GRAPHICAL MODELS AND IMAGE PROCESSING
Vol. 59, No. 4, July, pp. 205–220, 1997
ARTICLE NO. IP970431

Parameter Estimation in Hidden Fuzzy Markov Random Fields and Image Segmentation

Fabien Salzenstein and Wojciech Pieczynski

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

Received February 15, 1996; revised March 27, 1997; accepted April 9, 1997

This paper proposes a new unsupervised fuzzy Bayesian image segmentation method using a recent model using hidden fuzzy Markov fields. The originality of this model is to use Dirac and Lebesgue measures simultaneously at the class field level, which allows the coexistence of hard and fuzzy pixels in a same picture. We propose to solve the main problem of parameter estimation by using of a recent general method of estimation in the case of hidden data, called iterative conditional estimation (ICE), which has been successfully applied in classical segmentation based on hidden Markov fields. The first part of our work involves estimating the parameters defining the Markovian distribution of the noise-free fuzzy picture. We then combine this algorithm with the ICE method in order to estimate all the parameters of the fuzzy picture corrupted with noise. Last, we combine the parameter estimation step with two segmentation methods, resulting in two unsupervised statistical fuzzy segmentation methods. The efficiency of the proposed methods is tested numerically on synthetic images and a fuzzy segmentation of a real image of clouds is studied. © 1997 Academic Press

1. INTRODUCTION

This work addresses fuzzy statistical unsupervised image segmentation. Beyond probabilistic considerations, let us specify the interest in fuzzy segmentation in some real situations. Let us consider the problem of segmenting a satellite image into two classes: "houses" and "trees." There may be some pixels with only houses and others with only trees, but there may also be many pixels, as in suburbs, in which houses and trees are simultaneously present. Thus we have two hard classes, say 0 and 1, and a fuzzy class specified by $\varepsilon \in]0,1[$, which can be seen as the proportion of the area of class 1. Such a situation is intrinsically "fuzzy." Let us now consider the problem of segmenting into two classes a satellite image of a region containing lakes and forest. If the boundary pixels can be considered as negligible, each pixel clearly is "forest" or "lake." Such a situation is intrinsically "hard." In both cases the classes are observed with noise and the segmentation problem is to decide, from the observed image, in which class each pixel lies. In the first case we speak of fuzzy segmentation, and in the second case of hard segmentation. As we can see, fuzzy and hard segmentations are not competing but correspond to two different situations. Now, if we wish to use some statistical method we have to introduce random variables and probability distributions. We insist that from the viewpoint we adopt there is no connection between fuzziness and the stochastic modeling that can be used, although as suggested by some authors [42], probability measures can be considered as modeling fuzziness. The aim of this work is to propose a Markovian-model-based unsupervised method of satistical fuzzy segmentation that is able to cope with situations such as those in the first example above.

Thus fuzzy segmentation of images consists in allowing each pixel to belong to numerous classes simultaneously. Let $\Omega = \{\omega_1, \ldots, \omega_k\}$ be the set of classes. The problem is to associate to each pixel a vector $(x_1, \ldots, x_k) \in [0,1]^k$ with $x_1 + \cdots + x_k = 1$. Classical, or "hard," segmentation then appears as a particular case: all x_i are null except one, which is 1. This generalization turns out to be very pertinent in varied situations. Following Kent and Mardia [23], there are three ways of using fuzzy segmentations:

(i) One may use a visual representation of fuzzy reality

(ii) Considering that pixels have unit area and $x_{s,i}$ is the proportion of the class ω_i at the pixel s, the total area of the class ω_i is the sum of $x_{s,i}$

(iii) It is always possible to harden a fuzzy partition by choosing, for each pixel s, the class maximizing $x_{s,1}, \ldots, x_{s,k}$.

Our work basically uses the second viewpoint, with the main concern being to find the hard pixels and the fuzzy ones again following unsupervised segmentation. As an example, let us consider the problem of segmentation of a satellite image into two classes: forest and water. Obviously there may be pixels containing simultaneously some trees and some water and in this case the fuzzy segmentation gives in each pixel the proportion of forest and water.

1077-3169/97 $25.00
Copyright © 1997 by Academic Press
All rights of reproduction in any form reserved.

Numerous techniques for such fuzzy partitioning have been proposed [3, 4, 15, 17, 20] and an overview of the different methods can be found in [27].

On the other hand, statistical methods of segmentation, especially those using hidden Markov models, can turn out to be of exceptional efficiency in several situations [2, 5, 6, 9–12, 14, 16, 18, 21, 22, 24–26, 28, 29, 32, 36, 39, 41].

Finally, fuzzy and statistical aspects of methods can be merged, resulting in fuzzy statistical segmentations [7, 8, 23, 37]. We insist on the fact that by merging these two aspects of things one obtains an original modeling method, different from both probabilistic and fuzzy modeling. Indeed, fuzziness models the imprecision and probability models the uncertaintly. Roughly speaking, in the first case one can clearly see the pixel but it is impossible to clearly determine what class it belongs to, and, in the second case, the pixel clearly belongs to one class but cannot be clearly seen. When merging the two approaches we are faced with pixels which cannot be clearly seen and may not clearly belong to any class. See also the remark at the end of Section 3.

The aim of our paper is to present a statistical fuzzy segmentation method which may be considered a generalization of the well-known Markovian-model-based algorithms.

The unsupervised approach we present is based on a recent models using hidden fuzzy Markov random fields. In order to solve the main problem of parameter estimation, we propose using a recent general method of estimation in the case of hidden data, called iterative conditional estimation (ICE [29]). ICE is an iterative method which has been successfully applied in classical hidden Markov fields based segmentations [5, 6, 31] and the same ICE method can be used as the previous parameter estimation step in local Bayesian unsupervised segmentations [28].

As usual, we consider two random fields $X = (X_s)_{s\in S}$ and $Y = (Y_s)_{s\in S}$. The image to be segmented is a realization $Y = y$ of Y and the desired picture is the realization $X = x$ of the field X. So the values of Y are real. Let us consider the case of two classes. In the classical case, which will be called *hard* in what follows, the X_s take their values into $\Omega_h = \{0,1\}$, where the numbers 0 and 1 correspond to the hard classes (for instance the classes pure forest and pure city). As mentioned above, in the fuzzy model we take $\Omega_f = [0,1]$, where the numbers 0 and 1 correspond to the hard classes and $]0,1[$ to the fuzzy ones. Otherwise, if $X_s = x_s \in]0,1[$ then x_s indicates the proportion of the class 1 in the value of the pixel, and so $1 - x_s$ is the proportion of the class 0. The statistical approach requires a definition of priors, which is a probability distribution on $\Omega_f = [0,1]$. The originality of the model proposed in [7] is that the distribution of each X_s is given by a density h with respect to the measure $\nu = \delta_0 + \delta_1 + \mu$, which includes a hard component (Dirac functions δ_0, δ_1 on $\{0,1\}$) and a fuzzy one (the Lebesque measure μ on $]0,1[$). Thus, the probability of having hard classes can be positive, which is in harmony with the intuitive feeling we have about the image of classes. In contrast, when the pixel is fuzzy, the proportion of a given class varies continuously and thus its probability distribution is given by a density with respect to the Lebesgue measure. This model, which is of the pixel-by-pixel kind, can be refined in order to take into account the information lying in a context of small size [8]. Using such local models, it is possible to devise unsupervised local methods of fuzzy segmentation, the principle of which is the same as in local hard segmentation methods [6, 18, 26, 28]. Finally, the local model has been recently generalized by introducing the fuzzy Markov random fields described in [30]. However, we insist that local methods display some advantages in several situations [6, 28]. As in the hard case, the latter Markovian model allows one to take the entire information into account and, at each pixel, the distribution of each X_s is still given by a density h with respect to the measure $\nu = \delta_0 + \delta_1 + \mu$. