Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
45 (2007) 1–16
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Multisensor triplet Markov chains and theory of evidence

Wojciech Pieczynski *

GET/INT, Département CITI, CNRS UMR 5157 9, Rue Charles Fourier, 91000 Evry, France

Received 18 September 2003; received in revised form 9 May 2006; accepted 10 May 2006
Available online 13 June 2006

Abstract

Hidden Markov chains (HMC) are widely applied in various problems occurring in different areas like Biosciences, Climatology, Communications, Ecology, Econometrics and Finances, Image or Signal processing. In such models, the hidden process of interest X is a Markov chain, which must be estimated from an observable Y, interpretable as being a noisy version of X. The success of HMC is mainly due to the fact that the conditional probability distribution of the hidden process with respect to the observed process remains Markov, which makes possible different processing strategies such as Bayesian restoration. HMC have been recently generalized to "Pairwise" Markov chains (PMC) and "Triplet" Markov chains (TMC), which offer similar processing advantages and superior modeling capabilities. In PMC, one directly assumes the Markovianity of the pair (X, Y) and in TMC, the distribution of the pair (X, Y) is the marginal distribution of a Markov process (X, U, Y), where U is an auxiliary process, possibly contrived. Otherwise, the Dempster–Shafer fusion can offer interesting extensions of the calculation of the "a posteriori" distribution of the hidden data.

The aim of this paper is to present different possibilities of using the Dempster–Shafer fusion in the context of different multisensor Markov models. We show that the posterior distribution remains calculable in different general situations and present some examples of their applications in remote sensing area.

Keywords: Hidden Markov chains; Pairwise Markov chains; Triplet Markov chains; Multisensor data classification; Bayesian classification; Theory of evidence; Dempster–Shafer fusion

* Tel.: +33 1 60 76 44 25; fax: +33 1 60 76 44 33.
E-mail address: Wojciech.Pieczynski@int-evry.fr

0888-613X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2006.05.001

1. Introduction

Hidden Markov chains (HMC) are widely used in various problems comprising two stochastic processes $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$, in which $X = x$ is unobservable and must be estimated from the observed $Y = y$. The qualifier "hidden Markov" means that the hidden process X has a Markov distribution. When the distributions $p(y|x)$ of Y conditional on $X = x$ are simple enough, the pair (X, Y) retains the Markovian structure, and likewise 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 $Y = y$ even in the case of a large n. There are countless papers dealing with numerous problems in various areas like Biosciences [19,27], Climatology [39], Communications [8], Ecology [23], Econometrics and Finance [40], Image or Signal processing [7–9,35]. Let us also mention [2,14] as pioneering papers. More recently, HMC have been extended to "pairwise" Markov chains (PMC [30]) and "triplet" Markov chains (TMC [31,32]) and different recent studies show that these extensions can be useful in practical applications [6,13,21,22].

As this paper also address readers little familiar with HMC, let us illustrate their interest by considering the following simple situation, which will be used as an example through the whole paper. The points $(1, \ldots, n)$ are pixels of a line of a digital image, and each X_i takes its values in $\Omega = \{\omega_1, \omega_2\}$, where ω_1 is "forest" and ω_2 is "water". Otherwise, each Y_i takes its values in R and thus $Y = (Y_1, \ldots, Y_n) = (y_1, \ldots, y_n)$ is the observed line of the observed digital image. We wish to estimate $X = x$ from $Y = y$ in such a way that the proportion of wrongly estimated x_i would be minimal (an estimator verifying this property will be called "optimal"). The simplest way is to consider the distribution $p(x_i|y_i)$, called "posterior" distribution. Having the proportions of forest $p(x_i = \omega_1)$ and water $p(x_i = \omega_2)$, and having the two likelihoods $p(y_i|x_i = \omega_1)$, $p(y_i|x_i = \omega_2)$, the posterior distribution $p(x_i|y_i)$. is computed by the Bayes rule $p(x_i|y_i) = \frac{p(x_i)p(y_i|x_i)}{p(x_i=\omega_1)p(y_i|x_i=\omega_1)+p(x_i=\omega_2)p(y_i|x_i=\omega_2)}$. The optimal estimator $\hat{x} = \hat{s}(y)$ is then given by $\hat{x}_i = \hat{s}(y_i) = \arg\max_{\omega\in\Omega} p(x_i = \omega|y_i)$, for each $i = 1, \ldots, n$. This $\hat{s}$ minimizes the error probability $ER = P[(X_i = \omega_1, \hat{s}(Y_i) = \omega_2) \cup (X_i = \omega_2, \hat{s}(Y_i) = \omega_1)]$. In other words, when x_i is searched from y_i in some way, the error probability (and thus, via the large numbers law and for n large enough, the proportion of wrongly estimated pixels) must be superior or equal to ER. We can notice that this optimal method is relatively an intuitive one which consists of taking for $\hat{x}_i$ the class whose probability conditional to the observation y_i is maximal. How to diminish ER? When the random variables $Y_1, \ldots, Y_n$ are dependent, one can search x_i from (y_i, y_{i+1}) by computing $p(x_i|y_i, y_{i+1})$, or from (y_{i-1}, y_i, y_{i+1}) by computing $p(x_i|y_{i-1}, y_i, y_{i+1}), \ldots$, and so on. The drawback is that the computational problems arise when the number of y_j used to estimate x_i increases, and it is difficult to exceed about a dozen. Here we arrive at the interest of the hidden Markov models: they define a distribution $p(x, y) = p(x_1, \ldots, x_n, y_1, \ldots, y_n)$ of all the variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ in such a way that the distributions $p(x_i|y) = p(x_i|y_1, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_n)$ are computable (for every $i = 1, \ldots, n$), even for very large n (one million, or more...). The estimator $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n) = \hat{s}_{\mathrm{MPM}}(y)$, with $\hat{x}_i = \arg\max_{\omega\in\Omega} p(x_i = \omega|y)$, is then workable and can produce significantly better results than $\hat{x}_i = \hat{s}(y_i) = \arg\max p(x_i = \omega|y_i)$ above. For example, we can simultaneously have the optimal ER of about 45% and the error produced by $\hat{s}_{\mathrm{MPM}}$ of about 1%. The distributions $p(x_i|y)$ are called "posterior marginal distributions'', and $\hat{s}_{\mathrm{MPM}}$ is called ''Bayesian maximum of posterior marginals'', or ''Bayesian MPM''.

Otherwise, the theory of evidence [1,11,12,16–18,36–38,42] provides models which can be seen, in some situations, as extensions of probabilistic models and ensure some elegant formulations and solutions of different classification problems. For example, in image processing area we focus our examples on in this paper, this theory allows one to deal with medical images classification [5], radar and optical images fusion [15,24], change detection [25], or still different detectors fusion in SAR images [41]. To make a first link between the Bayesian classification and the theory of evidence, let us consider the posterior distribution $p(x_i|y_i)$ mentioned in the ''forest and water'' example above. It can be seen as the ''Dempster–Shafer'' fusion (DS fusion) of two distributions: the ''prior'' distribution $p(x_i)$, and the distribution $p^{y_i}(x_i)=\frac{p(y_i|x_i)}{p(y_i|x_i=\omega_1)+p(y_i|x_i=\omega_2)}$, defined by the observation y_i and the likelihoods $p(y_i|x_i)$ (note that $p^{y_i}(x_i)$ is not the posterior distribution $p(x_i|y_i)$). Such a DS fusion remains valid in more general context of the ''theory of evidence'', where the probabilities $p(x_i)$ and $p^{y_i}(x_i)$ are extended to ''belief functions''. Therefore, the result of the DS fusion of two belief functions, which is a probability distribution or not, can be seen as an extension of the posterior distribution $p(x_i|y_i)$, and used to estimate x_i from y_i. Such an extension is of interest in numerous situations as, for example, in the following example drawn from [22]. Let us consider the example of ''water'' and ''forest'' above. Let us assume that $p_i=p(x_i=\omega_1)$ depends on i, but the two distributions $p(y_i|x_i=\omega_1)$ and $p(y_i|x_i=\omega_2)$ do not depend on i. Moreover, imagine that for each $i=1,\ldots,n$ the probability $p_i=p(x_i=\omega_1)$ is sampled in $[0,1]$ with respect to some law whose expectation is 0.5. Imagine that we can not know the parameters $p_1,\ldots,p_n$, and thus we consider them as being equal and their common value is estimated in some way. So, we will use a false $s=P[X_i=\omega_1]$ instead of $p_1,\ldots,p_n$, which gives for each $i=1,\ldots,n$ an error probability $ER(s)$. According to the theory of evidence, we can replace the false $s=p(x_i=\omega_1)$, $1-s=p(x_i=\omega_2)$ by a ''weakened'' mass function $m(\{\omega_1\})=s-w$, $m(\{\omega_2\})=1-s-w$, $m(\{\omega_1,\omega_2\})=2w$. The DS fusion $m\oplus p^{y_i}$ is then a probability distribution on Ω and its use instead of $p(x_i|y_i)$ based on $s=p(x_i=\omega_1)$, $1-s=p(x_i=\omega_2)$ can diminish the classification error [22].

