Second IEEE Interantional Conference on Intelligent Processing Systems (ICIPC'98), Gold Coast, Australia, 4-7 August, 1998.

Multisensor Evidential Hidden Markov Fields and Image Segmentation

Azzedine Bendjebbour
Laboratoire de Statistique Théorique et Appliquée
Université Paris VI
4, Place Jussieu, 75005 Paris
France

Wojciech Pieczynski
Département Signal et Image
Institut National des Télécommunications
9, rue Charles Fourier, 91000 Evry
France

Abstract **– This paper deals with the statistical segmentation of multisensor images. In a Bayesian context the interest of using Hidden Markov Random Fields, which allow one to take contextual information into account, has been well known for about twenty years. In other situations, the Bayesian context is inadequate and one has to make use of the theory of evidence. The aim of our work is to propose an evidential model which can take into account contextual information via Markovian fields. We define a general evidential Markovian model and show that it is usable in practice. Some simulation results attest to the practical interest of the proposed model.**

1. INTRODUCTION

Multisensor image segmentation can take advantage of the theory of evidence [1, 4, 9, 10, 12-16], which allows one to take into account sensors of different nature. The segmentation step is performed after merging different sensors by applying the Dempster-Shafer combination rule. The theory of evidence approach can be seen as a generalization of the classical Bayesian approach, in the sense that in some situations one obtains the classical probabilistic framework. In particular, when at least one sensor is probabilistic, the Dempster-Shafer combination rule gives a probability distribution and the segmentation is then performed by some classical Bayesian rule. Otherwise, it is well known that in a Bayesian context the Hidden Markov Field model based segmentation methods may be of exceptional efficiency [5, 7, 11].
The aim of this work is to study how the advantages of the two approaches can be merged. In other words, the problem is to examine how the Dempster-Shafer combination rule can be used in a Markovian framework.
We proposed in [3] some heuristic manners of fusion. In this paper we introduce a new "evidential Markovian" model and show that the Dempster-Shafer fusion rule can be applied. Furthermore, we show that the most effective method presented in [3] is in fact a classical fusion method applied to the model we propose (subsection 2.3). The efficiency of the related segmentation method is illustrated by some simulations extracted from [3].
The organization of the paper is as follows:
In the next section we introduce a new model for evidential Markov fields and discuss its use in the problem of multisensor image segmentation. Section 3 is devoted to simulations and Section 4 concludes the paper.

II. MULTISENSOR EVIDENTIAL HIDDEN MARKOV FIELDS

A. Classical Hidden Markov Field Model

Classically, a random field model is as follows: given the set S of pixel, we consider two sets of random variables $X = (X_s)_{s \in S}$, $Y = (Y_s)_{s \in S}$ called "random fields". For m sensors, each X_s takes its values in a finite set of classes $\Omega = \{\omega_1, \ldots, \omega_k\}$ and each $Y_s = (Y_s^1, \ldots, Y_s^m)$ takes its values in R^m. The segmentation problem consists in estimating the unobserved realization $X = x$ of the field X from the observed realization $Y = y$ of the field Y, where $y = (y_s)_{s \in S}$ are m digital images representing the same scene. It is then generally solved by the use of a Bayesian strategy which is optimal with respect to some criterion.
The field $X = (X_s)_{s \in S}$ is said to be Markovian with respect to a neighbourhood V if its distribution can be written as

$$P_X[x] = \gamma e^{-U(x)} \qquad (2.1.1)$$

with

$$U(x) = \sum_{e \in E} \Psi_e(x_e) \qquad (2.1.2)$$

where E is the set of cliques (a clique being a subset of S which is either a singleton or a set containing mutual neighbors with respect to 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. In order to define the distributions of $Y = (Y_s)_{s \in S}$ conditional on $X = (X_s)_{s \in S}$, we will assume that the following three conditions hold:

(i) the random variables (Y_s) are independent conditionally on X,

(ii) the distribution of each Y_s conditional on X is its distribution conditional on X_s,

(iii) the random variables $Y_s^1, \ldots, Y_s^m$ are independent conditionally on X_s (i.e., the sensors are independent).

