Multisensor Image Segmentation Using Dempster–Shafer Fusion in Markov Fields Context

Azzedine Bendjebbour, Yves Delignon, Laurent Fouque, Vincent Samson, and Wojciech Pieczynski

***Abstract*—This paper deals with the statistical segmentation of multisensor images. In a Bayesian context, the interest of using hidden Markov random fields, which allows one to take contextual information into account, has been well known for about 20 years. In other situations, the Bayesian framework is insufficient and one must make use of the theory of evidence. The aim of our work is to propose evidential models that can take into account contextual information via Markovian fields. We define a general evidential Markovian model and show that it is usable in practice. Different simulation results presented show the interest of evidential Markovian field model-based segmentation algorithms. Furthermore, an original variant of generalized mixture estimation, making possible the unsupervised evidential fusion in a Markovian context, is described. It is applied to the unsupervised segmentation of real radar and SPOT images showing the relevance of the proposed models and corresponding segmentation methods in real situations.**

***Index Terms*—Bayesian segmentation, data fusion, Dempster–Shafer combination rule, generalized mixture estimation, hidden Markov fields (HMF), iterative conditional estimation ICE, multisensor image segmentation, theory of evidence.**

I. Introduction

THIS PAPER addresses the problem of unsupervised statistical segmentation of multisensor images. We place ourselves in the context of hidden Markov field (HMF) models, which have shown their effectiveness over the past two decades. The success of these models is mainly due to the following two properties. First, they allow one to take into account the spatial interactions between pixels, which can render different Bayesian methods of segmentation very effective. The pioneering papers [2], [13], [21], [23] have been followed by many others and, understandably, only a small part of this rich bibliography can be mentioned (numerous references can be found in [6], [8], [16]). Second, the parameter estimation problem can be solved by different general methods like estimation–maximization [10], stochastic gradient [43], or iterative conditional estimation [25] among others. This results in numerous possibilities of unsupervised segmentation methods, in which all hidden Markov model (HMM) parameters are estimated in a previous step. Theoretical studies of these different methods prove rather tedious. However, different simulations and real image segmentation results presented in different papers attest that the parameter estimation step does not lessen the effectiveness of the real parameter based methods. We may mention [7] and [44] for EM-based methods, [43] for SG-based methods, [5], [19], [24], [31], and [32] for iterative conditional estimation (ICE)-based ones, and [2], [21] for different other methods. Most of these papers treat one sensor image. However, the multisensor case is quite similar to the one sensor case, at least in the Gaussian and reasonable SNR case, which is usually considered ([41], among others).

So roughly speaking, when a HMM, possibly with unknown parameters, is well suited to the data considered and when the noise is not too strong, there is no serious difficulty in performing segmentation.

The purpose of our work is to extend, using the theory of evidence [1], [15], [34], [37], these well known methods to some situations in which the use of classical HMFs poses difficulties. Of course, the use of theory of evidence in image processing is not new and has already given satisfactory results in various problems, like medical image classification [4] , or SAR image interpretation [39]. So the originality of our approach is to use this theory in a Markovian context. In fact, when the "pixel-by-pixel" multisensor image segmentation is concerned, the theory of evidence can be useful in numerous situations. For example, if we have two classes, "forest" and "water," and if we know that the proportion of forest is between 20% and 30% (so the proportion of water is between 70% and 80%), we can model this knowledge putting 20% on the class forest, 70% on the class water, and 10% on the class "water-or-forest." This prior knowledge is then merged with the knowledge provided by the observation via the Dempster–Shafer combination rule, resulting in a probability measure, which then appears as a generalization of the classical posterior probability. An analogous process can be applied when the priors are known exactly but one of sensors is very noisy and its probabilistic model (distribution of the noise) is very unreliable. Another example, which will be treated in some detail in this paper, is the following: the proportions of the two classes and the noise distributions are well known. However, there are some clouds. As we will see in the following, the presence of clouds can be modeled by a probability on the set (forest, water, and water-or-forest).

Let us remark that the generalization considered in this paper is one possible generalization among many others [1], whose respective extensions to Markov models may be considered topics for further work.

The organization of the paper is as follows. In the next section, we briefly recall the classical HMF model. In Section III,

Manuscript received April 4, 2000; revised November 8, 2000.

A. Bendjebbour is with the Laboratoire de Statistique Théorique et Appliquée, Université Paris VI, 75005 Paris, France.

Y. Delignon is with the Département Electronique, École Nouvelle d'Ingénieurs en Communication, 59650 Villeneuve d'Ascq, France.

L. Fouque is with the Département DTIM, Office National d'Études et Recherches Aérospatiales, 92323 Châtillon, France.

V. Samson and W. Pieczynski are with the Département CITI, Institut National des Télécommunications, 91000 Evry, France.

Publisher Item Identifier S 0196-2892(01)05474-2.

we specify some elements of the theory of evidence and specify, in the simple pixel-by-pixel context, how they can generalize the classical Bayesian image segmentation methods. Section IV is devoted to a new "evidential" HMF model we propose. Different experiments, related with synthetic or real images, are presented in Section V. Section VI concludes the paper.

II. Multisensor Hidden Markov Fields

We recall in this section the classical multisensor hidden Markov field (MHMF). Given the set S of pixels, 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 of 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 neighborhood V if its distribution can be written as

$$P_X[x] = \gamma e^{-U(x)} \tag{2.1}$$

with

$$U(x) = \sum_{e\in E} \Psi_e(x_e) \tag{2.2}$$

where

- E set of cliques (a clique being a subset of S that is either a singleton or a set of pixels mutually neighbors with respect to V);
- x_e restriction of x to e;
- Ψ_e 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.

1) The random variables (Y_s) are independent conditionally on X.
2) The distribution of each Y_s conditional on X is its distribution conditional to X_s.
3) The random variables $Y_s^1, \ldots, Y_s^m$ are independent conditionally on X_s (i.e., the sensors are independent).

Due to theses 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 to $X_s = \omega_i$. Thus, the distribution of (X, Y) is defined by the functions Ψ_e and the densities f_i^j. The posterior distribution of X is

$$P[X = x/Y = y] = \gamma(y)\exp\left(-\left[U(x) - \sum_{s\in S}\log\left[\prod_{j=1}^{m} f_{x_s}^j(y_s^j)\right]\right]\right). \tag{2.3}$$

It is then possible to perform the segmentation by the maximum *a posteriori* (MAP) method

$$\hat{s}_{\mathrm{MAP}}(y) = \underset{x\in\Omega^{Card(S)}}{\arg\max}\, P[X = x/Y = y] \tag{2.4}$$

or the maximum posterior mode (MPM) method

$$\forall\, s \in S \;\; \hat{s}_{\mathrm{MPM}}(s, y) = \underset{\omega\in\Omega}{\arg\max}\, P[X_s = \omega/Y = y]. \tag{2.5}$$

The first problem can be solved by the simulated annealing of Geman *et al.* [13] and the second one by the algorithm of Marroquin *et al.* [23]. Let us also mention the iterated conditional mode (ICM) of Besag [2], which is a fast approximation of the MAP.

Remark 2.1: Let us comment on the hypotheses 1)– 3) mentioned previously, some of which appear as rather strong, at least in some situations. Although quite unrealistic, 1) generalizes the "white noise," notion that is widely used in signal processing. In fact, in Gaussian case, if we consider $B = (B_s)_{s\in S}$ the random field of white Gaussian noise with unit variance for each variable B_s, we have $Y_s = m_{X_s} + \sigma_{X_s}B_s$. Although open to criticism, such models are currently used in image processing based on HMF models and seem to be very robust. In other words, this obviously wrong hypothesis does not significantly undermine, at least in situations described in different papers, the applicability of the different processing steps. So we will keep it for this paper and offer some remarks about possible extensions in perspectives, Section VI. The hypothesis 2) does not seem overly strong. The hypothesis 3) can be removed at this stage, however, it will be useful when considering "evidential" sensors. Now, sensors can really be independent in some situations, and thus in such situations 3) is justified. Some further remarks on this hypothesis will be developed in Section VI.

