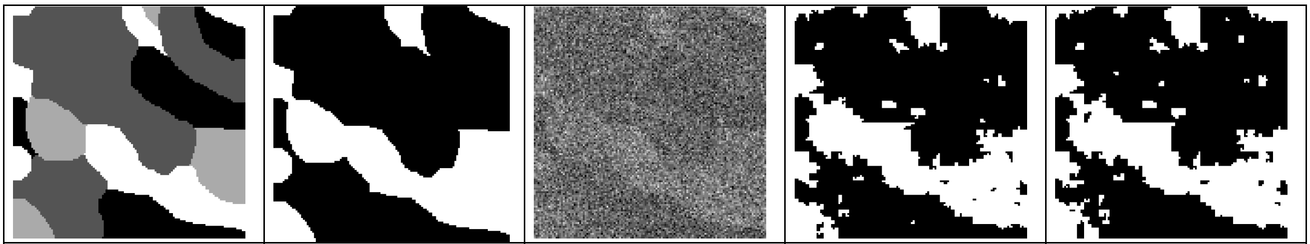
Tab. 4 : Estimated $p(u_1, x_1, u_2, x_2)$

Case 2				
	v_1	v_2	v_3	v_4
v_1	0.243	0.005	0.001	0.007
v_2	0.005	0.190	0.000	0.003
v_3	0.001	0.000	0.186	0.006
v_4	0.007	0.003	0.006	0.338

Tab. 5 : Estimated $p(u_1, x_1, u_2, x_2)$

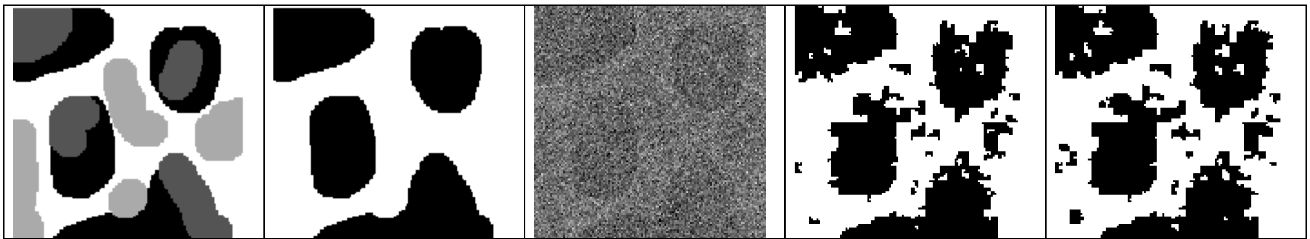
Case 1			
$p(y_1 \| (u_1, x_1) = (\omega_1, \lambda_1))$	$p(y_1 \| (u_1, x_1) = (\omega_1, \lambda_2))$	$p(y_1 \| (u_1, x_1) = (\omega_2, \lambda_1))$	$p(y_1 \| (u_1, x_1) = (\omega_2, \lambda_2))$
Mean=0, Var=1	Mean=0.25, Var=1	Mean=0.75, Var=1	Mean=1, Var=1
Case 2			
$p(y_1 \| (u_1, x_1) = (\omega_1, \lambda_1))$	$p(y_1 \| (u_1, x_1) = (\omega_1, \lambda_2))$	$p(y_1 \| (u_1, x_1) = (\omega_2, \lambda_1))$	$p(y_1 \| (u_1, x_1) = (\omega_2, \lambda_2))$
Mean=0, Var=1	Mean=0.25, Var=1	Mean=0.5, Var=1	Mean=0.75, Var=1

Tab. 6 : Parameters of the Gaussian densities $p(y_1 | (u_1, x_1)$ in two TMC studied.



$(X, U) = (x, u)$	Real classes image	Noisy image	TMC MPM:7.26%	HMC MPM: 8.23%

Fig. 2 : Case 1, Supervised segmentation using TMC based MPM and HMC based one. Parameters specified in Tab. 4, and Tab. 6.



$(X, U) = (x, u)$	Real classes image	Noisy image	TMC MPM:11.87%	HMC MPM: 13.06%

Fig. 3 : Case 2, Supervised segmentation using TMC based MPM and HMC based one. Parameters specified in Tab. 5 and Tab. 6.

Remark 3.2
We have tried to use the EM to perform unsupervised segmentation from the noisy images above but the results are not good. This is undoubtedly due to the high level of the noise; in fact, the introduction of the auxiliary classes has as

effect to increase the noise level. Further studies are needed to determinate in what kind of situations the TMC and EM based MPM unsupervised image segmentation is of interest with respect to HMC based one.

4. CONCLUSIONS

The aim of this paper was to propose some further developments and some first experiments related to a recent Triplet Markov Chain (TMC) model [25], which generalize the Pairwise Markov Chain model [10, 11], the latter generalizing the classical Hidden Markov Chain (HMC) model. We showed, via simulation study, that when the data suit a TMC the Bayesian TMC based restoration is more efficient than the HMC based one. Furthermore, the same is true in an unsupervised restoration, where parameters are estimated by a variant of the general Expectation-Maximization (EM) method. Likely to HMC and PMC, TMC can be used in image segmentation. We studied some simple examples and TMC Bayesian MPM segmentation method still works better s than the HMC based one; however, the differences are less striking and further studies are needed to understand in which situations TMC are to be used instead of HMC or PMC.

Let us mention some possible directions of further investigations.

(i) The two TMC used in our simulations are relatively simple; in particular, the possible correlation of the variables Y_1, …, Y_n conditionally on (X,U) has not been considered. More complex cases, with correlated and possibly non Gaussian noise could be considered. We have used here the EM method, well suited to the Gaussian case. When the noise is not Gaussian, the Iterative Conditional Estimation (ICE) method [23], which is more flexible than the EM method [8, 20] and which has been successfully used in Pairwise Markov Chains (PMC [10, 11]), which are a particular case of TMC, could be used;

(ii) the links with the Dempster-Shafer theory of evidence, discussed here in a simple mono sensor case could be complicated by considering the Dempster-Shafer fusion of numerous, possibly "evidential", sensors with possibly "evidential" priors, in the context of Markov modeling;

(iii) in an analogous way Triplet Markov Fields (TMF) model is proposed in [28] and first experiments show its interest in image segmentation. So, the generalization of some further Markov models like Markov trees [16, 24], or more complex grahical models [34], to Triplet Markov Models could possibly be of interest.

Références

1. K. Aas, L. Eikvil, and R. B. Huseby, Applications of hidden Markov chains in image analysis, *Pattern recognition*, Vol. 32, No. 4, pp. 703-713, 1999.
2. A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, *Revue Scientifique et Technique de la Défense*, No. 11, pp. 27-40, 1991.
3. L. E. Baum, T. Petrie, G. Soules, N. Weiss. A maximization technique occuring in the statstical analysis of probabilistic functions of Markov chains, *Ann, Math. Statistic*., 41, pp. 164-171, 1970.
4. A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski, Multisensor images segmentation using Dempster-Shafer fusion in Markov fields context, *IEEE Trans. on GRS*, Vol. 39, No. 8, pp. 1789-1798, 2001.
5. B. Benmiloud et 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.
6. I. Bloch, Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account, *Pattern Recognition Letters*, No. 17, pp. 905-919, 1996.
7. M.-Y. Chen, A. Kundu, and J. Zhou, Off-line handwritten word recognition using a hidden Markov model type stochastic network, *IEEE Trans. on PAMI*, Vol. 16, No. 5, pp. 481-496, 1994.
8. J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, *IEEE Transactions on SP*, Vol. 45, No. 10, pp. 2613-2615, 1997.
9. A.P. Dempster, N.M. Laird, and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *JRSS*, Series B, 39, pp. 1-38, 1977.
10. S. Derrode and W. Pieczynski, SAR image segmentation using using generalized Pairwise Markov Chains, SPIE's International Symposium on Remote Sensing, September 22-27, Crete, Greece, 2002.
11. S. Derrode and W. Pieczynski, Unsupervised Signal and Image Segmentation Using Pairwise Markov Chains, submitted to *IEEE Trans. on SP*.
12. S. Foucher, M. Germain, J.-M. Boucher, and G. B. Benié, Multisource classification using ICM and Dempster – Shafer theory, *IEEE Trans. on IM*, Vol. 51, No. 2, pp. 277-281, 2002.

