IMAGE AND SIGNAL RESTORATION USING PAIRWISE MARKOV TREES

E. Monfrini[1], *J. Lecomte*[2], *F. Desbouvries*[2], *and W. Pieczynski*[2],

[1] Université Claude Bernard Lyon 1, ISFA,
43, Bd du 11 novembre 1918, 69622 Villeurbanne.

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

ABSTRACT

This work deals with the statistical restoration of a hidden signal using Pairwise Markov Trees (PMT). PMT have been introduced recently in the case of a discrete hidden signal. We first show that PMT can perform better than the classical Hidden Markov Trees (HMT) when applied to unsupervised image segmentation. We next consider a PMT in a linear Gaussian model with continuous hidden data, and we give formulas of an original extension of the classical Kalman filter.

1. INTRODUCTION

Hidden Markov models (HMM), like hidden Markov chains (HMC), hidden Markov fields (HMF), or hidden Markov trees (HMT) admit a lot of applications in various domains, and in particular in signal and image processing [12]. These models have been recently generalized to pairwise Markov chains (PMC) [10], pairwise Markov fields (PMF) [6], and pairwise Markov trees (PMT) [7]. The aim of this paper is to present some further properties of the PMT introduced in [7] (in French). On the one hand, we present an application to unsupervised image segmentation of PMT with discrete hidden process. On the other hand, we consider PMT with continuous hidden processes, and we propose an original extension of the well known Kalman filter, which generalizes similar results proposed in the case of PMC [5] [9].

2. HIDDEN, PAIRWISE, AND TRIPLET MARKOV TREES

Let S be a finite set of points and $X=(X_s)_{s\in S}$, $Y=(Y_s)_{s\in S}$ two stochastic processes indexed on S. Each X_s takes its values in Ω (which will be a finite set in the next section and R^p in section 4), and Y_s takes its values in the set of observations, which will be the set R of real numbers in the next section, and R^q in section 4. Let S^1, …, S^N be a partition of S representing different « generations ». Each $s\in S^i$ admits $s^+\subset S^{i+1}$ (called its « children ») in such a way that every element $t\in S^{i+1}$ has a unique « parent » $t^-\in S^i$. We assume that S^1 is a singleton (its element s_r is called « root »). Then the distribution $p(x,y)$ of (X,Y) can be defined by four embedded models with increasing generality :

(i) the classical Hidden Markov Tree with independent noise (HMT-IN [1,4]). In this model,

$$p(x) = \ = p(x_{s_r})\prod_{s\in S\setminus S^1} p(x_s \mid x_{s^-}) \ , \qquad (1)$$

(i.e., X is a Markov Tree), $p(y|x)=\prod_{s\in S} p(y_s|x_s)$, and thus, putting $z=(x,y)$:

$$p(z)=p(x_{s_r})p(y_{s_r}|x_{s_r})\prod_{s\in S\setminus S^1} p(x_s|x_{s^-})p(y_s|x_s) \ ; \qquad (2)$$

(ii) the Hidden Markov Tree (HMT). In this model, X is a Markov tree as above, and the pairwise process $Z=(Z_s)_{s\in S}$, where $Z_s=(X_s,Y_s)$, is a Pairwise Markov Tree (PMT), which means that its distribution satisfies :

$$p(z)=p(z_{s_r})\prod_{s\in S\setminus S^1} p(z_s|z_{s^-}) \ ; \qquad (3)$$

(iii) the PMT $Z=(Z_s)_{s\in S}$, satisfying (3);

(iv) the Triplet Markov Tree (TMT) [8]. In this model, one introduces a latent variable $U=(U_s)_{s\in S}$ and assumes that the triplet $T=(X,U,Y)$ is a Markov Tree (i.e., T satisfies (3) with $t=(x,u,y)$ instead of $z=(x,y)$).

Let us remark that the greater generality of PMC with respect to HMT-IN appears locally at the transition probability level. In fact, $p(z_s|z_{s^-})$ in (3) can be written as

$$p(z_s|z_{s^-}) = p(x_s, y_s|x_{s^-}, y_{s^-}) = p(x_s|x_{s^-}, y_{s^-})p(y_s|x_s, x_{s^-}, y_{s^-}).$$

So we see that an HMT-IN is a PMT such that $p(x_s|x_{s^-}, y_{s^-}) = p(x_s|x_{s^-})$ and $p(y_s|x_s, x_{s^-}, y_{s^-}) = p(y_s|x_s)$.