III. Dempster–Shafer Combination Rule in Pixel-by-Pixel Context

Let us position ourselves at one pixel and consider the following problem. We have two classes $\Omega = \{\omega_1, \omega_2\}$ with the prior probability given by $\pi_1 = P_{X_s}[\omega_1]$, $\pi_2 = P_{X_s}[\omega_2]$ and two images of a same scene: a radar image and an optical one. The radar image is very noisy, which is modeled by two noise distributions g_1^1 and g_2^1. The optical image is not very noisy, but there are some clouds. So in spots without clouds, the grey level is distributed according to a probability density g_1^2 for the first class, and it is distributed according to a probability density g_2^2 for the second class. In spots hidden by the clouds, the grey level is distributed according to a probability density g_c^2. The problem is to decide, from the observed $\binom{Y_s^1}{Y_s^2} = \binom{y_s^1}{y_s^2}$, whether the pixel s belongs to the class one or to the class two. If there were no clouds, i.e., if $g_c^2 = 0$, the fused information would be expressed, according to the classical probabilistic model, by the probability $p^{y_s^1, y_s^2}$

$$p^{y_s^1, y_s^2}(\omega_1) = \frac{g_1^1(y_s^1)g_1^2(y_s^2)}{g_1^1(y_s^1)g_1^2(y_s^2) + g_2^1(y_s^1)g_2^2(y_s^2)}$$
$$p^{y_s^1, y_s^2}(\omega_2) = \frac{g_2^1(y_s^1)g_2^2(y_s^2)}{g_1^1(y_s^1)g_1^2(y_s^2) + g_2^1(y_s^1)g_2^2(y_s^2)}. \tag{3.1}$$

Furthermore, in the Bayesian context, the probability $p^{y_s^1, y_s^2}$ can be fused with the prior probability $\pi = (\pi_1, \pi_2)$, where $\pi_1 = \pi(\omega_1)$ and $\pi_2 = \pi(\omega_2)$, resulting in the posterior probability $\pi^{y_s^1, y_s^2}$

$$\pi^{y_s^1, y_s^2}(\omega_1) = \frac{\pi_1 g_1^1(y_s^1) g_1^2(y_s^2)}{\pi_1 g_1^1(y_s^1) g_1^2(y_s^2) + \pi_2 g_2^1(y_s^1) g_2^2(y_s^2)}$$

$$\pi^{y_s^1, y_s^2}(\omega_2) = \frac{\pi_2 g_2^1(y_s^1) g_2^2(y_s^2)}{\pi_1 g_1^1(y_s^1) g_1^2(y_s^2) + \pi_2 g_2^1(y_s^1) g_2^2(y_s^2)}. \quad (3.2)$$

Finally, we can say that we have three probability distributions defined on $\Omega = \{\omega_1, \omega_2\}$: the prior probability π, the probability $p^{y_s^1}$ defined by the observation in the first sensor by

$$p^{y_s^1}(\omega_1) = \frac{g_1^1(y_s^1)}{g_1^1(y_s^1) + g_2^1(y_s^1)}, \quad p^{y_s^1}(\omega_2) = \frac{g_2^1(y_s^1)}{g_1^1(y_s^1) + g_2^1(y_s^1)}$$

and the probability $p^{y_s^2}$ defined by the observation in the second sensor by

$$p^{y_s^2}(\omega_1) = \frac{g_1^2(y_s^2)}{g_1^2(y_s^2) + g_2^2(y_s^2)}, \quad p^{y_s^2}(\omega_2) = \frac{g_2^2(y_s^2)}{g_1^2(y_s^2) + g_2^2(y_s^2)}.$$

The posterior probability (3.2) is then simply the normalized product of these three probabilities: $\pi^{y_s^1, y_s^2}(\omega_1) \propto \pi_1 p^{y_s^1}(\omega_1) p^{y_s^2}(\omega_1)$, and $\pi^{y_s^1, y_s^2}(\omega_2) \propto \pi_2 p^{y_s^1}(\omega_2) p^{y_s^2}(\omega_2)$. So the "Bayesian fusion" simply is a normalized product. How does one generalize this fusion to take the presence of clouds into account? One possible way of using the theory of evidence is the following. For the optical sensor, there are three classes: $\Omega^* = \{\omega_1, \omega_2, c\}$, where c is the class "clouds." However, we are only interested in whether a given pixel is ω_1 or ω_2. Thus, if it is a "cloud" pixel, we have no information about ω_1 or ω_2. In the theory of evidence, such a situation is modeled by considering a probability M on $\Omega^* = \{\{\omega_1\}, \{\omega_2\}, \Omega\}$, which is called a "mass function." The probability of Ω, which is the probability of c, models the uncertainty due to the presence of clouds. Here this mass function is naturally defined by

$$M^{y_s^2}(\{\omega_1\}) = \frac{g_1^2(y_s^2)}{g_1^2(y_s^2) + g_2^2(y_s^2) + g_c^2(y_s^2)}$$

$$M^{y_s^2}(\{\omega_2\}) = \frac{g_2^2(y_s^2)}{g_1^2(y_s^2) + g_2^2(y_s^2) + g_c^2(y_s^2)}$$

and

$$M^{y_s^2}(\Omega) = \frac{g_c^2(y_s^2)}{g_1^2(y_s^2) + g_2^2(y_s^2) + g_c^2(y_s^2)}.$$

Furthermore, the Dempster–Shafer combination rule, which is specified in the general case below, allows one to fuse π, $p^{y_s^1}$, and $M^{y_s^2}$ in the following manner:

$$\left(\pi \oplus p^{y_s^1} \oplus M^{y_s^2}\right)(\omega_1) \propto \pi_1 p^{y_s^1}(\omega_1) \left[M^{y_s^2}(\omega_1) + M^{y_s^2}(\Omega)\right]$$

$$\left(\pi \oplus p^{y_s^1} \oplus M^{y_s^2}\right)(\omega_2) \propto \pi_2 p^{y_s^1}(\omega_2) \left[M^{y_s^2}(\omega_2) + M^{y_s^2}(\Omega)\right]. \quad (3.3)$$

So in the presence of clouds, the probability (3.2) becomes

$$\left(\pi \oplus p^{y_s^1} \oplus M^{y_s^2}\right)(\omega_1) = \frac{\pi_1 g_1^1(y_s^1)[g_1^2(y_s^2) + g_c^2(y_s^2)]}{\pi_1 g_1^1(y_s^1)[g_1^2(y_s^2) + g_c^2(y_s^2)] + \pi_2 g_2^1(y_s^1)[g_2^2(y_s^2) + g_c^2(y_s^2)]}$$

$$\left(\pi \oplus p^{y_s^1} \oplus M^{y_s^2}\right)(\omega_2) = \frac{\pi_2 g_2^1(y_s^1)[g_2^2(y_s^2) + g_c^2(y_s^2)]}{\pi_1 g_1^1(y_s^1)[g_1^2(y_s^2) + g_c^2(y_s^2)] + \pi_2 g_2^1(y_s^1)[g_2^2(y_s^2) + g_c^2(y_s^2)]}. \quad (3.4)$$

We can see, according to (3.2) and (3.4), how $\pi \oplus p^{y_s^1} \oplus M^{y_s^2}$ generalizes the classical posterior distribution. When there are no clouds, i.e., $g_c^2 = 0$, the Dempster–Shafer fusion result $\pi \oplus p^{y_s^1} \oplus M^{y_s^2}$ becomes the classical posterior probability $\pi^{y_s^1, y_s^2}$.