Due to these hypotheses all the distributions of Y conditional on X are defined, for k classes, by $k \times m$ distributions on R. To be more precise, let f_i^j denote the density of the distribution of Y_s^j conditional on $X_s = \omega_i$. Thus the distribution of (X,Y) is defined by the functions Ψ_e and the densities f_i^j. It is then possible to perform the segmentation by the maximum posterior mode (MPM) method:

$$\forall s \in S \;\; \hat{s}_{MPM}(s,y) = \arg\max_{x \in \Omega} P[X_s = x_s | Y = y] \quad (2.1.3)$$

by using the algorithm of Marroquin *et al.* [11]. Roughly speaking, realizations of X according to its posterior distribution are simulated and serve to estimate the posterior marginals in (2.1.3) with empirical frequencies.

B. Dempster-Shafer combination rule of evidential sensors.

In this subsection we set ourselves at one pixel. There are three equivalent ways to introduce the evidential measures, which are also called "fuzzy" measures on $\Omega = \{\omega_1, \ldots, \omega_k\}$: plausibilities, belief functions or mass functions. In this work we will adopt the representation by mass functions. Let us denote by $\Omega^* = \{\Omega_1, \ldots, \Omega_{2^k}\}$ the set of subsets of $\Omega = \{\omega_1, \ldots, \omega_k\}$. A mass function is a probability on Ω^*. Let us consider $m+1$ mass functions $M_0^s, M_1^s, \ldots, M_m^s$. Roughly speaking, M_0 models the prior information and $M_1^s, \ldots, M_m^s$ model the information contained in the observation of m sensors. The Dempster-Shafer combination rule, which enables one to aggregate theses different pieces of information, is as follows:

$$M^s(A) = \frac{1}{1-H} \sum_{A_0 \cap \ldots \cap A_m = A \neq \emptyset} \left[\prod_{j=0}^{m} M_j^s(A_j) \right] \quad (2.2.1)$$

with H the normalizing constant. The probability M^s is generally denoted by $M^s = M_0^s \otimes M_1^s \otimes \ldots \otimes M_m^s$. The mass functions can be seen as generalizations of the probability distributions in the following way: when the mass of every set but the singletons is null, it can be assimilated to a probability distribution. We will say that such a mass is "probabilistic", or "Bayesian". An important property is that if at least one mass function among $M_0^s, M_1^s, \ldots, M_m^s$ is probabilistic, then $M^s = M_0^s \otimes M_1^s \otimes \ldots \otimes M_m^s$ is also probabilistic.

Returning to our segmentation problem, we simply replace, at a given pixel s, $\Omega = \{\omega_1, \ldots, \omega_k\}$ by $\Omega^* = \{\Omega_1, \ldots, \Omega_{2^k}\}$. Thus, in the same manner as in the classical case, we now consider that we have $K = 2^k$ classes. The $K \times m$ densities of the distributions of Y_s^j conditional on $X_s = \Omega_i$, which correspond to the densities f_i^j of the classical model, will be denoted by g_i^j. For a given observation $y_s = (y_s^1, \ldots, y_s^m)$, the mass functions $M_1^s, \ldots, M_m^s$ are defined by

$$M_j^s(\Omega_i) = \frac{g_i^j(y_s^j)}{\sum_{q=1}^{K} g_q^j(y_s^j)} \quad (2.2.2)$$

and M_0^s, modelling the prior information, is independent of the observation $y_s = (y_s^1, \ldots, y_s^m)$. The whole information, from which we have to perform the segmentation, is then represented by $M^s = M_0^s \otimes M_1^s \otimes \ldots \otimes M_m^s$. It is important to recall that this evidential model is a generalization of the classical Bayesian model in the following way: when all mass functions $M_0^s, M_1^s, \ldots, M_m^s$ are probabilistic, then the mass function $M^s = M_0^s \otimes M_1^s \otimes \ldots \otimes M_m^s$ is simply the posterior distribution of X_s.