3. DISCRETE HIDDEN PROCESS

Let us assume that $\Omega = \{\omega_1,...,\omega_k\}$, which means that the hidden process is a finite valued one. Let Z be a PMT defined in (3). Then the distribution $p(x|y)$ of X conditional to $Y = y$ keeps the same form (1). More precisely, for s child of s^-, we have [7] :

$$p(x_s|x_{s^-}, y) = \frac{\beta(x_s)p(z_s|z_{s^-})}{\sum_{\omega_s\in\Omega}\beta(\omega_s)p(\omega_s, y_s|z_{s^-})}, \quad (4)$$

in which the "backward" probabilities $\beta(x_s) = p(y_{s^{++}}|z_s)$ can be computed recursively by

$$\beta(x_s) = 1 \text{ for } s^+ = \varnothing,$$
$$\beta(x_s) = \prod_{u\in s^+}(\sum_{\omega_u\in\Omega}\beta(\omega_u)p(\omega_u, y_u|z_s)) \text{ for } s^+ \neq \varnothing \quad (5)$$

On the other hand, we have the following result, showing the greater generality of PMT with respect to HMT [7] :

Proposition 1

Let Z be a PMT defined in (3) and let (P) be the following property :

(P) For all $s \in S\setminus S^1, x_{s^-}, x_s \in \Omega$ and $y_{s^-} \in R$,

$$p(x_s|x_{s^-}, y_{s^-}) = p(x_s|x_{s^-}).$$

Then :

1. (P) implies that Z is an HMT (i.e., X is a Markov Tree);
2. Assume that each $s \in S$ has at least two children and that for each $t_1 \in s^+$ there exists $t_2 \in s^+$ such that $p(z_{t_1}|z_s) = p(z_{t_2}|z_s)$ (i.e., the distributions of Z_{t_1} and Z_{t_2} conditional on Z_s are equal). Then "Z is an HMT" implies (P).

In particular, (P) is a necessary and sufficient condition when $p(z_s|z_{s^-})$ does not depend on $s \in (s^-)^+$.

An analogous result shows the greater generality of TMT with respect to PMT [8].

Now, let Z be a PMT defined in (3), and let us consider the problem of computing the distribution of X_s conditional on $Y = y$ (the marginal « a posteriori » distribution), needed for instance when using Bayesian Maximum a Posteriori (MPM) segmentation. This distribution $p(x_s|y)$ can be calculated in the following way. Let $s \in S$, and let $s_1 = s_r, \cdots, s_m = s$ be the unique path (for each $2 \le i \le m$, s_{i-1} is the unique parent of s_i) relating s_r and s. We have

$$p(x_s \mid y) = \frac{\alpha^s(x_s)\beta^s(x_s)}{\sum_{\omega_s\in\Omega}\alpha^s(\omega_s)\beta^s(\omega_s)}, \quad (6)$$

in which $\alpha^s(x_s)$ is computed along the path $s_1 = s_r$, …, $s_m = s$ by

$$\alpha^{s_1}(x_s) = \beta^{s_1}(x_s),$$
$$\alpha^{s_i}(x_{s_i}) = \sum_{\omega_{s_{i-1}}\in\Omega} p(x_{s_i}|x_{s_{i-1}}, y)\alpha^{s_{i-1}}(\omega_{s_{i-1}}), \quad (7)$$

and $p(x_{s_i}|x_{s_{i-1}}, y)$ are computed via (4).

Next, once $p(x_s|y)$ has been computed for each $s \in S$, one can use the classical Bayesian MPM segmentation method in which $\hat{x} = (\hat{x}_s)_{s\in S}$ is obtained as $\hat{x}_s = \arg\max_{\omega\in\Omega} p(x_s = \omega|y)$. When the segmentation is performed in an unsupervised manner, which is important in real applications, one has to estimate the model parameters from $Y = y$. General methods like Expectation-Maximization (EM) [4] or Iterative Conditional Estimation (ICE) [10] have been applied in HMT and can be extended to the PMT and TMT cases. Classical HMT prove useful in statistical unsupervised segmentation problems [4]. The aim of the example presented in Figure 1 is to show that the greater generality of PMT can improve the results obtained with HMT. The class image is a 128x128 image and the Markov tree structure is a quad-tree [4]. So, we have the root and seven "generations", with the last generation S^7 being the set of 128x128 pixels. The noisy image $y^7 = (y_s)_{s\in S^7}$ is obtained by simulating a classical Gaussian noise on the generation S^6, and then using (2) to obtain $(y_s)_{s\in S^7}$. In the classical HMT case, we consider that only the last generation S^7 is noisy according to (2). The results presented in Figure 1 are obtained in an unsupervised manner, the parameters being estimated with ICE.