In a general way, let us consider a set of classes $\Omega = \{\omega_1, \ldots, \omega_k\}$, the power set $\Omega^* = \{\Omega_1, \ldots, \Omega_{2^k}\}$ of Ω, and $m+1$ mass functions $M_0, M_1, \ldots, M_m$, which are probabilities on Ω^*. Recall that if a mass function only charges the singletons, it can be assimilated to a classical probability on Ω. Such a mass function will be called "Bayesian" or "probabilistic." Roughly speaking, M_0 models the prior information and $M_1, \ldots, M_m$ models the information contained in the observation of m sensors. The Dempster–Shafer combination rule, which enables one to aggregate these different pieces of information, is as follows:

$$M(A) = \frac{1}{1-H} \sum_{A_0 \cap \cdots \cap A_m = A \neq \varnothing} \left[\prod_{j=0}^{m} M_j(A_j)\right] \quad (3.5)$$

with

$$H = \sum_{A_0 \cap \cdots \cap A_m = \varnothing} \left[\prod_{j=0}^{m} M_j(A_j)\right]. \quad (3.6)$$

This defines a probability on Ω^*, which will be denoted by $M = M_0 \oplus M_1 \oplus \cdots \oplus M_m$. We have the following well known property.

Proposition 3.1: If at least one mass function among $M_0, M_1, \ldots, M_m$ is probabilistic, $M = M_0 \oplus M_1 \oplus \cdots \oplus M_m$ is probabilistic.

Returning to our segmentation problem, we simply replace, at a given pixel, $\Omega = \{\omega_1, \ldots, \omega_k\}$ by the power set Ω^* of the power set $\Omega^* = \{\Omega_1, \ldots, \Omega_{2^k}\}$ of Ω. Thus, in the same manner as in the classical case, we now consider that we have $K = 2^k - 1$ classes. The $K \times m$ densities of the distributions of Y_s^j conditional to $X_s = \Omega_i$, 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 $M_1, \ldots, M_m$ are defined by

$$M_j(\Omega_i) = \frac{g_i^j(y_s^j)}{\sum_{q=1}^{K} g_q^j(y_s^j)} \quad (3.7)$$

and M_0, modeling the prior information, is independent of the observation $y_s = (y_s^1, \ldots, y_s^m)$. However, remember that for a given Ω_i, we can have $M_j(\Omega_i) = 0$ for some sensors j, and $M_j(\Omega_i) \neq 0$ for some others. The whole information, from which we must perform the segmentation, is then represented by $M = M_0 \oplus M_1 \oplus \cdots \oplus M_m$. There exist different decision rules, for instance the classical Bayesian decision, in the case of probabilistic M. In order to simplify things we will consider in this paper that either M_0, or at least one mass function among $M_1, \ldots, M_m$, is probabilistic. Thus, M is probabilistic by virtue of Proposition 3.1, and our segmentation rule will be the Bayesian rule corresponding to a given loss function.

Remark 3.1: As noticed in the simple case above, the evidential model is a generalization of the classical Bayesian model in the following way. When all mass functions $M_0, M_1, \ldots, M_m$ are probabilistic, then the mass function $M = M_0 \oplus M_1 \oplus \cdots \oplus M_m$ is simply the posterior distribution of X_s.

Remark 3.2: The pixel-based approach described previously can actually be applied to other problems of classifying objects that are "spatially" independent.

Remark 3.3: Assuming M_0 probabilistic, we obtain a "generalization" of the classical Bayesian theorem. In fact, the probability distribution we obtain generalizes the classical posterior distribution. However, the situation here is different from the context of the Generalized Bayesian Theorem of Smets [36]: $M_1 \oplus \cdots \oplus M_m$ is defined by the observation $(y_s^1, \ldots, y_s^m)$, and it does not depend on the classes in Ω.

IV. Dempster–Shafer Fusion in Markov Field Context

Let us consider that M_0 is a classical Markov field, as in Section II, and we have m sensors, which possibly are evidential. We are looking for a consistent generalization of the classical HMM recalled in Section II. In the classical case and according to the hypotheses 1)– 3), the distribution of the field (X, Y) is

$$\begin{aligned} P_{(X,Y)}(x, y) &= P_X(x) f_x(y) \\ &= P_X(x) \prod_{s \in S} \prod_{j=1}^{m} f_{x_s}^j(y_s^j). \end{aligned} \tag{4.1}$$

First, according to the pixel-by-pixel approach specified in the previous section, we propose replacing the Bayesian fusion $\prod_{j=1}^m f_{x_s}^j(y_s^j)$ by the "evidential fusion" of the previous section: $M_{y_s} = M_s^1 \oplus \cdots \oplus M_s^m$, with each M_s^j defined by (3.7). Second, the consistency with the classical model leads us to keep the product $\prod_{s \in S}$, which models the spatial independence, conditionally on the class process, of the observations.

Finally, the information contained in the observation $Y = y$ will be modeled by the mass function M_y defined on each $A = (A_s)_{s \in S}$, with each A_s in Ω by

$$M_y[A] = \prod_{s \in S} M_{y_s}[A_s] = \prod_{s \in S} (M_s^1 \oplus \cdots \oplus M_s^m)[A_s]. \tag{4.2}$$

So the product $P_X(x) f_x(y)$, which is in the classical case a "Bayesian fusion" of the prior probability with the piece of evidence provided by the observation, must be replaced by $P_X \oplus M_y$, which is a probability according to the Proposition 3.1, and which extends to the evidential sensors case the classical posterior probability.

We have seen that in the classical HMFs case, the use of different Bayesian classification methods was possible because of the Markovianity of the posterior distribution of X. So we have to verify that the probability distribution $P_X \oplus M_y$ is a Markov distribution.

We have the following.

Proposition 4.1: Let us assume that M_0 is a Markov field with $M_0[x] = \gamma e^{-U(x)}$ and $U(x) = \sum_{e \in E} \Psi_e(x_e)$ [(2.1) and (2.2)], and let M_y be the mass function defined by (3.7) and (4.2). Then $M_0 \otimes M_y$ is a Markov field whose distribution is the same as the posterior distribution of M_0 classically corrupted with the independent noise

$$h_i(y_s) = \sum_{A/\omega_i \in A_s} g_{A_s}(y_s) \tag{4.3}$$

where

$$g_{A_s}(y_s) = \sum_{A_{i_1} \cap \ldots \cap A_{i_m} = A_s \neq \oslash} \Big/ \left[\prod_{j=1}^{m} g_{i_j}^j(y_s^j) \right]. \tag{4.4}$$

So the energy of $M_0 \oplus M_{s \in n}$ is generally computable and thus, the classical segmentation methods can be used.

Proof:

$$\begin{aligned} &M_0 \oplus M_y[x] \\ &\quad \propto \sum_{x \in B} \left(M_0[x] \times \prod_{s \in S} g_{B_s}(y_s) \right) \\ &\quad \propto \sum_{x \in B} e^{-\sum_{e \in E} \psi_e(x_e)} \times \prod_{s \in S} g_{B_s}(y_s) \\ &\quad = e^{-\sum_{e \in E} \psi_e(x_e)} \sum_{x \in B} \prod_{s \in S} g_{B_s}(y_s). \end{aligned} \tag{4.5}$$

On the other hand, knowing that $[x \in B] \Leftrightarrow [\forall s \in S \quad x_s \in B_s]$, we have

$$\sum_{x \in B} \left(\prod_{s \in S} g_{B_s}(y_s) \right) = \prod_{s \in S} \left(\sum_{x_s \in B_s} g_{B_s}(y_s) \right). \tag{4.6}$$

Equations (4.5) and (4.6) give

$$\begin{aligned} &M_0 \oplus M_y[x] \\ &\quad \propto e^{-\sum_{e \in E} \psi_e(x_e)} \prod_{s \in S} \left(\sum_{x_s \in B_s} g_{B_s}(y_s) \right) \\ &\quad = e^{-\sum_{e \in E} \psi_e(x_e)} \prod_{s \in S} h_{x_s}(y_s) \end{aligned} \tag{4.7}$$

which completes the proof.

Finally, when at least one sensor is evidential, their fusion is easily done in the pixel-by-pixel way. When at least one of them is probabilistic (as in the simple example of clouds in optical and radar images above), the result of the fusion is a probability measure and then the posterior Markov distribution of X is obtained as in the classical case, with $\prod_{j=1}^m f_{x_s}^j(y_s^j)$ in (2.3)

replaced by the fused probability M_{y_s}. If all sensors are evidential, $\prod_{j=1}^{m} f_{x_s}^{j}(y_s^j)$ in (2.3) is replaced with (4.3).

Remark 4.1: Let us notice that the model we propose appears, with respect to the richness of possibilities of the theory of evidence [11], [35], [38], [40], as a quite particular one. This particularism is mainly due to the particular form of M_y defined by (4.2). However, even in this particular setting, the fusion is workable and it extends the possibilities of the widely used HMFs.

Remark 4.2: We have assumed that all sensors were concerned with the same frame of discernment Ω. However, when sensors are concerned with partially overlapping frames, there exist different, more general methods of evidential fusion [17], [38] that may be used in the context considered in this paper.

Remark 4.3: The Markovian approach described in this section can also be applied to other problems of classifying objects. The main property to be verified is some spatial interaction among classes of considered objects, and this interaction can be reasonably modeled by a Markov field distribution.

V. Experiments

This section is devoted to three series of experiments. In all series, we consider four algorithms: two local algorithms and two global ones. Algorithm global Bayesian (GB) is the classical probabilistic MPM algorithm, which uses the sole Bayesian sensor. Algorithm global evidential (GE) is the new algorithm described in the previous section. In the same way, blind Bayesian (BB) will denote the classical Bayesian, which will use the sole Bayesian sensor, pixel-by-pixel method. Finally, we will call blind evidential (BE) the method which will use the fused information from the both Bayesian and Evidential sensors, in the pixel-by-pixel way. The aim of the experiments is to study how the use of the evidential sensor can improve the segmentation methods which use the only Bayesian sensor, in pixel-by-pixel way as well as in the Markovian one.

The first series of experiments, described in the Section V-A, is concerned with some hand-drawn images. Section V-B is devoted to the unsupervised fusion using an original variant of the generalized mixture estimation [9], [14], and in the last Section V-C, we present some results of real world images segmentation.

A. Simulated Images

Let $\Omega = \{a, b, c\}$ be a set of three classes and $\mathrm{p}\{\Omega\}$ the power set of Ω. We will consider two cases. In the first one, the evidential sensor is a consonant one, which means that the mass function defined by this sensor is a probability on $\Omega^* = \{A, B, C\}$ with $A = \{a\}$, $B = \{a, b\}$, and $C = \{a, b, c\}$. According to the general fusion rule, for p a probability on $\Omega = \{a, b, c\}$ and q, a probability on $\Omega^* = \{A, B, C\}$, the result of the Dempster–Shafer fusion $r = p \oplus q$ is the probability on $\Omega = \{a, b, c\}$ given by $r(a) \propto p(a)[q(A) + q(B) + q(C)]$, $r(b) \propto p(b)[q(B)+q(C)]$, and $r(c) \propto p(c)q(C)$. The Bayesian image (Image 1, Fig. 1) is a realization of a Markov field.

The results concerning the second case are presented in Table II and Fig. 2 (they are extracted from [3], where we have

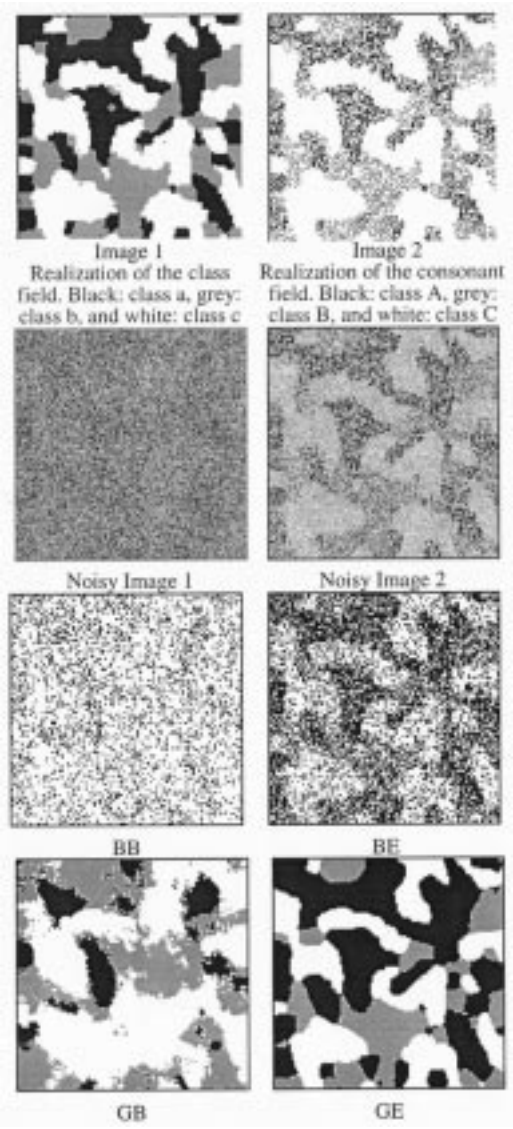

Fig. 1. Segmentation of noisy images corresponding to the case 9, Table I.

TABLE I
Results of Segmentations with Different Methods in the Consonant Case: $A = \{a\}$, $B = \{a, b\}$, and $C = \{a, b, c\}$. Noise Standard Deviation Equal to One. BM: Bayesian Means, CM: Consonant Means

Case	BM			CM			Error Rate			
	a	b	c	A	B	C	BB	BE	GB	GE
1	0	3	6	0	3	6	8.11	6.87	0.65	0.45
2	0	2	4	0	2	4	19.37	17.91	2.01	1.75
3	0	1	2	0	1	2	37.75	38.68	7.26	10.04
4	0	1	2	0	2	4	37.75	34.05	7.26	6.65
5	0	1	2	0	4	8	37.75	29.46	7.26	4.50
6	0	0.5	1	0	2	4	49.01	43.65	20.70	16.89
7	0	0.5	1	0	4	8	49.01	37.54	20.70	8.62
8	0	0.2	0.4	0	2	4	55.73	48.32	46.16	30.24
9	0	0.2	0.4	0	3	6	55.73	42.86	46.16	17.02
10	0	0.2	0.4	0	4	8	55.73	40.43	46.16	12.52
11	0	0.1	0.2	0	2	4	57.23	47.98	55.21	33.33
12	0	0.1	0.2	0	3	6	57.23	41.97	55.21	18.87
13	0	0.1	0.2	0	4	8	57.23	38.90	55.21	14.31

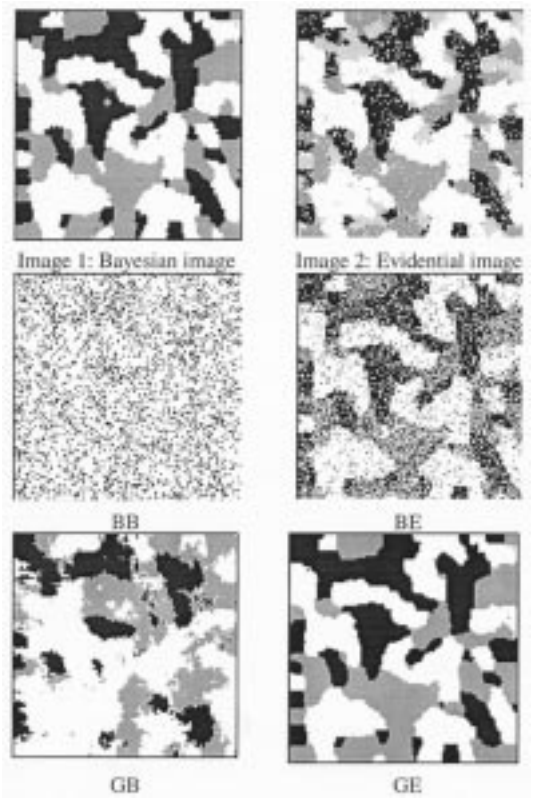

Fig. 2. Segmentation of noisy images corresponding to the case 4, Table II. BM: Bayesian means and CE: evidential means.

proposed the algorithm GE in a heuristic manner). These results concern the case $\Omega = \{a, b, c\}$, and $\Omega^* = \{A, B, C, D, E\}$, with $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $D = \{a, b\}$, and $E = \{a, b, c\}$. The Bayesian image is the same as Image 1, Fig. 1, and the evidential image (Image 2, Fig. 2) is obtained from Image 1 by a stochastic sampling analogous to the sampling used in the first case. As above, the noise variance is equal to one, and the different means are presented in Table II.

These different results show that the use of the evidential sensor can be beneficial, as well in the blind case as in the Markovian one. The more interesting cases are those in which the Bayesian images are much noisier that the evidential ones.

Of course, other situations could be considered in an analogous manner. According to the generalized setting of Sections III and IV, analogous segmentations could be used in every situation where we have a Bayesian sensor and an evidential one, the latter being able to distinguish some elements of the power set Ω^*.

B. Parameter Estimation and Unsupervised Image Fusion

When dealing with real images, the parameter estimation problem becomes an important factor. Although the considerations to follow are quite general, in order to simplify things, we shall place ourselves in a particular, relatively simple, case. Concerning the Markov class field we take a very simple Potts model, in which the energy [see (2.2)] is a sum over the couples of neighbor pixels: $U(x) = \sum_{(s,t)} \varphi(x_s, x_t)$. We take the potential function φ defined by $\varphi(x_s, x_t) = -\alpha$ if $x_s = x_t$, and $\varphi(x_s, x_t) = \alpha$ if $x_s \neq x_t$. Furthermore, let us consider

TABLE II
RESULTS OF SEGMENTATIONS WITH DIFFERENT METHODS. NOISE STANDARD DEVIATION EQUAL TO ONE. SAME NOTATIONS AS IN TABLE I, WITH $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $D = \{a, b\}$, AND $E = \{a, b, c\}$

Case	BM			EM					Error Rate			
	a	b	c	A	B	C	D	E	BB	BE	GB	GE
1	0	2	4	0	2	4	1	6	19.92	13.49	2.02	1.09
2	0	1	2	0	2	4	1	6	37.70	23.68	7.11	3.13
3	0	0.5	1	0	2	4	1	6	48.91	26.13	20.97	4.85
4	0	0.2	0.4	0	2	4	1	6	55.84	26.61	46.07	5.85
5	0	0.1	0.2	0	2	4	1	6	57.26	26.28	54.09	6.02
6	0	0.1	0.2	0	3	6	1	6	57.26	18.14	54.09	4.76

$\Omega = \{a, b, c\}$ and $\Omega^* = \{A, B, C, D\}$, with $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, and $D = \Omega = \{a, b, c\}$. Thus we have to determine g_a, g_b, and g_c concerning the Bayesian sensor, and g_A, g_B, g_C, and g_D concerning the Evidential one. Estimating these functions from the observations is the mixture estimation problem and we propose to solve it by applying a new variant of a recent method of "generalized" mixture estimation [14]. A mixture is called generalized when the form of each component is not known; however, it belongs to a given set of forms. For instance, if each of the densities g_a, g_b, and g_c can be Normal or exponential, we have eight possibilities of classical mixtures and the additional difficulty is to determine in which case we lie.

Remark 5.2: The generalized mixture estimation is more general than classical mixture estimation in that when the two classes have the same form of noise, it remains valid. Now, it is possible to have two classes with different forms of noise. For instance, the form of the sea surface variability can depend on weather and so, when the weather changes, the "noise" form can evolve. Furthermore, the noise for a given class can already be a mixture of distributions; the "vegetation" class can contain "grass," "trees," "bushes," and thus the distribution of "vegetation" can no longer be of a classical form. When trying to find a good approximation for this distribution, the general mixture offers more possibilities than the classical one.

Of course, if we know the forms of different noise processes, we have to use a classical mixture estimation like Gibbsian expectation-estimation method (GEM [7]), ICE [25], or SG [43].

The proposed method is as follows.

1) Consider (X, Y^1), which is a classical HMF described in Section II. We assume that each of the densities g_a, g_b, and g_c can be a normal, beta, or gamma density. We apply our method, whose novelty is specified in the following, to estimate this Markovian generalized mixture.
2) Consider (Y_s^2) without any Markovian structure. The distribution of Y_s^2 is thus a classical mixture on R of four distributions g_A, g_B, g_C, and g_D. We could still apply our method to treat this mixture as a generalized mixture, although in simulations below, we consider that it is a Gaussian mixture and we estimate it with the classical ICE.

The general mixture estimation method proposed in [14], called the ICE-GEMI algorithm, is an iterative method. At step q, let α^q and g_a^q, g_b^q, g_c^q be current prior

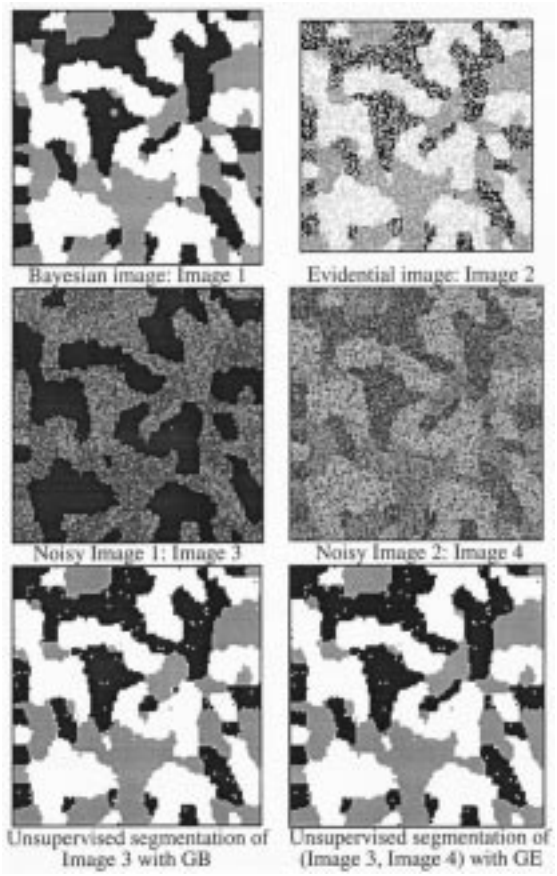

Fig. 3. Example of unsupervised segmentations of two sensor hidden evidential Markov field images, case 4, Table III.

parameters and current densities g_a, g_b, and g_c. The updating is as follows.

a) Simulate x^q, a realization of X, according to its α^q and g_a^q, g_b^q, g_c^q-based distribution conditional to $Y^1 = y^1$. Calculate $\alpha^{q+1} = \hat{\alpha}(x^q)$, with $\hat{\alpha}$ the Younes method [42].

b) For $i = a, b, c$, consider $S_i^q = \{s \in S / x_s^q = i\}$. Let $y_i^q = (y_s)_{s \in S_i^q}$. For each $i = a, b, c$ estimate from y_i^q the three "candidate" densities: f_i^1 (normal), f_i^2 (beta), and f_i^3 (gamma).

c) For $i = a, b, c$, choose between f_i^1, f_i^2, and f_i^3 using a decision rule D that gives $D(y_i^q) \in \{f_i^1, f_i^2, f_i^3\}$.

d) Update g_a, g_b, g_c by putting $(g_a^{q+1}, g_b^{q+1}, g_c^{q+1}) = (D(y_a^q), D(y_b^q), D(y_c^q))$.

The novelty of our method is situated at the decision rule D level. In [9], the rule D is based on the use of the Pearson system, in which one calculates the skewness and the kurtosis, and in [14], the rule D is the minimization of the Kolmogorov distance. The decision rule we propose here is based on kernel estimation, and the step c) becomes

i) For $i = a, b, c$, calculate

$$\hat{f}_i(y) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{y - y_j}{h_n}\right)$$

where (y_j) are in y_i^q, and K is the normal kernel $K(y) = (1/\sqrt{2\pi})e^{-(y^2/2)}$;

ii) take among f_i^1, f_i^2, and f_i^3 the density f_i^r such that $\|f_i^r - \hat{f}_i\|_\infty = \min_{1 \le t \le 3} \|f_i^t - \hat{f}_i\|_\infty$.

This procedure, which will be called ICE-KERNEL, turns out to perform better, at least in the setting of our experiments, than the Pearson system-based method (ICE-PEAR [9]).

TABLE III
B, G, N: BETA, GAMMA, AND NORMAL DISTRIBUTIONS, RESPECTIVELY, WITH THE CORRESPONDING PARAMETERS. MPM (Y^1): ERROR RATIO OBTAINED BY THE MPM METHOD USING THE ONLY PROBABILISTIC SENSOR Y^1. FUSION MPM: ERROR RATIO OBTAINED BY THE MPM METHOD AFTER THE DEMPSTER–SHAFER FUSION OF THE SENSORS Y^1 AND Y^2. ICE: THE CLASSICAL ICE ASSUMING ALL DISTRIBUTIONS NORMAL. τ: ERROR RATIO

Case	Class	a	b	c	τ(GB)	τ(GE)
1	True Laws	B(7.0, 7.0, 0.0, 20.0)	G(1.5, 2.0, 5.0)	N(2.0, 1.0)	3.63	1.86
2	ICE	N(9.8, 5.4)	N(7.8, 5.6)	N(1.5, 1.0)	4.54	2.23
3	ICE-PEAR	N(8.7, 5.8)	G(1.2, 2.6, 6.0)	N(2.4, 1.3)	9.93	5.69
4	ICE-KERNEL	B(4.3, 6.0, 1.8, 23.2)	G(1.4, 2.2, 4.7)	N(1.9, 0.9)	3.98	2.18

Remark 5.3: The ICE-GEMI algorithm can be used in the most "pessimistic" context, in which one has no training data. Of course, if one has a set of training data $(x_1, y_1), \ldots, (x_n, y_n)$, the data $x = (x_1, \ldots, x_n)$ can be used in different, and simpler, estimation procedures. In particular, if we have a programmed some version of ICE-GEMI, we can use it by removing the iterations and using x instead of x^q.

As an example, let us consider g_A, g_B, g_C, and g_D Normal, with variance 1 and means 0, 2, 4, and 1, respectively, whose parameters are estimated by classical ICE. Thus, concerning these densities, we consider a particular case, though a generalized ICE could also be applied. The forms (beta B, gamma G, or Normal N, see, for instance, [18] for different precise forms) and parameters of g_a, g_b, g_c are given in Table III, which also contains the Bayesian error ratio of GB and GE segmentations. Having the true distributions and the optimal segmentation results based on them, we then evaluate the parameter estimation effectiveness on the one hand, and the effectiveness of the estimated parameter based MPM segmentations, on the other hand.

Concerning the detection of the distribution forms, we note that ICE-KERNEL finds the right forms, and ICE-PEAR makes one mistake. It is interesting to note that ICE-based segmentation, which is necessarily based on normal distributions, here gives better results than the ICE-PEAR-based segmentation. This is not surprising because very badly estimated parameters can still give good segmentation results.

Finally, we note that the ICE-KERNEL-based fused MPM segmentation is close to the True Laws-based fused MPM segmentation.

C. Real Images Fusion

Let us consider a radar and an optical image, given on Fig. 4, of a same scene. We have four classes, and the presence of clouds in the SPOT image will be modeled by an evidential model. So we have $\Omega = \{a, b, c, d\}$, and the mass function concerning the evidential sensor will be defined on $\Omega^* = \{A, B, C, D, E\}$, with $A = \{a\}$, $B = \{b\}$, $C = \{c\}$, $D = \{d\}$, and $E = \{a, b, c, d\}$. Let us assume that all noise

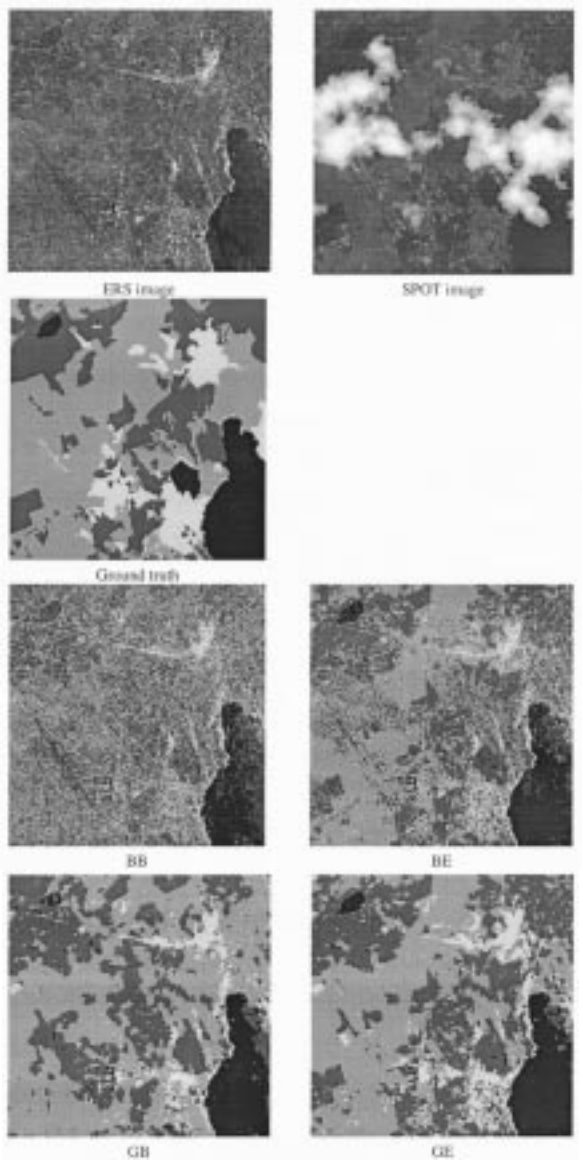

Fig. 4. Real images and different segmentation results.

distributions are Gaussian. Thus, we have nine means and nine variances to estimate. Concerning the Markov class field we take the same Potts model as in the previous section. Thus, the prior distribution is defined by the only real parameter α. We estimate all these parameters by a particular ICE. The parameters of the four Bayesian Gaussian densities and α are estimated from the Bayesian image $Y^1 = y^1$ and the five Gaussian densities corresponding to the evidential sensor are estimated from the evidential one $Y^2 = y^2$.

The results of estimation and rates of wrongly classified pixels are presented in Table IV, and the real images and different segmentation results are given on Fig. 4.

We have to remark that the results of the real image segmentation methods presented here may not be optimal. In fact, as the prior Markov field is quite rudimentary, the model considered is a rather simple one. However, our main goal was just to verify whether the general trends noticed in the simulation studies previously were preserved in an unsupervised segmentation of real

TABLE IV
ESTIMATES OF MEANS AND VARIANCES IN BAYESIAN AND EVIDENTIAL IMAGES (THE ESTIMATE OF THE PRIOR MARKOV FIELD IS $\hat{\alpha} = 0,59$) AND ERROR RATES OF UNSUPERVISED SEGMENTATION RESULTS

Class	*a*	*b*	*c*	*d*	
Mean and variance	(32,119)	(59,179)	(76,324)	(123,1820)	
Class	*A*	*B*	*C*	*D*	*E*
Mean and variance	(38,2)	(48,23)	(65,24)	(72,127)	(159,1691)
Algorithms	BB	BE	GB	GE	
Error rate	0.45	0.31	0.34	0.23	

TABLE V
ESTIMATES OF FORMS AND PARAMETERS IN BAYESIAN AND EVIDENTIAL IMAGES (THE ESTIMATE OF THE PRIOR MARKOV FIELD IS $\hat{\alpha} = 0,59$) AND RATES OF WRONGLY CLASSIFIED PIXELS IN UNSUPERVISED ALGORITHMS BASED ON GENERAL MIXTURES ESTIMATION

Class	*a*	*b*	*c*	*d*
Form and parameters	B(5.6,6.9,−4,77)	Γ(12.9,4.2,8)	Γ(8.7,6.2,22)	B(0.8,2.9,91,300)
Class	*A*	*B*	*C*	*D*
Form and parameters	B(3.6,3.5,33,42)	Γ(6.2,2.0,36)	N(64,34)	Γ(6.5,4.9,43)
Class	*E*			
Form and parameters	N(162,1700)			
Algorithms	BB	BE	GB	GE
Error rate	0.43	0.29	0.33	0.23

images. According to the results of Table IV, we may say that they are.

Concerning the possibly non-Gaussian forms of the noise densities, the results of generalized mixture estimation and the rates of wrongly classified pixels are given in Table V.

We can see that the generalized mixture estimation gives different forms which are not necessarily Gaussian. However, in the problem of interest, we notice that the error rates in Table V are not significantly different from the error rates in Table IV. Thus, at least in cases treated here, one may use classical Gaussian mixture estimation.

Let us briefly mention an another example, which is somewhat different from all examples studied in this section. We have always assumed having one Bayesian sensor and an evidential one. However, according to the Proposition 4.1, all sensors can be evidential in such a way that their fusion remains evidential. In particular, we can consider just one evidential sensor. We present in Fig. 5 a radar image of a chalky plateaux in southern France, which mainly contain pastures and woods. Because of the relief, some spots, which are in black in Fig. 5, do not produce any radar echo. We still consider the simple Ising model, in which the parameter is estimated from the ground truth by the stochastic gradient [42], which gives $\hat{\alpha} = 2.7$. We then consider two models. In the first one, we do not consider the ignorance spots and assume that we have two classes. The two noise densities, assumed Gaussian, are then estimated by the SEM [22], and the real image is segmented into two classes by the classical MPM. The error ration is 21.7%. In the second model, we consider three classes and, as before, we estimate three Gaussian noise densities f_1, f_2, and f_3 with the SEM. The density f_3 corresponding to the unobserved spots, according to (4.3) in Proposition 4.1, the MPM is applied using the two "artificial" noise terms $h_1 = f_1 + f_3$ and $h_2 = f_2 + f_3$. The new error ratio is 16.6%, which improves the quality of the segmentation.

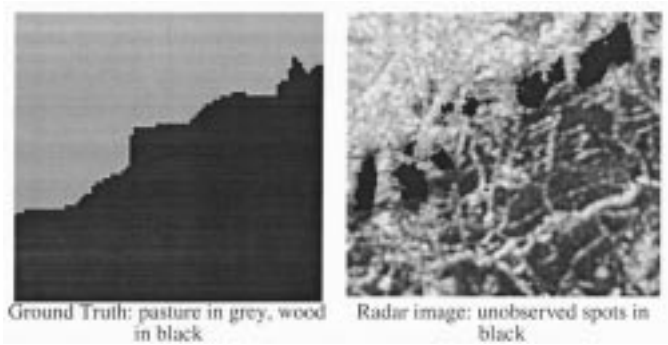

Fig. 5. Radar images of two classes "pasture" and "wood" and the ground truth.

VI. Conclusions

We proposed in this paper a new HMM, extending the classical Bayesian HMM, allowing one to segment multisensor images. Its main novelty was to show how a link with the theory of evidence can be advantageously made. In fact, in some real situations, as in the case of cloud presence in optical satellite images, modeling the resulting uncertainty with evidential mass functions is quite straightforward and improves the segmentation results. Furthermore, different simulation studies showed that there exist other situations in which taking into account an additional evidential sensor can improve the segmentation performed from the sole Bayesian sensor. Concerning the possibly unsupervised segmentation in the Markovian context considered, we have proposed an original method of generalized mixture estimation. Finally, two examples of real image segmentations showed the favorable behavior of the proposed methods in real situations.

We have used HMFs in this paper, but the use of other Markov models for prior distribution, like Markov chains [14], [30], Markov trees [20], or still more general Markov networks [33] could be considered. As mentioned in [26], the Markov trees model seems particularly well suited to the situations when we have to deal with evidential priors.

As perspectives for further work, we may view two directions. The first is to extend different Bayesian Markov models toward "correlated" sensors. Such extensions, which would relax the hypothesis 3) of Section II, have been briefly discussed in [29], which extends the classical probabilistic case studied in [27]. The second is to try to relax, in the evidential sensor context, the hypothesis 1) of Section II. We have recently proposed a solution of this problem in the classical probabilistic context by introducing "pairwise" Markov fields [28]. Such models, in which the couple (classes, observations) is a Markov field, are more general than the classical HMFs. In fact, the class field distribution is not necessarily a Markov distribution. So, some "evidential" pairwise Markov fields models could possibly be studied to take into account some spatial dependence of observations conditionally on classes.

References

[1] A. Appriou, "Probabilités et incertitude en fusion de données multisenseurs," *Rev. Sci. Tech. Déf.*, no. 11, pp. 27–40, 1991.

[2] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc.*, vol. 48, pp. 259–302, 1986.

[3] A. Bendjebbour and W. Pieczynski, "Segmentation d'images multisenseur par fusion évidentielle dans un contexte markovien," *Trait. 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 Recognit. Lett.*, no. 17, pp. 905–919, 1996.

[5] J. M. Boucher and P. Lena, "Unsupervised Bayesian classification, application to the forest of Paimpont (Brittany)," *Photo Interprétation*, vol. 32, no. 1994/4, 1995/1–2, pp. 79–81, 1995.

[6] R. Chellapa and A. Jain, Eds., *Markov Random Fields, Theory and Application*. San Diego, CA: Academic, 1993.

[7] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognit.*, vol. 23, no. 6, pp. 747–761, 1989.

[8] N. A. C. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993.

[9] Y. Delignon, A. Marzouki, and W. Pieczynski, "Estimation of generalized mixture and its application in image segmentation," *IEEE Trans. Image Processing*, vol. 6, pp. 1364–1375, Oct. 1997.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 39, pp. 1–38, 1977.

[11] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster–Shafer theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, pp. 804–813, May 1995.

[12] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 39–55, jan. 1987.

[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, June 1984.

[14] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 465–475, May 1997.

[15] J. Guan and D. A. Bell, *Evidence Theory and Its Applications*. Amsterdam, The Netherlands: North-Holland, 1991.

[16] X. Guyon, "Random fields on network. Modeling, statistics, and applications," in *Probability and its Applications*. New York: Springer-Verlag, 1995.

[17] F. Janez and A. Appriou, "Théorie de l'évidence et cadres de discernements non exhaustifs," *Trait. Signal*, vol. 13, no. 3, pp. 237–250, 1996.

[18] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*. New York: Wiley, 1970, vol. 1.

[19] Z. Kato, J. Zeroubia, and M. Berthod, "Unsupervised parallel image classification using Markovian models," *Pattern Recognit.*, vol. 32, pp. 591–604, 1999.

[20] J.-M. Laferté, P. Pérez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Trans. Image Processing*, vol. 9, pp. 390–404, Mar. 2000.

[21] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 799–813, Aug. 1989.

[22] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 618–633, May 1993.

[23] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Assoc.*, vol. 82, pp. 76–89, 1987.

[24] M. Mignotte, C. Collet, P. Pérez, and P. Bouthémy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Processing*, vol. 9, pp. 1216–1231, July 2000.

[25] W. Pieczynski, "Statistical image segmentation," *Mach. Graph. Vis.*, vol. 1, no. 1/2, pp. 261–268, 1992.

[26] ——, "Hidden evidential Markov trees and image segmentation," in *Proc. Int. Conf. Image Processing'99*, Kobe, Japan, 1999.

[27] W. Pieczynski, J. Bouvrais, and C. Michel, "Estimation of generalized mixture in the case of correlated sensors," *IEEE Trans. Image Processing*, vol. 9, pp. 308–311, Feb. 2000.

[28] W. Pieczynski and A.-N. Tebbache, "Pairwise Markov random fields and segmentation of textured images," *Mach. Graph. Vis.*, vol. 9, no. 3, pp. 705–718, 2000.

[29] W. Pieczynski, "Unsupervised Dempster–Shafer fusion of dependent sensors," in *Proc. IEEE Southwest Symp. Image Analysis and Interpretation (SSIAI'2000)*, Austin, TX, Apr. 2–4, 2000, pp. 247–251.

[30] W. Qian and D. M. Titterington, "On the use of Gibbs Markov chain models in the analysis of images based on second-order pairwise interactive distributions," *J. Appl. Statist.*, vol. 16, no. 2, pp. 267–281, 1989.
[31] P. Rostaing, J.-N. Provost, and C. Collet, "Unsupervised multispectral image segmentation using generalized Gaussian noise model," in *Proc. Energy Minimization Methods in Computer Vision and Pattern Recognition'99*, York, U.K., July 1999.
[32] F. Salzenstein and W. Pieczynski, "Parameter estimation in hidden fuzzy Markov random fields and image segmentation," *Graphical Models Image Process.*, vol. 59, no. 4, pp. 205–230, 1997.
[33] R. Serfozo, *Introduction to Stochastic Networks*. New York: Springer-Verlag, 1999.
[34] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[35] P. Smets, "Belief functions versus probability functions," in *Uncertainty and Intelligent Systems*, B. Bouchon, L. Saitta, and R. R. Yager, Eds. Berlin, Germany: Springer-Verlag, 1988, pp. 17–24.
[36] ——, "Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem," *Int. J. Approx. Reas.*, vol. 9, pp. 1–35, 1993.
[37] ——, "Belief functions (with discussion)," in *Non-Standard Logics for Automated Reasoning*, P. Smets, A. Mamdani, D. Dubois, and H. Prade, Eds. New York: Academic, 1988, pp. 252–286.
[38] ——, "Data fusion in the transferable belief model," in *Proc. 3rd Int. Conf. Information Fusion, FUSION 2000*, vol. 1, Paris, France, July 10–13, 2000, pp. PS-21–PS-33.
[39] F. Tupin, I. Bloch, and H. Maître, "A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors," *IEEE Trans. Geosci Remote Sensing*, vol. 37, pp. 1327–1343, May 1999.
[40] R. R. Yager, J. Kacprzyk, and M. Fedrizzi, *Advances in the Dempster–Shafer Theory of Evidence*. New York: Wiley, 1994.
[41] T. Yamazaki and D. Gingras, "Image classification using spectral and spatial information based on MRF models," *IEEE Trans. Image Processing*, vol. 4, pp. 1333–1339, Sept. 1995.
[42] L. Younes, "Estimation and annealing for Gibbsian fields," *Ann. Inst. Henri Poincaré*, vol. 24, pp. 269–294, 1988.
[43] ——, "Parametric inference for imperfectly observed Gibbsian fields," *Probability Theory Related Fields 82*, pp. 625–645, 1989.
[44] J. Zhang, J. W. Modestino, and D. A. Langan, "Maximum likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 404–420, Apr. 1994.

Azzedine Bendjebbour was born in Algeria in 1969. He received the Ph.D. degree from the Université Pierre et Marie Curie, Paris, France, in 2000.

His research interests include mathematical statistics, stochastic processes, and theory of evidence.

Yves Delignon received the Ph.D. degree in the statistics of radar images from the Image Processing Department, Ecole Nationale Supérieure des Télécommunications de Bretagne, Bretagne, France, in 1993.

He is currently an Assistant Professor with Image and Signal Processing Deaprtment, Ecole Nouvelle d'Ingénieurs en Communication, Lille, France. From 1990 to 1993, he was with the Image Processing Department, Ecole Nationale Supérieure des Télécommunications de Bretagne. He is currently in charge of teaching on the mobile communications systems at the Ecole Nouvelle d'Ingénieurs en Communication. His present research activities are the statistical modeling and the signal processing.

Laurent Fouque was born in France in 1973. He received the Diplôme d'Etude Approfondies degree from the Université Pierre et Marie Curie, Paris, France, in 1997, and is currently pursuing the Ph.D. degree at the Institut National des Télécommunications, Evry, France, and Office National d'Études et Recherches Aérospatiales, Châtillon, France.

His research interests include statistical pattern recognition, multisensor classification, theory of evidence, and satellite imaging.

Vincent Samson was born in Caen, France, in 1976. He received the degree from the Ecole Nationale de la Statistique et de l'Analyze de l'Information, Paris, France, in 1999, and the M.Sc. degree in applied mathematics from the University of Rennes, Rennes, France. He is currently pursuing the Ph.D. degree in physics in the Image Processing Research Unit of the Office National d'Etudes et de Recherches Aérospatiales in collaboration with the Laboratoire des Signaux et Systèmes, Châtillon, France.

In 1999, he spent five months as a Trainee with the Institut National des Télécommunications, Evry, France, within the Statistical Image Processing Group.

Wojciech Pieczynski received the Doctorat d'État degree in mathematical statistics from the Université de Paris VI, Paris, France, in 1986.

He is currently Professor and Head of Image and Optimization Group, Institut National des Télécommunications, Evry, France. His research interests include mathematical statistics, statistical image processing, and theory of evidence.