Hidden Evidential Markov Trees and Image Segmentation

Wojciech Pieczynski
Département Signal et Image, Institut National des Télécommunications,
9, rue Charles Fourier, 91000 Evry, France, e-mail Wojciech.Pieczynski@int-evry.fr

Abstract

The problem addressed in this paper is that of statistical segmentation of images using hidden Markov models. The problem is to introduce a prior evidential knowledge, defined by a mass function, or equivalently, by a belief function. We notice that the result of the Dempster-Shafer fusion of an evidential Markov field with a probability provided by the observations is not necessarily a Markov field. Thus using classical Bayesian segmentation as MPM or MAP is not tractable. In order to solve this problem, we show that the use of Markov trees, which is an another way of modelling the spatial dependence of the class random process, leads to tractable segmentation methods. In fact, the Dempster-Shafer fusion does not destroy the Markovianity in the *a posteriori* distribution and thus the classical Bayesian segmentation methods like as MPM or MAP may used. Furthermore, some ways of the model parameter estimation are indicated.

1 Introduction

The Dempster-Shafer theory of evidence [2, 11, 12], which can be seen as a kind of generalization of classical probabilistic models, can be useful in Bayesian image segmentation. More precisely, when the priors are "evidential" (i.e., the prior knowledge about the classes is modeled with a belief function, or, equivalently, with a mass function) and the observations are "classical" (i.e. modeled with the classical probability densities), their Dempster-Shafer fusion gives a probability measure, which is the posterior distribution in the classical purely probabilistic model. When wishing to take spatial interaction among pixels into account in such situations, the first idea is to consider a Markov field distribution for priors [4, 5]. Unfortunately, the result of fusion is no longer a classical Markov field, and thus the use of classical segmentation methods is not possible.

The original contribution of this paper is to show that when replacing the Markov field by a Markov tree, the result of the Dempster-Shafer fusion stays a Markov tree, which makes possible the use of classical Bayesian methods (like MAP or MPM) for segmenting the image.

The organization of the paper is as follows. In the next section we specify the problem in a simple context, without Markov models. The third section contains the main results concerning the hidden Markov trees. Conclusions are to be found in the fourth section.

2. Pixel by pixel context

Let S be the set of pixels, $X = (X_s)_{s \in S}$ the random field of classes (each X_s takes its values in $\Omega = \{\omega_1, \omega_2\}$), and $Y = (Y_s)_{s \in S}$ the random field of observations (each Y_s takes its values in R). Let $\pi_1 = P_{X_s}[\omega_1]$ and $\pi_2 = P_{X_s}[\omega_2]$ be priors, and f_1, f_2 be densities of P_{Y_s} conditional on $X_s = \omega_1, \omega_2$, respectively. For observed $Y_s = y_s$, the posterior distribution of X_s is then :

$$\pi_1^{y_s} = \frac{\pi_1 f_1(y_s)}{\pi_1 f_1(y_s) + \pi_2 f_2(y_s)}$$
$$\pi_2^{y_s} = \frac{\pi_2 f_2(y_s)}{\pi_1 f_1(y_s) + \pi_2 f_2(y_s)} \tag{1}$$

According to the Dempster-Shafer theory of evidence, let us assume now that our prior knowledge about classes in not precise enough to be modeled by a probability distribution, and instead an evidential measure (which equivalently can be a mass function, a plausibility function, or a belief function) is used [2, 11, 12]. We use a mass function, which is a probability Π on $\Omega^* = \{\{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}\}$:

$\Pi_1 = \Pi(\{\omega_1\})$, $\Pi_2 = \Pi(\{\omega_2\})$,
$\Pi_{12} = \Pi(\{\omega_1, \omega_2\})$.

To illustrate such situations, let us consider the case of a satellite picture taken in a region A (with probability 0.5), or in a region B (also with probability 0.5). We know that the proportions of classes ω_1, ω_2 are $0.2, 0.8$ in a region A, but we know nothing about these proportions in a region B. Such a knowledge can then be modeled with $\Pi_1 = 0.1$, $\Pi_2 = 0.4$, and $\Pi_{12} = 0.5$.