The aim of this paper is to answer the following question. Is it possible to simultaneously benefit from Markov models and theory of evidence? More precisely, we will focus on the following question: is it possible to extend the calculus of the posterior marginals $p(x_i|y)$, that is classical in the hidden Markov chains context, to a theory of evidence context? We provide different results specifying how the DS fusion can be performed in multisensor HMC and PMC.

Let us notice that similar problems have been recently considered in the context of pairwise [28] and triplet [3] Markov fields and different results, somewhat similar to the results of the present paper, are described in [33]. However, although the ideas are similar, the practical solutions are very different in Markov fields context considered in [33], and in Markov chains context considered here. In fact, $p(x_i|y)$ have to be estimated by some Monte Carlo Markov chains (MCMC) method in the Markov field context, while they can be, as specified in the present paper, explicitly computed in the Markov chains one.

The organization of the paper is the following. Classical HMC, PMC, and TMC are recalled in the next section, with a slight novelty concerning the characterization of stationary HMC. Basic notions of the theory of evidence are specified in Section 3. Section 4 is devoted to the case of ''evidential priors'' and contains an extension to PMC of the results presented in the case of HMC in [22]. The "evidential observations" and the general case are dealt with in Section 5, while Section 6 contains concluding remarks and some perspectives.

2. Triplet Markov chains (TMC)

2.1. Pairwise Markov chains

Let us briefly present the Pairwise Markov chains (PMC) model introduced in [30]. Let $X=(X_1,\dots,X_n)$ and $Y=(Y_1,\dots,Y_n)$ be two stochastic processes, where each X_i takes its values in a finite set $\Omega=\{\omega_1,\dots,\omega_k\}$ and each Y_i takes its values in R. Let $Z=(Z_1,\dots,Z_n)$ be the "pairwise" process, with $Z_i=(X_i,Y_i)$. The processes X and Y are said to be a PMC when Z is a Markov process, which means that its distribution is written

$$p(z)=p(z_1)p(z_2|z_1)\cdots p(z_n|z_{n-1}) \tag{2.1}$$

The classical "hidden Markov chain", which will be denoted by CHMC in the whole paper, then appears as the following particular case of (2.1):

$$p(z)=p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1})p(y_1|x_1)\cdots p(y_n|x_n) \tag{2.2}$$

The greater generality of PMC over CHMC can be seen at a "local" and at a "global" level. Recalling that $z_i=(x_i,y_i)$, the general form of the transitions in (2.1) is $p(z_{i+1}|z_i)=p(x_{i+1}|x_i,y_i)p(y_{i+1}|x_{i+1},x_i,y_i)$, and in (2.2) this form is $p(z_{i+1}|z_i)=p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1})$. Therefore the CHMC is a PMC in which $p(x_{i+1}|x_i,y_i)=p(x_{i+1}|x_i)$ and $p(y_{i+1}|x_{i+1},x_i,y_i)=p(y_{i+1}|x_{i+1})$, which shows the greater generality of PMC at the "local" level. Concerning the "global" level we remark that $p(y|x)$ is a Markov chain in PMC, while it is given by the very simple formula $p(y|x)=p(y_1|x_1)\cdots p(y_n|x_n)$ in CHMC. Other differences between stationary CHMC and stationary PMC are specified in Proposition 2.1 below.

The large success of CHMC is due to the fact that $p(x|y)$ is a Markov chain, with computable transitions $p(x_{i+1}|x_i,y)$. This remains true in PMC; in fact, the following "forward" $\alpha(x_i)=p(y_1,\dots,y_{i-1},z_i)$ and "backward" $\beta(x_i)=p(y_{i+1},\dots,y_n|z_i)$ probabilities (which give the classical ones when PMC is a CHMC) are calculable recursively with the formulas (2.3) and (2.4). This makes possible the calculation of the transitions $p(x_{i+1}|x_i,y)$ and the marginals $p(x_i|y)$ associated with the distribution of X conditional on $Y=y$ via formulas (2.5) and (2.6).

$$\alpha(x_1)=p(z_1),\ \text{and}\ \alpha(x_{i+1})=\sum_{x_i\in\Omega}\alpha(x_i)p(z_{i+1}|z_i)\quad \text{for } 2\leqslant i\leqslant n \tag{2.3}$$

$$\beta(x_n)=1,\ \text{and}\ \beta(x_i)=\sum_{x_{i+1}\in\Omega}\beta(x_{i+1})p(z_{i+1}|z_i)\quad \text{for } 1\leqslant i\leqslant n-1 \tag{2.4}$$

$$p(x_{i+1}|x_i,y)=\frac{p(z_{i+1}|z_i)\beta(x_{i+1})}{\beta(x_i)} \tag{2.5}$$

$$p(x_i|y)=\frac{\alpha(x_i)\beta(x_i)}{\sum_{x_i'\in\Omega}\alpha(x_i')\beta(x_i')} \tag{2.6}$$

The formulas (2.3)–(2.6) are generalizations of the well known CHMC formulas, which are obtained by taking $p(z_1)=p(x_1)p(y_1|x_1)$ and $p(z_{i+1}|z_i)=p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1})$.

So, roughly speaking, PMC are more general than CHMC, but the interesting properties that are at the origin of the CHMC's success are retained. Let us notice that this theoretical greater generality of PMC over CHMC can imply significant practical superiority of PMC based unsupervised segmentation methods over the CHMC-based ones [13].

Let us call "hidden Markov chain" (HMC) every PMC $Z = (X, Y)$ in which X is a Markov chain (thus CHMC are HMC in which, in addition, $p(y|x) = p(y_1|x_1) \cdots p(y_n|x_n)$). The greater generality of stationary PMC over HMC is stated in the following proposition (the equivalence of (i) and (ii) is showed in [30], while the condition (iii) is new).

Proposition 2.1. *Let $Z = (X, Y)$ be a PMC verifying*

(a) $p(z_i, z_{i+1})$ *does not depend on* $1 \leqslant i \leqslant n-1$;
(b) $p(z_i = a, z_{i+1} = b) = p(z_i = b, z_{i+1} = a)$ *for each* $1 \leqslant i \leqslant n-1$, *a, and b.*

Then the three following conditions

(i) *X is a Markov chain (i.e., $Z = (X, Y)$ is a HMC);*
(ii) *for each* $2 \leqslant i \leqslant n$, $p(y_i|x_i, x_{i-1}) = p(y_i|x_i)$;
(iii) *for each* $1 \leqslant i \leqslant n$, $p(y_i|x) = p(y_i|x_i)$

are equivalent.

Proof. The equivalence of (i) and (ii) is showed in [30]; let us show the equivalence between (ii) and (iii).