C. Markovian context

We show in this Section that such a model can be used in the Markovian context. We assume that $X = (X_s)_{s \in S}$ is a classical Markovian field and the observation fields $Y = (Y_s)_{s \in S}$ can be evidential. In accordance with (2.2.2) we define the m mass functions $M_1, \ldots, M_m$ corresponding to the m sensors $Y^1,, \;, Y^m$ with

$$M_j(A) = \frac{\prod_{s \in S} g_{A_s}^j(y_s^j)}{\sum_{B \in [\Omega^*]^N} \left[\prod_{s \in S} g_{B_s}^j(y_s^j) \right]} \quad (2.2.3)$$

where $N = Card(S)$ and $A = (A_s)_{s \in S}$ is every element of $(\Omega^*)^N$.

First, one can see that the mass function defined with (2.2.3) is the product, over $s \in S$, of the mass functions defined with (2.2.2) : $M_j = \prod_{s \in S} M_j^s$. Second, considering $M_{sen}^s = M_1^s \otimes \ldots \otimes M_m^s$ and $M_{sen} = M_1 \otimes \ldots \otimes M_m$, we have $M_{sen} = \prod_{s \in S} M_{sen}^s$. Thus, we can say that M_{sen} can be computed "pixel by pixel", which is feasible.

Keeping in mind that M_0 is equal to the distribution P_X defined by (2.1.1) and (2.1.2) of the Markov Field X, we now have to calculate $M_0 \otimes M_{sen}$. We can consider that M_{sen} is the mass function corresponding to one sensor. Let us denote by g_A the conditional densities defining it. We can then state the following:

Proposition

Let M_0 be a Markov Field with $M_0[x] = \gamma e^{-U(x)}$ and $U(x) = \sum_{e \in E} \Psi_e(x_e)$ (formulas (2.1.1) and (2.1.2)). Then $M_0 \otimes M_{sen}$ is a Markov Field, whose distribution coincides with the posterior distribution of M_0 classically corrupted with the independent noise

$$h_i(y_s) = \sum_{A/\omega_i \in A} g_A(y_A) \qquad (2.2.4)$$

As a consequence, the energy of $M_0 \otimes M_{sen}$ is computable and the classical segmentation methods, like MPM defined with (2.1.3), can be used.

Proof
We have

$$\begin{aligned} M_0 \otimes M_{sen}[x] &\propto \sum_{x \in B} (M_0[x] \times \prod_{s \in S} g_{B_s}(y_{B_s})) \propto \\ &\propto \sum_{x \in B} e^{-\sum_{e \in E} \psi_e(x_e)} \times \prod_{s \in S} g_{B_s}(y_{B_s}) = \\ &= e^{-\sum_{e \in E} \psi_e(x_e)} \sum_{x \in B} \prod_{s \in S} g_{B_s}(y_{B_s}) = \\ &= e^{-\sum_{e \in E} \psi_e(x_e)} \prod_{s \in S} (\sum_{x_s \in B_s} g_{B_s}(y_{B_s})) = \\ &= e^{-\sum_{e \in E} \psi_e(x_e)} \prod_{s \in S} h_{x_s}(y_s) \end{aligned} \qquad (2.2.5)$$

which completes the proof.

Let us notice that when M_{sen} is probabilistic, $M_0 \otimes M_{sen}$ is the classical posterior distribution of X.