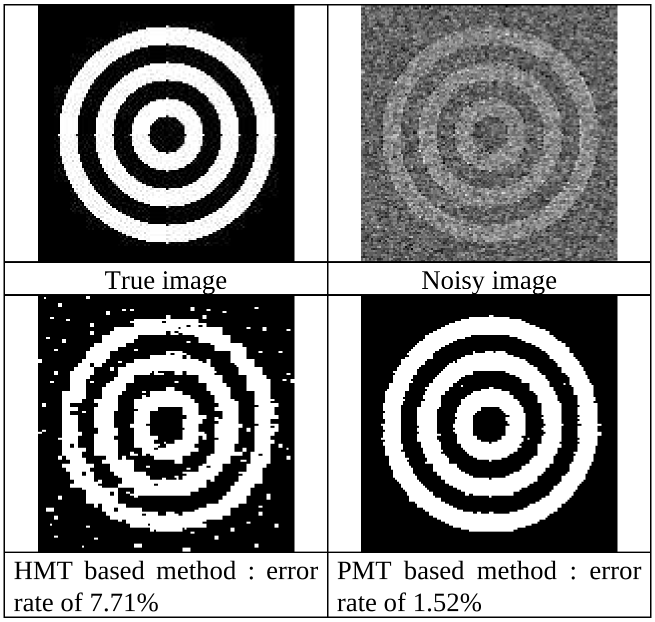

Figure 1. HMT and PMT based unsupervised Bayesian segmentation results.

4. CONTINUOUS HIDDEN PROCESS

Let us now consider a PMT $Z = (X, Y)$, in which each X_s takes its values in R^p and each Y_s takes its values in R^q. Equations (4-7) still hold (with the obvious difference that sums should be replaced by integrals), but may be difficult to compute in the general case.

So, let us now address the particular case in which in addition Z is a Gaussian process. Injecting this assumption in the algorithm of section 3 immediately leads to a Kalman-like smoothing algorithm which is omitted here for want of space.

In this section, we will rather show that in the Gaussian case, it is also possible to develop a Kalman-like adaptive filtering algorithm for PMT. Recall that S^1, …, S^N denotes the partition of S in terms of its successive « generations ». Let us put $X_n = (X_s)_{s \in S^n}$ and $X_{1:n} = (X_1, ..., X_n)$. The random vectors Y_n, $Y_{1:n}$, Z_n, and $Z_{1:n}$ are defined similarly. Since $Z = (Z_s)_{s \in S}$ is a Markov Tree, the time-varying process $Z = (Z_n)_{n \geq 1}$ is a Markov Chain. This observation enables us to adapt to PMT the classical Kalman filtering methodology which is valid in the context of Markov Chains. More precisely, our aim consists in recursively estimating (as new data become available) the p.d.f. of the last "leaves" X_{n+1} given all observed variables up to level $n+1$, i.e. we want to compute $p(x_{n+1} | y_{1:n+1})$ in terms of $p(x_n | y_{1:n})$ and of y_{n+1}.

Our assumptions are as follows. We assume that the model is linear and Gaussian :

$$\underbrace{\begin{bmatrix} X_s \\ Y_s \end{bmatrix}}_{Z_s} = \underbrace{\begin{bmatrix} F_s^1 & F_s^2 \\ H_s^1 & H_s^2 \end{bmatrix}}_{F_s} \underbrace{\begin{bmatrix} X_{s^-} \\ Y_{s^-} \end{bmatrix}}_{Z_{s^-}} + \underbrace{\begin{bmatrix} G_s^{11} & G_s^{12} \\ G_s^{21} & G_s^{22} \end{bmatrix}}_{G_s} \underbrace{\begin{bmatrix} U_s \\ V_s \end{bmatrix}}_{W_s}, \quad (8)$$

in which $W = (W_s)_{s \in S \setminus S^1}$ are random zero-mean vectors, independent and independent of Z_1. We also assume that W is Gaussian and that Z_1 is Gaussian with mean $\bar{z}_1$ and variance-covariance matrix P_1, which is denoted by $Z_1 \sim \mathrm{N}(\bar{z}_1, P_1)$. Then Z is a Gaussian process and consequently $p(x_n | y_{1:n})$ and $p(x_{n+1} | y_{1:n})$ are also Gaussian. Let us set $p(x_n | y_{1:n}) \sim \mathrm{N}(\hat{x}_{n|n}, P_{n|n})$ and $p(x_{n+1} | y_{1:n}) \sim \mathrm{N}(\hat{x}_{n+1|n}, P_{n+1|n})$, and let

$$E(W_s W_s^T) = Q_s, \ \tilde{Q}_s = \begin{bmatrix} \tilde{Q}_s^{11} & \tilde{Q}_s^{12} \\ \tilde{Q}_s^{21} & \tilde{Q}_s^{22} \end{bmatrix} = G_s Q_s G_s^T. \quad (9)$$

We shall also need the following notation : For n fixed, let $S^n = (s_1, ..., s_k)$, and let $s_i^+ = \{s_{i,p}^+\}_{p=1}^{j}$ (i.e. $s_{i,p}^+$ is the p^{th} son of node s_i). For $l, m \in \{1, 2\}$, let $\mathbf{F}_{n+1}^l$, $\mathbf{H}_{n+1}^l$, and $\tilde{\mathbf{Q}}_{n+1}^{l,m}$ be the following block-diagonal matrices :

$$\begin{aligned} \mathbf{F}_{n+1}^l &= \mathrm{diag}(F_{s_1^+}^l, ..., F_{s_k^+}^l), \\ \mathbf{H}_{n+1}^l &= \mathrm{diag}(H_{s_1^+}^l, ..., H_{s_k^+}^l), \\ \tilde{\mathbf{Q}}_{n+1}^{l,m} &= \mathrm{diag}(\tilde{Q}_{s_1^+}^{l,m}, ..., \tilde{Q}_{s_k^+}^{l,m}), \end{aligned} \quad (10)$$

in which

$$F_{s_i^+}^l = \begin{bmatrix} F_{s_{i,1}^+}^l \\ \vdots \\ F_{s_{i,j}^+}^l \end{bmatrix}, \quad H_{s_i^+}^l = \begin{bmatrix} H_{s_{i,1}^+}^l \\ \vdots \\ H_{s_{i,j}^+}^l \end{bmatrix}, \quad \tilde{Q}_{s_i^+}^{l,m} = \begin{bmatrix} \tilde{Q}_{s_i^+,1}^{l,m} & & 0 \\ & \ddots & \\ 0 & & \tilde{Q}_{s_i^+,j}^{l,m} \end{bmatrix}. \quad (11)$$

The following result is an extension of the classical Kalman filter [3] :

Proposition 2 (Kalman filter for PMT).

Let us assume that Z is a PMT and that model (8) holds. Suppose that $Z_1 \sim \mathrm{N}(\bar{z}_1, P_1)$ and $W_s \sim \mathrm{N}(0, Q_s)$.

Then $\hat{x}_{n+1|n+1}$ and $P_{n+1|n+1}$ can be computed from $\hat{x}_{n|n}$ and $P_{n|n}$ via :

Time-update equations

$$\hat{x}_{n+1|n} = \mathbf{F}^1_{n+1}\hat{x}_{n|n} + \mathbf{F}^2_{n+1}y_n \,, \quad (12)$$

$$P_{n+1|n} = \tilde{\mathbf{Q}}^{1,1}_{n+1} + \mathbf{F}^1_{n+1}P_{n|n}(\mathbf{F}^1_{n+1})^T \quad (13)$$

Measurement-update equations

$$\tilde{y}_{n+1} = y_{n+1} - \mathbf{H}^1_{n+1}\hat{x}_{n|n} - \mathbf{H}^2_{n+1}y_n \quad (14)$$

$$L_{n+1} = \tilde{\mathbf{Q}}^{2,2}_{n+1} + \mathbf{H}^1_{n+1}P_{n|n}(\mathbf{H}^1_{n+1})^T \quad (15)$$

$$K_{n+1|n+1} = (\tilde{\mathbf{Q}}^{1,2}_{n+1} + \mathbf{F}^1_{n+1}P_{n|n}(\mathbf{H}^1_{n+1})^T)L^{-1}_{n+1} \quad (16)$$

$$\hat{x}_{n+1|n+1} = \hat{x}_{n+1|n} + K_{n+1|n+1}\tilde{y}_{n+1} \quad (17)$$

$$P_{n+1|n+1} = P_{n+1|n} - K_{n+1|n+1}L_{n+1}K^T_{n+1|n+1} \quad (18)$$

Remarks

- The algorithm in Proposition 2 requires the inversion of the square matrix L_{n+1} defined in (15), the dimension of which is proportional to the number of variables in generation $n+1$ of the tree. However, this full-size matrix inversion can be avoided by conditioning with respect to each variable in y_{n+1} one after the other;
- The algorithm is valid under the implicit assumption that each node has at least one child, but can easily by adapted to the general case where some node(s) has(ve) no child;
- If each node has exactly one child, then the PMT reduces to a particular case of the Pairwise Markov Chain model introduced in [5, corollary 1 p. 72], and the algorithm of Proposition 2 reduces to the algorithm proposed for this model [5, eqs. (13.56) and (13.57)];
- For more general PMT which are neither linear nor Gaussian, one could consider Monte-Carlo based algorithms, which would extend the solutions already proposed in the context of PMC and TMC [2], and would provide a suboptimal solution to the filtering problem.

5. CONCLUDING REMARKS

The recently introduced PMT, which are strictly more general that HMT, can be used either in discrete or continuous hidden signal restoration, as well as in a supervised or an unupervised manner. Its greater generality can lead to an improvement of the results obtained with the classical HMT. For further research we may mention the possibilities of extending PMT to Pairwise Markov Graphical models, together with the associated methods of hidden process restoration and parameter estimation [11].

6. REFERENCES

[1] K. C. Chou, A. S. Willsky & A. Benveniste, Multiscale recursive estimation, data fusion, and regularization, IEEE tr. Autom. Control, Vol. 39, N. 3, pp. 464-78, March 1994

[2] F. Desbouvries and W. Pieczynski, Particle Filtering in Pairwise and Triplet Markov Chains, Proceedings of the IEEE – EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP 2003), Grado-Gorizia, Italy, June 8-11, 2003.

[3] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng., Trans. ASME, Series D, Vol. 82, N. 1, pp. 35-45, 1960

[4] J.-M. Laferté, P. Pérez, and F. Heitz, Discrete Markov image modeling and inference on the quadtree, IEEE Trans. on Image Processing, Vol. 9 , No 3, pp. 390 - 404, 2000.

[5] R. S. Lipster and A. N. Shiryaev, Statistics of Random Processes, Vol. 2 : Applications, chapter 13 : "Conditionally Gaussian sequences : filtering and related problems", Springer Verlag, Berlin, 2001.

[6] W. Pieczynski and A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, Machine Graphics and Vision, Vol. 9, No. 3, pp. 705-718, 2000.

[7] W. Pieczynski, Arbres de Markov Couple – Pairwise Markov Trees, Comptes Rendus de l'Académie des Sciences – Mathématique, Paris, Ser. I 335, pp. 79-82, 2002.

[8] W. Pieczynski, Arbres de Markov Triplet et fusion de Dempster-Shafer –Triplet Markov Trees and Dempster-Shafer fusion, Comptes Rendus de l'Académie des Sciences – Mathématique, Ser. I 336, pp. 869-872, 2003.

[9] W. Pieczynski and F. Desbouvries, Kalman Filtering using Pairwise Gaussian Models, International Conference on Acoustics, Speech and Signal Processing (ICASSP 03), Hong-Kong, April 2003.

[10] W. Pieczynski, Pairwise Markov Chains, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 25, No. 5, pp. 634-639, 2003.

[11] J. Whittaker, Graphical models in applied multivariate statistics, Wiley, 1990.

[12] A. S. Willsky, Multiresolution Markov models for signal and image processing, Proceedings of IEEE, Vol. 90, No. 8, pp. 1396-1458, 2002.