(ii) implies (iii). The distribution of $Z = (X, Y)$ can be written

$$
\begin{aligned}
p(z) &= \frac{p(z_1, z_2) \cdots p(z_{n-1}, z_n)}{p(z_1) \cdots p(z_{n-1})} \\
&= \underbrace{\left[\frac{p(x_1, x_2) \cdots p(x_{n-1}, x_n)}{p(x_2) \cdots p(x_{n-1})}\right]}_{a(x)} \underbrace{\left[\frac{p(y_1, y_2|x_1, x_2) \cdots p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(y_2|x_2) \cdots p(y_{n-1}|x_{n-1})}\right]}_{b(x,y)}
\end{aligned} \tag{2.7}
$$

As $p(y_j|x_{j-1}, x_j) = p(y_j|x_j, x_{j+1}) = p(y_j|x_j)$, the integration of $b(x, y)$ with respect to $y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n$ gives $p(y_i|x_i)$. Thus $p(y_i|x) = \frac{p(x,y_i)}{a(x)} = \frac{a(x)p(y_i|x_i)}{a(x)} = p(y_i|x_i)$. Conversely, (iii) implies that $p(x, y_i) = p(y_i|x_i)p(x)$ and so, integrating the latter with respect to $x_1, \ldots, x_{i-2}, x_{i+1}, \ldots, x_n$, we obtain $p(x_{i-1}, x_i, y_i) = p(y_i|x_i)p(x_{i-1}, x_i)$. As $p(x_{i-1}, x_i, y_i) = p(y_i|x_{i-1}, x_i)p(x_{i-1}, x_i)$, we have $p(y_i|x_i, x_{i-1}) = p(y_i|x_i)$, which ends the proof. □

Let us briefly comment the interest of Proposition 2.1 by remarking the following points:

(i) The equality (2.7) is always true, in any PMC. Thus when the PMC $Z = (X, Y)$ is a HMC, we have $a(x) = p(x)$ and $b(x, y) = p(y|x)$. The interesting point is to remark that when the PMC $Z = (X, Y)$ is not a HMC, then $a(x) \neq p(x)$ and $b(x, y) \neq p(y|x)$, but every $p(x_i, x_{i+1})$ in $a(x)$ is the distribution of (X_i, X_{i+1}) and every $p(y_i, y_{i+1}|x_i, x_{i+1})$ in $b(x, y)$ is the distribution of (Y_i, Y_{i+1}) conditional on $(X_i, X_{i+1}) = (x_i, x_{i+1})$ (similarly, every $p(x_i)$ in $a(x)$ is the distribution of X_i and every $p(y_i|x_i)$ in $b(x, y)$ is the distribution of Y_i conditional on $X_i = x_i$).

(ii) The CHMC have been extended in different directions in many papers; however, at our knowledge, except the PMC proposed in [30] (also see [32] for continuous hidden chain) all the extensions are HMC. Therefore, accordingly to Proposition 2.1, none of these extensions can take into account the fact that the aspect of the noise $p(y_i|x_i, x_{i-1})$ can depend on both x_i and x_{i-1} (at least, under hypotheses (a) and (b)). Considering again the example of "water" and "forest" given in Introduction, we can imagine that for $x_i =$ "forest", its visual aspect of the observed y_i, whose variability is modelled by the distribution $p(y_i|x_i, x_{i-1})$, also depends on x_{i-1}. In fact, trees beside "water" ($x_{i-1} =$ "water") can have different aspect from the trees elsewhere. More generally, we see that PMC can model the fact that the "noise" can be different on frontiers, while HMC cannot.

(iii) It is possible to show that in PMC $p(y_i|x)$ can depend on all $x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n$, which means that the noise at a given point can be modelled in a very complete manner.

2.2. Triplet Markov chains

Let $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$, and $Z = (Z_1, \ldots, Z_n)$, with $Z_i = (X_i, Y_i)$, be stochastic processes as above, with X the hidden process and Y the observed one. The problem remains the same: estimate X from Y. Considering a Triplet Markov Chain (TMC) consists of introducing a third process $U = (U_1, \ldots, U_n)$ such that the triplet $T = (X, U, Y)$ is a Markov chain. Assuming that the variables X_i, U_i, and Y_i take their values in $\Omega = \{\omega_1, \ldots, \omega_k\}$, $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$, and R, respectively, let us thus assume that $T = (X, U, Y)$ is a Markov chain. Putting $V_i = (X_i, U_i)$, $V = (V_1, \ldots, V_n)$, we can say that (V, Y) is a PMC, and thus all results of the previous subsection remain valid, with $v_i = (x_i, u_i)$ instead of x_i. In particular, the distributions $p(v_{i+1}|v_i, y)$ and $p(v_i|y)$ are calculable by formulas (2.3)–(2.6), with variables linked with x replaced by variables linked with v, and variables linked with z replaced by variables linked with t. This means that $p(x_i|y) = \sum_{u_i \in \Lambda} p(x_i, u_i|y) = \sum_{u_i \in \Lambda} p(v_i|y)$ is calculable, which makes possible the use of the Bayesian MPM restoration method. We also immediately see than TMC are more general than PMC, as the latter are obtained by taking $\Lambda = \Omega$ and $X = U$.

The following lemma will be very useful in Sections 4 and 5.

Lemma 2.1. *Let $V = (V_1, \ldots, V_n)$ be a random chain, each V_i taking its values in the same finite set* V. *Then V is a Markov chain if and only if there exist $n - 1$ positive functions $q_1, \ldots, q_{n-1}$ such that the law of V is proportional to the product $q_1(v_1, v_2) \times \cdots \times q_{n-1}(v_{n-1}, v_n)$:*

$$p(v) \propto q_1(v_1, v_2) \times \cdots \times q_{n-1}(v_{n-1}, v_n) \tag{2.8}$$

If (2.8) is verified, $p(v_1)$ and the transitions $p(v_i|v_{i-1})$ of the Markov chain V are given by

$$p(v_1) = \frac{\beta_1(v_1)}{\sum_{v'_1} \beta_1(v'_1)}, \quad p(v_i|v_{i-1}) = \frac{q_{i-1}(v_{i-1}, v_i)\beta_i(v_i)}{\beta_i(v_{i-1})} \quad \text{for } 2 \leqslant i \leqslant n-1 \tag{2.9}$$

where $\beta_1(v_1), \ldots, \beta_n(v_n)$ are calculated from $q_1, \ldots, q_{n-1}$ by the recursive formulas

$$\beta_n(v_n) = 1, \text{ and } \beta(v_{i-1}) = \sum_{v_i} \beta_i(v_i) q_{i-1}(v_{i-1}, v_i) \quad \text{for } 2 \leqslant i \leqslant n-1 \tag{2.10}$$

Having $p(v_1)$ and the transitions $p(v_i|v_{i-1})$, the margins $p(v_i)$ are classically calculated by the recursive formulas

$$p(v_1) \text{ given, and } p(v_i) = \sum_{v_{i-1}} p(v_{i-1})p(v_i|v_{i-1}) \quad \text{for } 2 \leqslant i \leqslant n \tag{2.11}$$

Proof. If V is a Markov chain, we can take $q_1(v_1, v_2) = p(v_1, v_2)$, $q_1(v_2, v_3) = p(v_3|v_2), \ldots, q_{n-1}(v_{n-1}, v_n) = p(v_n|v_{n-1})$. Then (2.8) is verified with $p(v) = q_1(v_1,v_2) \times \cdots \times q_{n-1}(v_{n-1}, v_n)$, and both (2.9) and (2.10) are verified with all $\beta_i(v_i)$ are equal to 1.

Conversely, let us assume that (2.8) is verified. Let $p(v) = q_1(v_1, v_2) \times \cdots \times q_{n-1}(v_{n-1}, v_n)/S$, where $S = \sum_{v_1,\ldots,v_n} q_1(v_1, v_2) \times \cdots \times q_{n-1}(v_{n-1}, v_n)$. V is a Markov chain if for every $3 \leqslant i \leqslant n$ $p(v_i|v_1, \ldots, v_{i-1})$ does not depend on $(v_1, \ldots, v_{i-2})$. We have

$$\begin{aligned} p(v_i|v_1, \ldots, v_{i-1}) &= \frac{p(v_1, \ldots, v_{i-1}, v_i)}{p(v_1, \ldots, v_{i-1})} \\ &= \frac{\sum_{v_{i+1},\ldots,v_n} q_1(v_1, v_2) \times \cdots \times q_{n-1}(v_{n-1}, v_n)/S}{\sum_{v_i,\ldots,v_n} q_1(v_1, v_2) \times \cdots \times q_{n-1}(v_{n-1}, v_n)/S} \\ &= \frac{q_{i-1}(v_{i-1}, v_i)\sum_{v_{i+1},\ldots,v_n} q_i(v_i, v_{i+1}) \times \cdots \times q_{n-1}(v_{n-1}, v_n)}{\sum_{v_i,\ldots,v_n} q_{i-1}(v_{i-1}, v_i) q_i(v_i, v_{i+1}) \times \cdots \times q_{n-1}(v_{n-1}, v_n)} \end{aligned} \tag{2.12}$$

On the one hand, we see that $p(v_i|v_1, \ldots, v_{i-1})$ only depends on v_{i-1} and thus (2.12) gives $p(v_i|v_{i-1})$. On the other hand, we can calculate the sum $\sum_{v_{i+1},\ldots,v_n} q_i(v_i, v_{i+1}) \times \cdots \times q_{n-1}(v_{n-1}, v_n)$ beginning with v_n, continuing with $v_{n-1}, \ldots$ and so on until v_{i+1}. Doing so we see, according to (2.10), that $\sum_{v_{i+1},\ldots,v_n} q_i(v_i, v_{i+1}) \times \cdots \times q_{n-1}(v_{n-1}, v_n) = \beta_i(v_i)$. Doing the same in the denominator of (2.12) until v_i, we obtain (2.9), which ends the proof. □

We will be interested on $p(x_i|y)$, and thus, according to this lemma, all we have to do is to show that $p(v|y)$ verifies (2.8), where y is a constant. $p(v_i|y)$ is then calculable with (2.9)–(2.11), and $p(x_i|y)$ is given by $p(x_i|y) = \sum_{u_i \in \Lambda} p(x_i, u_i|y) = \sum_{u_i \in \Lambda} p(v_i|y)$.

Remark 2.1. Let us denote by r the number of elements in the finite set V. According to (2.10), calculating all $\beta_1(v_1), \ldots, \beta_n(v_n)$ needs $r(n-1)$ additions and $r(n-1)$ multiplications. Then, according to (2.9), making $r(n-1)$ divisions gives $p(v_1)$ all the a posteriori transitions $p(v_i|v_{i-1})$ for $2 \leqslant i \leqslant n-1$. Otherwise, according to (2.11), $r(n-1)$ further additions and $r(n-1)$ additional multiplications give the margins $p(v_i)$. Finally, all $p(v_i)$ are computed after $5r(n-1)$ elementary operations and important is that their number is linear in n. When $V_i = (X_i, U_i)$ takes its values in $V \subset \Omega \times \Lambda$, with $\Omega = \{\omega_1, \ldots, \omega_k\}$ and $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$, we have $r \leqslant km$ and thus the computation of all the margins $p(x_i|y)$ needs no more than nm further additions. As a result, all margins $p(x_i|y)$ are calculated by less than $5km(n-1) + nm = (5k+1)mn - 5km$ elementary operations. Important is that this total number increases proportionally to n and thus remains generally workable for very large n.

3. Theory of evidence

Let us consider $\Omega = \{\omega_1, \ldots, \omega_k\}$, and its power set $\mathrm{P}(\Omega) = \{A_1, \ldots, A_q\}$, with $q = 2^n$. A function M from $\mathrm{P}(\Omega)$ to $[0, 1]$ is called a "basic belief assignment" (bba) if $M(\emptyset) = 0$ and

$\sum_{A\in P(\Omega)} M(A) = 1$. A bba M defines then a "plausibility" function Pl from P(Ω) to $[0,1]$ by $Pl(A) = \sum_{A\cap B\neq\emptyset} M(B)$, and a "credibility" function Cr from P(Ω) to $[0,1]$ by $Cr(A) = \sum_{B\subset A} M(B)$. For a given bba M the corresponding plausibility function Pl and credibility function Cr are linked by $Pl(A) + Cr(A^c) = 1$, so that each of them defines the other. Conversely, Pl and Cr can be defined by some axioms, and each of them defines then an unique corresponding bba M. More precisely, Cr is a function from P(Ω) to $[0,1]$ verifying $Cr(\emptyset) = 0$, $Cr(\Omega) = 1$, and $Cr(\bigcup_{j\in J} A_j) \geqslant \sum_{I\subset J, I\neq\emptyset} (-1)^{|I|+1} Cr(\bigcap_{j\in I} A_j)$, and Pl is a function from P(Ω) to $[0,1]$ verifying analogous conditions, with $\leqslant$ instead of $\geqslant$ in the third one. A credibility function Cr verifying such conditions also is the credibility function defined by the bba $M(A) = \sum_{B\subset A} (-1)^{|A-B|} Cr(B)$.

Finally, each of the three functions M, Pl, and Cr can be defined in an axiomatic way, and each of them defines the two others. Furthermore, a probability function p can be seen as a particular case in which $Pl = Cr = p$.

When two bbas M_1, M_2 represent two pieces of evidence, we can combine – or fuse – them using the so called "Dempster–Shafer fusion" (DS fusion), which gives $M = M_1 \oplus M_2$ defined by:

$$M(A) = (M_1 \oplus M_2)(A) \propto \sum_{B_1\cap B_2 = A} M_1(B_1) M_2(B_2) \tag{3.1}$$

We will say that a bba M is "Bayesian" when, being null outside singletons, it defines a probability and we will say that it is an "evidential" bba when it is not a Bayesian one. One can then see that when either M_1 or M_2 is Bayesian, then the fusion result M is Bayesian. In fact, for M_1 Bayesian $M_1(B_1)$ is null outside singletons, and thus $M_1(B_1)M_2(B_2)$ in (3.1) also is null outside singletons. This means that $M(A)$ in (3.1) is null outside singletons because if A is not a singleton, all $M_1(B_1)$ such that $A = B_1 \cap B_2$ (for some B_2) are null. Otherwise, as mentioned in Introduction, one may see that the calculus of the posterior probability is a DS fusion of two Bayesian bbas. Note that the different extensions of the present paper are based on this very important point.

Example 3.1. Let $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$ be a CHMC, with $p(x, y)$ given by (2.2). The posterior distribution $p(x|y)$ of X can then be seen as a normalized product of the probability $p_1(x) = p(x_1)p(x_2|x_1) \cdots p(x_n|x_{n-1})$ and the probability $p_2^y(x) = \frac{p(y_1|x_1)\ldots p(y_n|x_n)}{\sum_{x'\in\Omega^n} p(y_1|x'_1)\ldots p(y_n|x'_n)}$. As the probabilities p_1 and p_2^y are equivalent to Bayesian bba M_1 and M_2^y (we have $p_1(x_1, \ldots, x_n) = M_1(\{x_1\}, \ldots, \{x_n\})$ and $p_2^y(x_1, \ldots, x_n) = M_2^y(\{x_1\}, \ldots, \{x_n\})$, we can consider that the Bayesian bba M^y equivalent to the probability $p(x|y)$ is the result of DS fusion of M_1 and $M_2^y : M^y = M_1 \oplus M_2^y$.