III. SIMULATION RESULTS

A. Model considered

Let $\Omega = \{a, b, c\}$, $\Omega^* = \{A, B, C, D, F\}$ with $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $D = \{a, b\}$, and $F = \{a, b, c\}$. We consider two sensors; the first one Bayesian and the second one evidential. Thus the mass function defined by the second sensor is, for each pixel, a probability on $\Omega^* = \{A, B, C, D, F\}$. Let us denote the class field by $X^B = (X_s^B)_{s \in S}$ (B for "Bayesian"), and recall that each random variable X_s^B takes its values in $\Omega = \{a, b, c\}$. Let us artificially introduce an "evidential" field $X^E = (X_s^E)_{s \in S}$, i.e., such that each random variable X_s^E takes its values in $\Omega^* = \{A, B, C, D, F\}$. The field X^C does not intervene in the model proposed above, although, it will be useful in what follows. Fields $Y^B = (Y_s^B)_{s \in S}$ and $Y^E = (Y_s^E)_{s \in S}$ are the fields of observations corresponding to the two sensors. Thus the segmentation problem is the problem of estimating of the invisible realization of $X^B = (X_s^B)_{s \in S}$ from the realizations of $Y^B = (Y_s^B)_{s \in S}$ and $Y^E = (Y_s^E)_{s \in S}$.
The general problem we will study is to see how the observation of the consonant sensor Y^E can improve the quality of the segmentations obtained with the sole sensor Y^B. According to Section 2 the combination of the three mass functions considered gives a Markovian distribution with a computable energy. Thus it will be possible to perform the MPM from both sensors and compare its efficiency with the efficiency of the MPM obtained with the sole Bayesian sensor. Furthermore, we will compare the effectiveness of these two methods with the effectiveness of two blind methods.

B. Algorithms compared

(i) Algorithm $A1$

We will call $A1$ the algorithm MPM based on our model. Here M_0 is a classical Markov field, M_1 is a Bayesian sensor, and M_2 is an evidential sensor. Thus we have the following masses:

$$M_0(x) = \gamma e^{-\sum_{s \in S} \sum_{t \in V_s} \phi(x_s, x_t)} \qquad (3.1.1)$$

$$M_1(x) = \prod_{s \in S} \frac{g_{x_s}(y_s^1)}{\sum_{z_s \in \{a,b,c\}} g_{z_s}(y_s^1)} \qquad (3.1.2)$$

$$M_2(x^*) = \prod_{s \in S} \frac{g_{x_s^*}(y_s^2)}{\sum_{z_s^* \in \{A,B,C\}} g_{z_s^*}(y_s^2)} \qquad (3.1.3)$$

According to the results above, $M = M_0 \otimes M_1 \otimes M_2$ is the Markov distribution

$$M(x) = M_0 \otimes M_1 \otimes M_2(x) = \xi e^{-\eta(x)} \qquad (3.1.4)$$

with

$$\begin{aligned} \eta(x) = \sum_{s \in S} \sum_{t \in V_s} [\phi(x_s, x_t) - Log(g_{x_s}(y_s^1)) - \\ - Log(\sum_{x_s \in x_s^*} g_{x_s^*}(y_s^2))] \end{aligned} \qquad (3.1.5)$$

Simulating such Markov fields by using the Gibbs sampler is possible, so we can estimate the marginal distributions and apply the MPM as in (3.5).

(ii) $\hat{s}_{MPM}$ algorithm

Algorithm $\hat{s}_{MPM}$ is the classical MPM algorithm which uses only the Bayesian sensor.

(iii) Blind algorithms $BL1$, $BL2$

Algorithm $BL1$ is the simplest one: the segmentation is obtained by maximizing at each pixel

$$p(\omega) = P[X_s^B = \omega / Y_s^B = y_s^B]$$
for $\omega \in \{a, b, c\}$ (3.1.6)

Algorithm $BL2$ is obtained by maximizing at each pixel $r = p \otimes q$ obtained with Dempster-Shafer fusion from p, given by (3.1.6), and q, given by

$$q(\omega^*) = \frac{g_{\omega^*}(y_s)}{\sum_{\lambda \in \{A,B,C\}} g_\lambda(y_s)} \quad (3.1.7)$$
for $\omega^* \in \{A, B, C, D, F\}$

We can see that the difference between $A1$ and $\hat{s}_{MPM}$ is comparable to the difference between $BL2$ and $BL1$.