0-7803-5467-2/99/ $10.00 © 1999 IEEE

Then we consider that the information about the classes given with the observation $Y_s = y_s$ is modeled with the probability

$$q^{y_s}(\omega_1) = \frac{f_1(y_s)}{f_1(y_s) + f_2(y_s)}, \qquad q^{y_s}(\omega_2) = \frac{f_2(y_s)}{f_1(y_s) + f_2(y_s)} \tag{2}$$

The Dempster-Shafer combination rule applied to the mass function Π and the probability q^{y_s} gives a probability Π^{y_s}:

$$\Pi^{y_s}(\omega_1) = \frac{(\Pi_1 + \Pi_{12}) f_1(y_s)}{(\Pi_1 + \Pi_{12}) f_1(y_s) + (\Pi_2 + \Pi_{12}) f_2(y_s)}, \qquad \Pi^{y_s}(\omega_2) = \frac{(\Pi_2 + \Pi_{12}) f_2(y_s)}{(\Pi_1 + \Pi_{12}) f_1(y_s) + (\Pi_2 + \Pi_{12}) f_2(y_s)} \tag{3}$$

We see how (3) generalises (1) : when $\Pi_{12} = 0$ the mass function Π is comparable to the probability measure π, and the Dempster-Shafer combination rule gives the classical posterior probability (1). In other words, when our prior knowledge is "probabilistic", the posterior probability Π^{y_s} becomes the classical posterior probability.

3. Markovian context

3.1 Hidden Markov Fields

It is well known that the spatial interaction among the random variables (X_s) can be modeled via Markov random fields, and then different Bayesian segmentation, such as MAP [4] or MPM [8], are better than the Bayesian segmentations based on (1). How to introduce Markov models in the evidential context? We consider in [1] the case where the field $X = (X_s)_{s \in S}$ is classically probabilistic and Markovian, and the sensors are evidential (which mean, roughly speaking, that the probabilities (2) are replaced by some mass functions). Different simulation results presented in [1] show then that such a modeling is quite interesting. In this paper, we would like to replace the classical probabilistic priors with a Markov evidential prior. So the idea could be to consider a Gibbs measure P^* defined on $(\Omega^*)^{Card(S)}$ with :

$$P^*[x^*] = \gamma^* e^{-\sum_{e \in E} \Psi_e^*(x_e^*)} \tag{4}$$

where γ^* is a constant, E is the set of cliques (a clique being a subset of S which is either a singleton or a set containing mutual neighbours with respect to a neighbouring V), x_e^* is the restriction of x^* to e, and Ψ_e^* is a function, which depends on e only, and which takes its values in R.

Thus we have to fuse the mass function defined by (4) with the mass function defined by the observations. If the mass function (4) were probabilistic, that is to say if we had a classical hidden Markov field, the simplest way of modeling the probability measure provided by the observations is to assume that they are independent conditionally on the classes. Taking the same simple modeling, we have

$$p^y(\omega) = \prod_{s \in S} p^{y_s}(\omega_s) = \prod_{s \in S} \left(\frac{f_{\omega_s}(y_s)}{f_{\omega_1}(y_s) + f_{\omega_2}(y_s)} \right) \tag{5}$$

Putting $N = Card(S)$ and $\omega = (\omega_1, \ldots, \omega_N)$, the Dempster-Shafer fusion is written :

$$(P^* \oplus p^y)(\omega) \propto \sum_{\substack{\omega_1 \in x_1^* \\ \ldots \\ \omega_N \in x_N^*}} p^{y_1}(\omega_1) \ldots p^{y_N}(\omega_N) P^*(x_1^*, \ldots, x_N^*) \tag{6}$$

of the mass function P^* given on $(\Omega^*)^N$ by (4) and the probability distribution p^y given on Ω^N by (5) gives a probability distribution on Ω^N. This is a generalization of the classical hidden Markov model because when the mass function P^* is a classical Markov field, then $P^* \oplus p^y$ is the classical *a posteriori* distribution of this field. The drawback is that $P^* \oplus p^y$ can not be written as a simple Markov probability distribution.

In other words, in the hidden Markov field case the Markovianity is not preserved when fusing a Markovian evidential prior knowledge with classical probabilistic sensors observation knowledge.