Example 3.2. Let us consider the problem of satellite or airborne optical image segmentation into two classes $\Omega = \{\omega_1, \omega_2\}$ "forest" and "water", as considered in Introduction. However, let us imagine that there are clouds. Thus, at pixel i, we have a random variable X_i taking its values in $\Omega = \{\omega_1, \omega_2\}$, whose distribution is $p_1(x_i)$. The observed $Y_i = y_i \in R$ possibly follows three distributions: $p(y_i|\omega_1)$, $p(y_i|\omega_2)$, and $p(y_i|c)$, with c for "clouds". Thus y_i defines on $\{\omega_1, \omega_2, c\}$ the probability $p^{y_i}(\omega_1) = p(y_i|\omega_1)/a$, $p^{y_i}(\omega_2) = p(y_i|\omega_2)/a$, $p^{y_i}(c) = p(y_i|c)/a$, with $a = p(y_i|\omega_1) + p(y_i|\omega_2) + p(y_i|c)$. Therefore, we have two probability distributions: p_1 on $\Omega = \{\omega_1, \omega_2\}$, and p^{y_i} on $\{\omega_1, \omega_2, c\}$, and the problem is to fuse them to obtain a probability on Ω which would be an extension of the classical posterior distribution, and which could be used to perform some Bayesian classification. One possible way of performing such a fusion, that we will consider in this paper, uses the theory of evidence in the following way. Consider the bba M_1 defined on $\{\{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}\}$ by $M_1(\{\omega_1\}) = p_1(\omega_1)$, $M_1(\{\omega_2\}) = p_1(\omega_2)$, $M_1(\{\omega_1, \omega_2\}) = 0$ (note that M_1 simply is the Bayesian bba associated with p_1), and the bba $M_2^{y_i}$ defined on $\{\{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}\}$ by $M_2^{y_i}(\{\omega_1\}) = p^{y_i}(\omega_1)$, $M_2^{y_i}(\{\omega_2\}) = p^{y_i}(\omega_2)$, $M_2^{y_i}(\{\omega_1, \omega_2\}) = p^{y_i}(c)$. The fused bba $M = M_1 \oplus M_2^{y_i}$ is a probability on Ω and can be seen as an extension of the classical posterior probability $p(x_i|y_i)$; in fact, when there are no clouds, $M_2^{y_i}$ is defined on $\{\{\omega_1\}, \{\omega_2\}\}$ and the bba $M^{y_i}(x_i) = [M_1 \oplus M_2^{y_i}](x_i)$ is equivalent to the classical posterior distribution $p(x_i|y_i)$. Moreover, it is possible to attach an intuitive meaning to $M_2^{y_i}(\Omega)$: as we are only interested on "water" and "forest", $M_2^{y_i}(\Omega)$ models the ignorance attached with the fact that one can not see through clouds. How to pass from one pixel i to a Markov model defined on a line $(1, \ldots, n)$? Let $p_1(x) = p(x_1)p(x_2|x_1) \cdots p(x_n|x_{n-1})$ be a Markov chain distribution and let M_1 be the Bayesian bba associated with p_1 as in Example 3.1 above. Let us consider $M_2^y(u_1, \ldots, u_n) = M_2^{y_1}(u_1) \times \cdots \times M_2^{y_n}(u_n)$, where each u_i is in $P(\Omega) = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \Omega\}$ and $M_2^{y_1}, \ldots, M_2^{y_n}$ are the bba $M_2^{y_i}$ defined above. We see that if there are no clouds, $M_2^y(u_1, \ldots, u_n)$ simply is $M_2^y(\{x_1\}, \ldots, \{x_n\})$ in Example 3.1 above. Therefore $M^y(x) = [M_1 \oplus M_2^y](x)$ generalizes the Markov distribution $p(x|y)$ (probability distribution and Bayesian bba being equivalent, we note $M^y(x_1, \ldots, x_n)$ instead of $M^y(\{x_1\}, \ldots, \{x_n\})$ in order to simplify). Moreover, $M^y = M_1 \oplus M_2^y$ is itself a Markov law. In fact, we have $[M_1 \oplus M_2^y](x) \propto \sum_{x \in u} M_1(x) M_2^y(u)$ on the one hand (the sum is taken over $u = (u_1, \ldots, u_n)$ such that for the fixed $x = (x_1, \ldots, x_n)$ one has $x_1 \in u_1, \ldots, x_n \in u_n$), and $\sum_{x \in u} M_1(x) M_2^y(u) = M_1(x) \prod_{i=1}^n \left[\sum_{x_i \in u_i} M_2^{y_i}(u_i)\right]$, on the other hand. Therefore the DS fusion does not destroy the Markovianity in the simple case considered here; however, as we will see in the following, it often does.

Let us remark that in Example 3.2 above, the Markov chain $p_1(x) = p(x_1)$ $p(x_2|x_1) \cdots p(x_n|x_{n-1})$ can be replaced with a Markov field, and such a model has been successfully applied in synthetic and real image segmentation in [4].

4. Evidential priors in PMC

Let us return to the situation described in Example 3.1 above. When wishing to model the prior information M_1 by a "Markov" evidential distribution, things become less direct than in Example 3.2 because the DS fusion, even with a simple probabilistic M_2^y defined by a observation $Y = y$, destroys the Markovianity. However, Bayesian processing can still be applied because, as we are going to see in the following, the margins $p(x_i|y)$ remain computable.

Let us remark that the main interest of the TMC models $T = (X, U, Y)$ lies in the fact that the distribution of $V = (X, U)$ conditional on $Y = y$ is a Markov chain distribution. In other words, $p(x, u|y)$ is a PMC distribution. Therefore, the main problem is to show that for fixed $Y = y$ we are faced with some PMC $p(x, u|y)$. This is the reason that we will directly concentrate on this problem in this section, and in the following one as well. Thus TMC "disappear" and only PMC are dealt with; however, the content of both sections is valid for every fixed $Y = y$, and thus the stated properties are true because $T = (X, U, Y)$ is a Markov chain. Moreover, having in mind that $T = (X, U, Y)$ is a

Markov chain is determining when dealing with the parameter estimation problem; for example, this allowed us to use the "Expectation-Maximization" (EM) method in [22].

In this Section we specify the possibility of extending the CHMC and PMC, with along the corresponding Bayesian restorations, to the evidential priors case. We begin by CHMC (Proposition 3.1), which have already been studied [22]. However, we present here a rapid proof based on Lemma 2.1, and the same lemma is then used to present original result concerning PMC.

Definition 4.1. A bba M defined on $P(\Omega^n)$ will be called "Evidential Markov chain" (EMC) if it is null outside $[P(\Omega)]^n$ and if it can be written

$$M(A_1, A_2, \ldots, A_n) = M(A_1)M(A_2|A_1), \ldots, M(A_n|A_{n-1}) \tag{4.1}$$

where for each $i = 2, \ldots, n$ and $A_i \in P(\Omega)$, $M(|A_i)$ is a bba on $P(\Omega)$. Let us remark that a Bayesian EMC (which is null outside $A = (A_1, A_2, \ldots, A_n)$ such that all $A_1, A_2, \ldots, A_n$ are singletons) is equivalent to the classical Markov chain, which is obtained according to $p(x) = p(x_1)p(x_2|x_1) \cdots p(x_n|x_{n-1}) = M(\{x_1\})M(\{x_2\}|\{x_1\}), \ldots, M(\{x_n\}|\{x_{n-1}\})$.

Proposition 4.1. *Let M_1 be an EMC on $[P(\Omega)]^n$, and M_2^y a probability on Ω^n defined from the observed process $Y = y \in R^n$ by $M_2^y(x_1, \ldots, x_n) \propto p(y_1|x_1) \cdots p(y_n|x_n)$. Then the time requested to compute marginal distributions $M^y(x_i)$ of the probability distribution $M^y = M_1 \oplus M_2^y$ is linear in the number of observations.*

Proof. Let $\Lambda = P(\Omega)$, and $v_i = (x_i, u_i)$, where $u_i \in \Lambda$. For fixed $y = (y_1, \ldots, y_n)$, let $q_1, \ldots, q_{n-1}$ be functions defined on $(\Omega \times \Lambda)^2$ by

$$\begin{aligned}
q_1(v_1, v_2) &= 1_{[x_1 \in u_1]} M_1(u_1) p(y_1|x_1) 1_{[x_2 \in u_2]} M_1(u_2|u_1) p(y_2|x_2), \\
q_2(v_2, v_3) &= 1_{[x_3 \in u_3]} M_1(u_3|u_2) p(y_3|x_3), \\
&\ldots \\
q_{n-1}(v_{n-1}, v_n) &= 1_{[x_n \in u_n]} M_1(u_n|u_{n-1}) p(y_n|x_n).
\end{aligned}$$

On the one hand, $q_1(v_1, v_2)q_2(v_2, v_3) \cdots q_{n-1}(v_{n-1}, v_n)$ defines a Markov chain $V = (V_1, \ldots, V_n)$ by virtue of Lemma 2.1. On the other hand, the DS fusion $M^y(x_1, \ldots, x_n) = [M_1 \oplus M_2^y](x_1, \ldots, x_n)$ can be seen as the calculation of the distribution $p(x)$, which is the marginal distribution obtained by summing $p(x, u)$ with respect to $u = (u_1, \ldots, u_n)$, of the Markov distribution $p(v) = p(x, u)$. Therefore the margins $M^y(x_i)$ are computable from $q_1, \ldots, q_{n-1}$ as specified in Lemma 2.1, and the requested time is linear in the number of observations. □

Proposition 4.2. *Let M_1 be an EMC on $[P(\Omega)]^n$, and M_2^y a probability on Ω^n defined from the observed process $Y = y \in R^n$ by $M_2^y(x_1, \ldots, x_n) \propto \frac{p(y_1, y_2|x_1, x_2) \cdots p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(y_2|x_2) \cdots p(y_{n-1}|x_{n-1})}$ (see (2.6)).*

Then the time requested to compute marginal distributions $M^y(x_i)$ of the probability distribution $M^y = M_1 \oplus M_2^y$ is linear in the number of observations.

The proof is similar to the proof of the Proposition 4.1, with

$$q_1(v_1, v_2) = 1_{[x_1 \in u_1]} M_1(u_1) 1_{[x_2 \in u_2]} M_1(u_2|u_1) \frac{p(y_1, y_2|x_1, x_2)}{p(y_2|x_2)},$$

$$q_2(v_2, v_3) = 1_{[x_3 \in u_3]} M_1(u_3|u_2) \frac{p(y_2, y_3|x_2, x_3)}{p(y_3|x_3)},$$

$$\ldots,$$

$$q_{n-2}(v_{n-2}, v_{n-1}) = 1_{[x_{n-1} \in u_{n-1}]} M_1(u_{n-1}|u_{n-2}) \frac{p(y_{n-2}, y_{n-1}|x_{n-2}, x_{n-1})}{p(y_{n-1} \mid x_{n-1})},$$

$$q_{n-1}(v_{n-1}, v_n) = 1_{[x_n \in u_n]} M_1(u_n \mid u_{n-1}) p(y_{n-1}, y_n \mid x_{n-1}, x_n).$$

Finally, knowing that the extension of CHMC specified in Proposition 4.1 is of interest (reference [22]), and knowing that the extension of CHMC to PMC also is (reference [13]), we can conjecture that there are some situations in which the extension of PMC specified in Proposition 4.2 would be of interest.

5. Evidential observation information in PMC

Let us first consider the case of one sensor. Considering again (2.7), let us consider the case in which $a(x)$ remains a probability distribution (it is equivalent to an EMC M_0) and in which the Markov chain induced on Ω^n by $b(x, y)$ (for fixed y) is replaced by an EMC M_1^y. Likely to what is said in the previous Section, the DS fusion of M_0 with M_1^y generalizes then the posterior distribution induced by (2.7). Such generalization can be well suited to several physical situations; for example, see Example 3.2 in Section 3 and a similar example in [4]. The aim of this Section is to extend such modeling to the PMC context. We have:

Proposition 5.1. *Let $\Omega = \{\omega_1, \ldots, \omega_k\}$ be the set of classes and* P(Ω) *be the power of Ω. Let M_0 be a Bayesian bba equivalent to the Markov chain $a(x)$ in (2.7), and let M_1^y be an EMC defined on $[\mathrm{P}(\Omega)]^n$ from the observed process $Y = y \in R^n$ by $M_1^y(u_1, \ldots, u_n) \propto \frac{p(y_1, y_2|u_1, u_2) \ldots p(y_{n-1}, y_n|u_{n-1}, u_n)}{p(y_2|u_2) \ldots p(y_{n-1}|u_{n-1})}$ (which extends $b(x, y)$ in (2.7) to an EMC).*

Then the margins $M^y(x_i)$ of the probability distribution given on Ω^n by the Bayesian bba $M^y = M_0 \oplus M_1^y$ are calculable with a number of elementary operations inferior or equal to $N = (5k + 1)2^k n - 5k2^k$. Therefore, the Bayesian MPM segmentation is workable for convenient n and k.

The proof is similar to the proof of the Proposition 4.1, with

$$q_1(v_1, v_2) = 1_{[x_1 \in u_1]} M_1(x_1) 1_{[x_2 \in u_2]} M_1(x_2|x_1) \frac{p(y_1, y_2|u_1, u_2)}{p(y_2|u_2)},$$

$$q_2(v_2, v_3) = 1_{[x_3 \in u_3]} M_1(x_3|x_2) \frac{p(y_2, y_3|u_2, u_3)}{p(y_3|u_3)},$$

$$\ldots,$$

$$q_{n-2}(v_{n-2}, v_{n-1}) = 1_{[x_{n-1} \in u_{n-1}]} M_1(x_{n-1}|x_{n-2}) \frac{p(y_{n-2}, y_{n-1}|u_{n-2}, u_{n-1})}{p(y_{n-1}|u_{n-1})},$$

$$q_{n-1}(v_{n-1}, v_n) = 1_{[x_n \in u_n]} M_1(x_n|x_{n-1}) p(y_{n-1}, y_n|u_{n-1}, u_n).$$

As the number of sets in $P(\Omega)$ is 2^k, we find the maximal number of elementary operations $N = (5k+1)2^k n - 5k2^k$ by applying Remark 2.1 to $r = 2^k$ and $m < 2^k$.

Finally, let us consider the case of numerous sensors: each $Y_i = (Y_i^1, \ldots, Y_i^m)$ takes its values in R^m. In the probabilistic framework, when the sensors are independent conditionally on X, (2.7) is extended to

$$p(x,y) = \underbrace{\frac{p(x_1,x_2)\ldots p(x_{n-1},x_n)}{p(x_2)\ldots p(x_{n-1})}}_{a(x)} \times \underbrace{\frac{p(y_1^1,y_2^1|x_1^1,x_2^1)\ldots p(y_{n-1}^1,y_n^1|x_{n-1}^1,x_n^1)}{p(y_2^1|x_2^1)\ldots p(y_{n-1}^1|x_{n-1}^1)}}_{b^1(x,y^1)} \times \cdots \times \underbrace{\frac{p(y_1^m,y_2^m|x_1^m,x_2^m)\ldots p(y_{n-1}^m,y_n^m|x_{n-1}^m,x_n^m)}{p(y_2^m|x_2^m)\ldots p(y_{n-1}^m|x_{n-1}^m)}}_{b^m(x,y^m)} \quad (5.1)$$

In this general probabilistic model we can either extend $a(x)$ or one of $b^1(x, y^1), \ldots, b^m(x, y^m)$ to a bba, as described above. So, let us denote by M_0 the possible extension of $a(x)$, and $M_1, \ldots, M_m$ the m possible extensions of $b^1(x, y^1), \ldots, b^m(x, y^m)$. Therefore when all the bbas $M_0, M_1, \ldots, M_m$ are Bayesian, $M = M_0 \oplus M_1 \oplus \cdots \oplus M_m$ is Bayesian and is equivalent to the posterior distribution $p(x|y)$ of the probabilistic model (5.1). Moreover, we can consider such extensions simultaneously knowing that if one at least among the bbas is Bayesian, then $M = M_0 \oplus M_1 \oplus \cdots \oplus M_m$ is Bayesian. However, we will consider the most general case, in which no hypothesis is made on any bba $M_0, M_1, \ldots, M_m$ (each of them can be Bayesian or not). We have:

Proposition 5.2. *Let $\Omega = \{\omega_1, \ldots, \omega_k\}$ be the set of classes and $\Lambda = P(\Omega)$ be the power of Ω. Let $M_0, M_1, \ldots, M_m$ be EMCs, where M_0 (possibly) extends $a(x)$ and $M_1, \ldots, M_m$ (possibly) extend $b^1(x, y^1), \ldots, b^m(x, y^m)$ in (5.1).*

Then the number of elementary operations required for calculation of the margins $M(u_i)$ of the bba $M = M_0 \oplus M_1 \oplus \cdots \oplus M_m$ is linear in observations. More precisely, it is inferior or equal to $N = 5(2^{k(m+2)})(n-1)$.