C. Numerical results

We consider a realization of a Markov Bayesian field, which is called "Bayesian image" in Figure 1. The "Evidential image", Figure 1, is obtained from the Bayesian image by random sampling. At each pixel $s \in S$, one samples a value in $\{A, B, C, D, F\}$ in the following way:

Case	BM			EM				
	a	b	c	A	B	C	D	F
1	0	2	4	0	2	4	1	6
2	0	1	2	0	2	4	1	6
3	0	0.5	1	0	2	4	1	6
4	0	0.2	0.4	0	2	4	1	6
5	0	0.1	0.2	0	2	4	1	6
6	0	0.1	0.2	0	3	6	1	6

Case	Error Rate			
	$BL1$	$BL2$	$\hat{s}_{MPM}$	$A1$
1	19.92	13.49	2.02	1.09
2	37.70	22.68	7.11	3.13
3	48.91	26.13	20.97	4.85
4	55.84	26.61	46.07	5.85
5	57.26	26.28	54.09	6.02
6	57.26	18.14	54.09	4.76

Table 1

Segmentations of six images. Three basic classes a, b, c and five evidential classes $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $D = \{a, b\}$, and $F = \{a, b, c\}$. Noise standard deviation equal to one. BM: Bayesian means, EM: evidential means. $BL1$: blind Bayesian sensor based segmentation, $BL2$: fusioned blind Bayesian and evidential sensors based segmentation, $\hat{s}_{MPM}$: Bayesian sensor MPM segmentation, $A1$: fusioned evidential MPM segmentation.

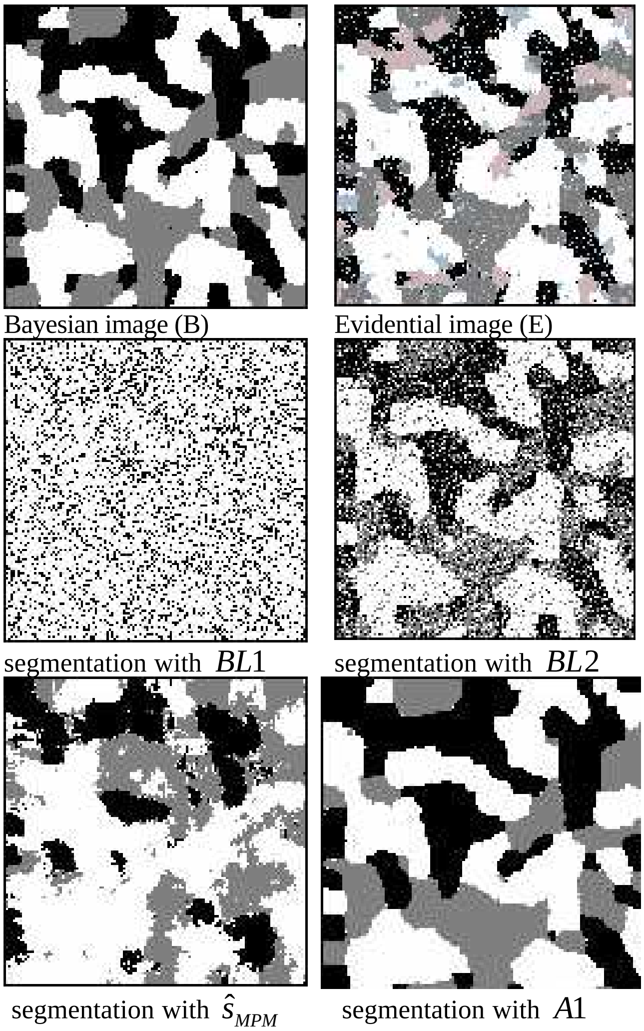

Figure 1
Segmentation of noisy images corresponding to the case 4, Table 1.