As we will see in the next section, this Markovianity is saved when using the hidden Markov tree model.

3.2 Hidden Markov Trees

Another way of taking into account the spatial interactions in the a priori evidential knowledge we propose in this paper consists of using the Markov tree models, which have been successfully considered in the classical probabilistic context [3, 6, 7]. Let us specify our method in the particular case of the model presented in Figure 1. We consider $Card(S) = 2^n$ (with $n = 2$ in Figure 1). The class field is here $X^2 = (X_s^2)_{s \in S}$, the other X_i^j being "artificial". The distribution of X^2 is then the marginal distribution of $X = (X^0, X^1, X^2)$, and the distribution of X is given with the distribution of X^0, the distributions of X^1 conditional on X^0, and the distributions of X^2 conditional on X^1. Different calculations are feasible once we assume that four "sons" are always independent conditionally to the "father". In such a model the components of the class field $X^2 = (X_s^2)_{s \in S}$ are dependent (let us insist on the fact that they are independent conditionally on X^1, but they are dependent when removing the conditioning), and thus it is an alternative to the hidden Markov field model. In particular, the simulations of X^2 according to its posterior, i.e. conditional on the observations $Y = (Y_1, ..., Y_{16})$, distribution is feasible.

Let us return to the problem of Dempster-Shafer fusion. We have to fuse the probabilities defined by (5), with the mass function of the tree X^*. We show that, on the contrary to the Markov hidden field case, such a fusion is feasible. After fusion, the distribution of X^3 is a tractable probability law and, when the tree is probabilistic, the distribution of X^3 is exactly its posterior distribution. So we obtain a model in which spatial interactions in the class field and an evidential prior knowledge can simultaneously be taken into account.

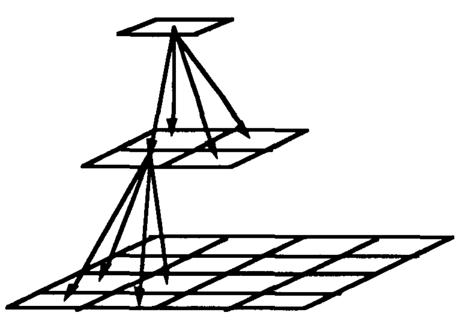

Figure 1
First two steps of the construction of "four children" pyramids.

So, let us consider n scales and $N = 4^n$ pixels. On each pixel $s \in S$, we consider a mass function q, which is a probability measure on Ω^*, and which models our prior evidential knowledge about the class lying on the pixel s. Spatial interactions among pixels are then modeled by an "evidential tree": we consider the scale 0 with a mass function q^0 and $2^k - 1$ mass functions $q(.|x^*)$, with $|x^* \in \Omega^* - \varnothing$, on Ω^*, each of which models the knowledge about the son conditionally on the father. So q^0 and $q(.|x^*)$ define a mass function $q^{(n)} = (q^0, q^1, ..., q^{n-1}, q^n)$ on the whole tree, which is a probability on $\Omega^* \times (\Omega^*)^4 \times ... (\Omega^*)^{4^{n-1}} \times (\Omega^*)^{4^n}$. Our evidential prior knowledge about classes is then modeled by q^n, the marginal probability of $q^{(n)}$ defined on $(\Omega^*)^{4^n}$.

The prior knowledge has to be fused with the knowledge provided by the observation $y = (y_s)_{s \in S}$. The latter observation knowledge will be classically modeled by the probability

$$p(x) = \prod_{s \in S} p(x_s) = \prod_{s \in S} \frac{f_{x_s}(y_s)}{\sum_{\omega \in \Omega} f_\omega(y_s)}, \qquad (7)$$

for $x = (x_s)_{s \in S}$

which is analogous to (5). According to the general Dempster-Shafer theory the result of fusion of the prior and the observation knowledge, modeled by

$$r = q^n \oplus p$$

is a probability on Ω^N.

Lemma

Let S be the set S of N pixels, $\Omega = \{\omega_1, ..., \omega_k\}$ the set of classes, and Ω^* the power set of Ω. Let us consider a probability q defined on $(\Omega^*)^N$ by

$$q(z_1, ..., z_N) = \sum_{t \in T} q(z_1|t) ... q(z_N|t) q'(t) \qquad (8)$$

where q' is a probability on a finite set T.