Proof. The EMCs M_j are of the form $M_j(u_1^j, \ldots, u_n^j) = M_j(u_1^j)M_j(u_2^j|u_1^j)M_j(u_n^j|u_{n-1}^j)$. As above, we will use the Lemma 2.1. Let us consider $q_1, q_2, \ldots, q_{n-1}$ defined on $[\Lambda \times \Lambda^{m+1}]^2$ (each $v_i = (u_i, w_i) \in \Lambda \times \Lambda^{m+1}$, with $w_i = (u_i^0, \ldots, u_i^m)$) by

$$q_1(v_1, v_2) = 1_{[u_1 = u_1^0 \cap \cdots \cap u_1^m]} \left[\prod_{j=0}^m M_j(u_1^j)\right] 1_{[u_2 = u_2^0 \cap \cdots \cap u_2^m]} \left[\prod_{j=0}^m M_j(u_2^j|u_1^j)\right],$$

$$q_2(v_2, v_3) = 1_{[u_3 = u_3^0 \cap \cdots \cap u_3^m]} \left[\prod_{j=0}^m M_j(u_3^j|u_2^j)\right],$$

$$\ldots$$

$$q_{n-1}(v_{n-1}, v_n) = 1_{[u_n = u_n^0 \cap \cdots \cap u_n^m]} \left[\prod_{j=0}^m M_j(u_n^j|u_{n-1}^j)\right].$$

By virtue of Lemma 2.1, the functions $q_1, q_2, \ldots, q_{n-1}$ define a Markov chain on $[\Lambda \times \Lambda^{m+1}]^n$, such that $M(v_i) = M(u_i, w_i)$ is computable and the number of elementary operations requested is linear in observations. Then $M(u_i)$ is obtained by $M(u_i) = \sum_{w_i} M(u_i, w_i)$.

Concerning the number of elementary operations, each q_i takes its values in $[\Lambda \times \Lambda^{m+1}]^2$, where Λ is the power set of $\Omega = \{\omega_1, \ldots, \omega_k\}$. Therefore $\Lambda \times \Lambda^{m+1}$ contains $r = 2^{k(m+2)}$ elements. Knowing that the calculation of all $p(v_i)$ requires $N = 5r(n-1)$ elementary operations (see Remark 2.1), the total number of elementary operations requested is inferior to $N = 5(2^{k(m+2)})(n-1)$. □

Therefore, the result $M(u_1, \ldots, u_n)$ in Proposition 5.2 is not necessarily a probability measure. However, it is still possible to perform statistical segmentation using one among different possible decision rules (see [12]). For example, the so-called "maximum of plausibility" works as follows. Once $M(u_i)$ is computed, we calculate the plausibility of each $x_i = \omega \in \Omega$ by $Pl(x_i = \omega) = \sum_{\omega \in u_i} M(u_i)$, and then the estimated $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ is given by $\hat{x}_i = \arg\max_{\omega} Pl(x_i = \omega)$. We obtain a compatible extension as when $M(u_1, \ldots, u_n)$ is a probability distribution, the maximum of plausibility $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ is the Bayesian MPM solution.

Remark 5.1. The number $r = 2^{k(m+2)}$ of elements in $\Lambda \times \Lambda^{m+1}$ strongly increases with km, which can pose problems. However, in practice different bba can be null on numerous elements of $\Lambda \times \Lambda^{m+1}$ and thus the number of elements effectively used can be much smaller than $2^{k(m+2)}$. In fact, the EMCs $M_0, M_1, \ldots, M_m$ can take their values in $(\Lambda_0)^n, (\Lambda_1)^n, \ldots, (\Lambda_m)^n$, respectively, where $\Lambda_0 \subset \Lambda, \Lambda_1 \subset \Lambda, \ldots, \Lambda_m \subset \Lambda$. Then $q_{i-1}(v_{i-1}, v_i)$ is non-null only for $v_i = (u_i, w_i) \in \Lambda \times \Lambda^{m+1}$, with $w_i = (u_i^0, \ldots, u_i^m)$, such that $u_i^0 \in \Lambda_0, \ldots, u_i^m \in \Lambda_m$, and $u_i = u_i^0 \cap \ldots \cap u_i^m$. In different examples specified in the paper, we can see that the number of elements in $\Lambda_0, \Lambda_1, \ldots$ effectively used is much smaller than $2^{k(m+2)}$.

Example 5.1. Let us consider the following example, already mentioned in [33] in the context of Markov fields. Let us imagine a satellite image representing a scene containing a river (ω_1), a sea (ω_2), urban area (ω_3), and forest (ω_4). Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\Lambda = \mathrm{P}(\Omega)$, and let n be the number of pixels. Imagine that we have some prior knowledge about the probabilistic distribution of "water", which is $\{\omega_1, \omega_2\}$, and "land", which is $\{\omega_3, \omega_4\}$, and that this distribution is a Markov chain. Thus M_0 is a Markov chain on $(\Lambda_0)^n$, with $\Lambda_0 = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$, and can be seen as an EMC on Λ^n. Furthermore, there are two sensors: an optical sensor Y^1, and an infrared sensor Y^2. The optical sensor can not see any difference between the river (ω_1) and the sea (ω_2), and there are some clouds hiding a part of the scene. Thus this sensor is sensitive to $\Lambda_1 = \{\{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_4\}, \Omega\}$ and, for every $2 \leqslant i \leqslant n$ and u_{i-1}^1, u_i^1 in Λ_1, this sensitivity is given by a likelihood $p(y_{i-1}^1, y_i^1 | u_{i-1}^1, u_i^1)$. This defines the function $b^1(u^1, y^1)$ as in (5.1), with $x \in \Omega^n$ replaced by $u^1 \in (\Lambda_1)^n$. The infrared sensor mainly detects temperature differences and can only detect a difference between the urban area and other classes; thus it is sensitive to $\Lambda_2 = \{\{\omega_3\}, \{\omega_1, \omega_2, \omega_4\}\}$. As above, this sensitivity is given by a likelihood $p(y_{i-1}^2, y_i^2 | u_{i-1}^2, u_i^2)$, which defines $b^2(u^2, y^2)$, with $u^2 \in (\Lambda_2)^n$. Then $b^1(u^1, y^1)$ defines Markov chain M_1 on $(\Lambda_1)^n$, and $b^2(u^2, y^2)$ defines a Markov chain M_2 on $(\Lambda_2)^n$ (both M_1 and M_2 are EMC on Λ^n). Finally, we have three EMC M_0, M_1, M_2 defined on Λ^n and, according to the Lemma 2.1, the margins $M(u_i)$ of the bba $M = M_0 \oplus M_1 \oplus M_2$ are computable. Let us see on which sets $A \in \Lambda$ the margins $M(u_i = A)$ are non-null. $M(u_i = A)$ is non-null if there exist $(A_0, A_1, A_2) \in \Lambda_0 \times \Lambda_1 \times \Lambda_2$ such that $A = A_0 \cap A_1 \cap A_2$ and $M_0(A_0)M_1(A_1)M_2(A_2)$ is non-null. Given the forms of Λ_0, Λ_1, and Λ_2 specified above, we find that $M(u_i = A)$ is non-null on $\Lambda^* = \{\{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_4\}\}$. Finally, having $M(u_i = A)$ on Λ^*, the corresponding plausibility on $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ is computed by $Pl(x_i = \omega_1) = Pl(x_i = \omega_2) \propto M(u_i = \{\omega_1, \omega_2\})$, $Pl(x_i = \omega_3) \propto M(u_i = \{\omega_3\})$, and $Pl(x_i = \omega_4) \propto M(u_i = \{\omega_4\})$, which is then used to estimate $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$ by $\hat{x}_i = \max_\omega Pl(x_i = \omega)$ for each $1 \leqslant i \leqslant n$.