13. L. Fouque, A. Appriou, and W. Pieczynski, An Evidential Markovian Model for Data Fusion and Unsupervised Image Classification, Proceedings of 3rd International Conference on Information Fusion, FUSION 2000, Vol. 1, July 10th-13th, 2000, Paris, France, pp. TuB4-25 - TuB4-31.
14. 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.
15. J. Guan and D.A. Bell, Evidence Theory and Its Applications, North-Holland, Amsterdam, 1991.
16. J.-M. Laferté, P. Pérez, and F. Heitz, Discrete Markov image modeling and inference on the quadtree, *IEEE Transactions on IP*, Vol. 9 , No 3, pp. 390 - 404, 2000.
17. S. Lakshmanan, and H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields, *IEEE Trans. on Trans. on PAMI*, Vol. 11, pp. 799-813, 1989.
18. McLachlan, EM Algorithm and Extensions, Wiley, Series in Probability and Statistics, 1997.
19. J. Marroquin, S. Mitter, T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *Journal of the American Statistical Association*, 82, pp. 76-89, 1987.
20. P. Pérez. Markov random fields and images. *CWI Quarterly*, 11(4):413-437, 1998.
21. W. Pieczynski, Hidden evidential Markov trees and image segmentation, Proceedings of the IEEE International Conference on Image Processing (ICIP' 99), 24-28 October 1999, Kobe, Japon.
22. W. Pieczynski, Unsupervised Dempster-Shafer fusion of dependent sensors, Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI' 2000), 2-4 April 2000, Austin, Texas, États-Unis, pp. 247-251.
23. W. Pieczynski, Statistical image segmentation, *Machine Graphics and Vision*, Vol. 1, No. 1/2, pp. 261-268, 1992.
24. W. Pieczynski, Arbres de Markov Couple; Pairwise Markov Trees, *C. R. Acad. Sci.* Paris, Ser. I 335, pp. 79-82, 2002.
25. W. Pieczynski, Chaînes de Markov Triplet; Triplet Markov Chains, *C. R. Acad. Sci.* Paris, Ser. I 335, pp. 275-278, 2002.
26. W. Pieczynski and A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, *Machine Graphics & Vision*, Vol. 9, No. 3, pp. 705-718, 2000.
27. W. Pieczynski, Pairwise Markov Chains, *IEEE Trans. on PAMI*, to appear.
28. W. Pieczynski, D. Benboudjema, and P. Lanchantin, Statistical image segmentation using Triplet Markov Fields, SPIE’s International Symposium on Remote Sensing, September 22-27, Crete, Greece, 2002.
29. G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, 1976.
30. P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, *International Journal of Approximate Reasoning*, Vol. 9, pp. 1-35, 1993.
31. P. Smets, Belief functions versus probability functions, In B. Bouchon, L. Saitta, and R. R. Yager (eds), Uncertainty and intelligent systems, Berlin, Springer-Verlag, pp. 17-24, 1988.
32. A.-N. Tebbache, J.-M. Boucher, and W. Pieczynski, Unsupervised textured image segmentation using Pairwise Markov random fields, submitted to *IEEE Trans. on IP*.
33. F. Tupin, H. Maitre, and I. Bloch, A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors, *IEEE Trans. on GRS*, Vol. 37, No. 3, pp. 1327-1343, 1999.
34. J. Whittaker, Graphical Models in Applied Multivariate Statistics, Wiley, Series in Probability and Mathematical Statistics, 1996
35. R. R. Yager, J. Kacprzyk, and M. Fedrizzi, Advances in the Dempster-Shafer theory of evidence, Wiley, 1994.