Let p be a probability on Ω^N verifying

$$p(x_1,\dots,x_N) = p(x_1)\dots p(x_N). \qquad (9)$$

The result of the Dempster-Shafer combination rule is then a probability on Ω^N verifying :

$$(q \oplus p)(x_1,\dots,x_N) \propto$$

$$\propto \sum_{t\in T}\left(\sum_{x_1\in z_1} q(z_1|t)\right)p(x_1)\dots\left(\sum_{x_N\in z_N} q(z_N|t)\right)p(x_N)q'(t)$$

Roughly speaking, the mixture form (8) is saved.

Proof.

We have

$$(q \oplus p)(x_1,\dots,x_N) \propto$$

$$\propto \sum_{\substack{x_1\in z_1,\\ \dots\\ x_N\in z_N}} q(z_1,\dots,z_N)p(x_1)\dots p(x_N) =$$

$$= \sum_{\substack{x_1\in z_1\\ \dots\\ x_N\in z_N}} \sum_{t\in T} q(z_1|t)\dots q(z_N|t)q'(t)p(x_1)\dots p(x_N) =$$

$$= \sum_{t\in T}\left(\sum_{\substack{x_1\in z_1\\ \dots\\ x_N\in z_N}} q(z_1|t)p(x_1)\dots q(z_N|t)p(x_N)\right)q'(t) =$$

$$= \sum_{t\in T}\left(\sum_{x_1\in z_1} q(z_1|t)p(x_1)\right)\dots$$

$$\dots\left(\sum_{x_N\in z_N} q(z_N|t)p(x_N)\right)q'(t) =$$

$$= \sum_{t\in T}\left(\sum_{x_1\in z_1} q(z_1|t)\right)p(x_1)\dots\left(\sum_{x_N\in z_N} q(z_N|t)\right)p(x_N)q'(t)$$

which ends the proof.

Proposition

Let us consider a mass function $q^{(n)} = (q^0, q^1,\dots,q^{n-1}, q^n)$ defined on the tree, which is a probability on $\Omega^* \times (\Omega^*)^4 \times \dots (\Omega^*)^{4^{n-1}} \times (\Omega^*)^{4^n}$, q the marginal probability on $(\Omega^*)^{4^n}$ induced from $q^{(n)}$, and the probability p defined on Ω^N from the observations $y_1,\dots,y_N$ by

$$p(x) = \prod_{s\in S} p(x_s) = \prod_{s\in S} \frac{f_{x_s}(y_s)}{\sum_{\omega\in\Omega} f_\omega(y_s)}$$

The probability $q^n \oplus p$ is then the marginal probability on Ω^N induced by $q^{(n),x} = (q^0, q^1,\dots,q^{n-1}, r^n)$ defined on the tree, which is a probability on $\Omega^* \times (\Omega^*)^4 \times \dots (\Omega^*)^{4^{n-1}} \times \Omega^{4^n}$ defined by $q^0, q^1,\dots,q^{n-1}$, and

$$r^n(x^n|x^{*n-1}) = \prod_{i=1}^{4^n}\left(p(x_i^n)\sum_{z_i^n\in\Omega^*/x_i^n\in z_i^n} q_i^n(z_i^n|x_{i^-}^{*n-1})\right)$$

Proof

The result is obtained by applying the Lemma to $T = \Omega^* \times (\Omega^*)^4 \times \dots (\Omega^*)^{4^{n-1}}$

As a consequence, stochastic realizations can be sampled in Ω^{4^n} with respect to $q^n \oplus p$ (which simply is the posterior probability distribution on Ω^{4N}). This allows one to estimate the marginal distributions of $q^n \oplus p$ and thus to segment the noisy image $y = (y_s)_{s\in S}$ by the classical MPM method.

When dealing with real images, the problem of the model parameter estimation may be of the most importance. For instance, in the simplest case the parameters are following. For the prior distribution on tree we have the mass function q^0 (three parameters with the sum equal to one), and three "father-son" conditional mass functions (for each possible father $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_1,\omega_2\}$ we have a mass function on $\{\{\omega_1\}, \{\omega_2\}, \{\omega_1,\omega_2\}\}$), which makes twelve real parameters. Concerning the noise parameters, all the distributions of Y conditional on X^n are given by two densities f_{ω_1}, f_{ω_2} which are densities of the distribution of each Y_s conditional on $X_s^n = \omega_1$ and $X_s^n = \omega_2$, respectively. Thus in the Gaussian case, the noise parameters are two means m_1, m_2 and two variances σ_1^2, σ_2^2. Finally, we have sixteen real parameters to estimate from $Y = y$. One possible way may then be to apply the Iterative Condition Estimation (ICE, [9]), which is a general estimation method in the case of hidden data. In a more complex case, in which the exact nature of densities f_{ω_1}, f_{ω_2} is not known (for instance, we do not know whether f_{ω_1} is Gaussian or Beta, and the same for f_{ω_2}) the generalized mixture ICE (ICE-GEMI, [5]) could possibly be used (ICE-GEMI has recently been successfully applied to the classical hidden Markov tree [10].

4 Conclusions

We have addressed in this paper the problem of Bayesian image segmentation. The problem was to introduce a prior evidential knowledge in a Markovian context. First, we have noticed that the hidden Markov field model is not adequate. In fact, the result of Dempster-Shafer fusion of the prior Markov mass function with the probabilistic mass function provided by the observations does not provide a Markov probability distribution. This is a drawback because the classical Bayesian segmentations with MPM or MAP are not feasible. Second, we showed that this drawback does not occur when the prior evidential knowledge is modeled by a hidden Markov tree (a Markov tree is a pyramid in which the base is made up with the image pixels) The probabilistic mass function provided by the observations can then be fused with an evidential Markov tree. The result is an evidential tree with a probability measure on the base. The latter probability, which is a generalisation of the posterior probability obtained in the classical probabilistic framework, may then be used to perform some Bayesian segmentation like MPM. Otherwise, let us recall that the MPM is much faster in the Markov tree context than in the Markov field case; in fact, there are no iterations in calculations.
The model parameter estimation does not pose a particular problem and some general methods like Iterative Conditional Estimation (ICE) may be used.

References

[1] A. Bendjebbour and W. Pieczynski, Segmentation d'images multisenseur par fusion évidentielle dans un contexte markovien, *Traitement du Signal*, Vol. 14, No. 5, pp. 453-464, 1997.

[2] 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.

[3] C. Bouman and M. Shapiro, A multiscale random field model for Bayesian image segmentation, *IEEE Transactions on IP*, Vol. 3, No. 2, pp. 162-177, 1994.

[4] S. Geman, and G. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on PAMI*, Vol. 6, pp. 721-741, 1984.

[5] N. Giordana and W. Pieczynski, Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation, *IEEE Transactions on PAMI* Vol. 19, No. 5, pp. 465-475, 1997.

[6] J.-M. Laferté, F. Heitz, P. Pérez, Modèles markoviens d'images et algorithmes d'estimation non linéaire sur le quadarbre, *Traitement du Signal*, Vol. 15, No. 3, pp. 213-230, 1998.

[7] M. R. Luentgen and A. S. Willsky, Likelihood calculation for a class of multiscale stochastic models with application to texture discrimination, *IEEE Transactions on IP*, Vol. 4, No. 2, pp. 194-207, 1995.

[8] J. Marroquin, S. Mitter, and T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *Journal of the American Statistical Association*, 82, pp. 76-89, 1987.

[9] W. Pieczynski, Statistical image segmentation, *Machine Graphics and Vision*, Vol. 1, No. 1/2, pp. 261-268, 1992.

[10] P. Rostaing, J.-N. Provost, and C. Collet, Unsupervised Multispectral Image Segmentation using Generalized Gaussian Noise Model, Proceedings of EMMCVPR'99, York, Englan, july 1999.

[11] G. Shafer, A mathematical theory of evidence, Princeton University Press, Prinston, 1976.

[12] R. R. Yager, J. Kacprzyk, and M. Fedrizzi, Advances in the Dempster-Shafer theory of evidence, Wiley, 1994.