6. Conclusions and perspectives

We dealt in this paper with different models allowing one simultaneously benefit from the hidden Markov chains (or their recent extensions) and the theory of evidence. We showed the calculability of the posterior marginal distributions, with immediate application to Bayesian hidden signal restoration. More precisely, we considered two cases: (i) the prior distribution of the hidden process becomes a "basic belief assignment" (bba); (ii) the distribution of the hidden process defined by the observed one becomes a bba. In both cases the Dempster–Shafer fusion generalizes the classical calculus of the posterior distribution, and thus is quite interesting in different restoration problems.

Let us mention some possible perspectives for further studies. The parameter estimation problem, whose solution is preliminary to unsupervised restoration methods, is undoubtedly among the most important, and difficult, problems. First applications of the "Expectation-Maximization" (EM [20]) or "Iterative Conditional Estimation" (ICE [4,10]) in triplet Markov models seem promising [3,21,22] and thus their applications in the general model proposed in the paper could be of interest. Otherwise, the independent sensors considered here could be extended to correlated sensors, as suggested in a simpler case in [29]. Finally, Markov chain model can be extended to a Markov tree model in a relatively straight manner [26,34], and thus the different results of the present paper could possibly be generalized to such models, with application to segmentation of multisensor and multiresolution data.

References

[1] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, Revue Scientifique et Technique de la Défense 11 (1991) 27–40.
[2] L.E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occuring in the statstical analysis of probabilistic functions of Markov chains, Annals of Mathematical Statistics 41 (1970) 164–171.
[3] D. Benboudjema, W. Pieczynski, Unsupervised image segmentation using triplet Markov fields, Computer Vision and Image Understanding 99 (3) (2005) 476–498.
[4] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, W. Pieczynski, Multisensor images segmentation using Dempster–Shafer fusion in Markov fields context, IEEE Transactions on Geoscience and Remote Sensing 39 (8) (2001) 1789–1798.
[5] 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 17 (8) (1996) 905–919.
[6] S. Bricq, C. Collet, J.-P. Armspach, Triplet Markov chain for 3D MRI brain segmentation using a probabilistic atlas, in: IEEE International Symposium on Biomedical Imaging, April 6–9, Arlington, Virginia, USA, 2006.
[7] N. Brunel, W. Pieczynski, Unsupervised signal restoration using hidden Markov chains with copulas, Signal Processing 85 (12) (2005) 2304–2315.
[8] O. Cappé, E. Moulines, T. Ryden, Inference in hidden Markov models, Springer, 2005.
[9] M. Chen, A. Kundu, J. Zhou, Off-line handwritten work recognition using a hidden Markov model type stochastic network, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (5) (1994).

[10] J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, IEEE Transactions on Signal Processing 45 (10) (1997) 2613–2615.
[11] T. Denoeux, Reasoning with imprecise belief structures, International Journal of Approximate Reasoning 20 (1) (1999) 79–111.
[12] T. Denoeux, Analysis of evidence-theoretic decision rules for pattern classification, Pattern Recognition 30 (7) (1997) 1095–1107.
[13] S. Derrode, W. Pieczynski, Signal and image segmentation using pairwise Markov chains, IEEE Transactions on Signal Processing 52 (9) (2004) 2477–2489.
[14] G.D. Fornay, The Viterbi algorithm, Proceedings of the IEEE 61 (3) (1973) 268–277.
[15] S. Foucher, M. Germain, J.-M. Boucher, G.B. Benié, Multisource classification using ICM and Dempster–Shafer theory, IEEE Transactions on Instrumentation and Measurement 51 (2) (2002) 277–281.
[16] L. Fouque, A. Appriou, W. Pieczynski, An evidential Markovian model for data fusion and unsupervised image classification, in: Proceedings of 3rd International Conference on Information Fusion, FUSION 2000 1, July 10th–13th, 2000, Paris, France, pp. TuB4-25–TuB4-31.
[17] J. Guan, D.A. Bell, Evidence Theory and its Applications, North-Holland, Amsterdam, 1991.
[18] F. Janez, A. Appriou, Theory of evidence and non-exhaustive frames of discernment: Plausibilities correction methods, International Journal of Approximate Reasoning 18 (1-2) (1998) 1–19.
[19] T. Koski, Hidden Markov models for bioinformatics, Kluwer Academic Publishers., 2001.
[20] G.J. McLachlan, T. Krishnan, EM algorithm and extensions, Wiley, Series in Probability and Statistics, 1997.
[21] P. Lanchantin, W. Pieczynski, Unsupervised non-stationary image segmentation using triplet Markov chains, Advanced Concepts for Intelligent Vision Systems (ACVIS 04), August 31–September 3, Brussels, Belgium, 2004.
[22] P. Lanchantin, W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chain using evidential priors, IEEE Transactions on Signal Processing 53 (8) (2005) 3091–3098.
[23] F. Le Ber, M. Benoît, C. Schott, J.-F. Mari, C. Mignolet, Studying crop sequences with CarrotAge, a HMM-based data mining software, Ecological Modelling 191 (1) (2006) 170–185.
[24] S. Le Hégarat-Mascle, I. Bloch, D. Vidal-Madjar, Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover, Pattern Recognition 31 (11) (1998) 1811–1823.
[25] S. Le Hégarat-Mascle, R. Seltz, Automatic change detection by evidential fusion of change indices, Remote Sensing of Environment 91 (3-4) (2004) 390–404.
[26] E. Monfrini, J. Lecomte, F. Desbouvries, and W. Pieczynski, Image and signal restoration using pairwise Markov trees, in: IEEE Workshop on Statistical Signal Processing (SSP 2003), Saint Louis, Missouri, September 28–October 1, 2003.
[27] P. Nicolas, L. Bize, F. Muri-Majoube, M. Hoebeke, F. Rodolphe, S. Dusko Ehrlich, B. Prum, P. Bessières, Mining bacillus subtilis chromosome heterogeneities using hidden Markov models, Nucleic Acid Research 30 (2002) 1418–1426.
[28] W. Pieczynski, A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, Machine Graphics & Vision 9 (3) (2000) 705–718.
[29] W. Pieczynski, Unsupervised Dempster–Shafer fusion of dependent sensors, in: Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation, SSIAI'2000, April, 2–4, Austin, Texas, USA, 2000, pp. 247–251.
[30] W. Pieczynski, Pairwise Markov Chains, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (5) (2003) 634–639.
[31] W. Pieczynski, C. Hulard, T. Veit, Triplet Markov chains in hidden signal restoration, in: SPIE's International Symposium on Remote Sensing, September 22–27, Crete, Greece, 2002.
[32] W. Pieczynski, F. Desbouvries, On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA 2005), Brest, France, May 2005.
[33] W. Pieczynski, D. Benboudjema, Multisensor triplet Markov fields and theory of evidence, Image and Vision Computing 24 (1) (2006) 61–69.
[34] J.-N. Provost, C. Collet, P. Rostaing, Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps, Computer Vision and Image Understanding 93 (2) (2004) 155–174.
[35] C. Raphael, Automatic segmentation of acoustic musical signals using hidden Markov models, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (4) (1999) 360–370.
[36] G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, 1976.

[37] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, International Journal of Approximate Reasoning 9 (1993) 1–35.
[38] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191–234.
[39] M. Thyer, G. Kuczera, Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model, Water Resources Research 36 (11) (2000) 3301.
[40] L.C. Thomas, D.E. Allen, N. Morkel-Kingsbury, A hidden Markov chain model for the term structure of bond credit risk spreads, International Review of Financial Analysis 11 (3) (2002) 311–329.
[41] F. Tupin, H. Maitre, I. Bloch, A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors, IEEE Transactions on Geoscience and Remote Sensing 37 (3) (1999) 1327–1343.
[42] R.R. Yager, J. Kacprzyk, M. Fedrizzi, Advances in the Dempster–Shafer theory of evidence, Wiley, 1994.