From the results of Table 1 we may put forth the following remarks:

Contribution of the consonant sensor in the blind context

We do not examine the contribution of the Markovianity here, and indeed limit our comments to the contribution of the consonant sensor in the "pixel by pixel" segmentation. Thus we have to compare results obtained with $BL2$ with those obtained with $BL1$. Since all parameters are known, the results obtained with $BL2$ are the best results one can expect when working "pixel by pixel". We note that the use of the consonant sensor always improves the results obtained with the sole Bayesian sensor, and in some situations its contribution is quite noticeable. The improvement in the results increases when the information tends to be concentrated on the consonant sensor; however, even when both sensors are equally noisy, the improvement can be non negligible.

Contribution of the contextual information in the Bayesian and Evidential contexts

The contribution of the contextual information in the Bayesian context is well known and it appears once again through results of Table 1. In fact, the error ratio obtained with $\hat{s}_{MPM}$ is always lower than the error ratio obtained with $BL1$. According to the same Table, we may say that this contribution is also sizeable in the evidential context; in fact, method $A1$ is always more effective than $BL2$.

Contribution of the consonant sensor in the Markovian context

We notice that the Markovian method $A1$ always performs better than $\hat{s}_{MPM}$. In certain situations, one of which is presented in Figure 1, the difference is visually quite significant. As a conclusion, we may say that $A1$, which is the classical fusion method directly associated with the new model proposed, can significantly improve the segmentation obtained with the only Bayesian sensor.

IV. CONCLUSIONS

The interest of the Hidden Markov Field models in image processing has been established for about twenty years [5, 7, 11]. Conversely, the use of "fuzzy measures", as plausibility or consonant measures, can be judicious in different image processing situations [1, 4, 9, 10, 12-16]. In particular, when one deals with the segmentation problem applied to multisensor images, considering some sensors as "evidential" can improve classical Bayesian segmentation. The aim of this work was to propose a model combining evidential and Markovian advantages. We showed that the method proposed heuristically in [3] can be seen as a classical Dempster-Shafer fusion method in the setting of the new model we propose. The latter model allows one to take into account evidential sensors, once the prior information can be modelled by a classical Markov field. In fact, the Dempster-Shafer combination rule is applicable, and the result is a classical Markov field. The latter allows one to apply some classical methods of segmentation, like Maximum Posterior Mode (MPM [11]), whose good behaviour in this context has been shown via simulations.

Let us mention that rendering our algorithms unsupervised is a perspective for future work. Different parameter estimation methods, which render the classical Markov models based methods unsupervised, have been proposed. Among others, it is possible to apply recent methods proposed in [6, 8], whose novelty is to allow one to also estimate the particular form of the noise densities g_i^j.

V. REFERENCES

[1] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, *Revue Scientifique et Technique de la Défense*, No. 11, pp. 27-40, 1991.

[2] J. Besag, On the statistical analysis of dirty pictures, *Journal of the Royal Statistical, Society*, Series B, 48, pp. 259-302, 1986.

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

[4] 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.

[5] R. Chellapa, and A. Jain Ed., Markov Random Fields, *Academic Press*, 1993.

[6] Y. Delignon, A. Marzouki, and W. Pieczynski, Estimation of Generalized Mixture and Its Application in Image Segmentation, *IEEE Transactions on Image Processing*, Vol. 6, No. 10, pp. 1364-1375, 1997.

[7] 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.

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

[9] J. Guan and D.A. Bell, Evidence Theory and Its Applications, North-Holland, Amsterdam, 1991.

[10] G. J. Klir and T. A. Folger, Fuzzy sets, uncertainty, and information, Prentice Hall, 1988.

[11] 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.

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

[13] P. Walley, Statistical reasoning with imprecise probabilities, Chapmann and Hall, London, 1991.

[14] R.R. Yager, A general approach to decision making with evidential knowledge, Uncertainty in Artificial Intelligence, L.N. Kanal and J.F. Lemmer ed., Elsevier Science Publishers, BV North-Holland, 1986.

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

[16] L.A. Zadeh, A simple view of the Dempster-Shafer theory of evidence, AI Magazine, Summer, pp. 85-90, 1986.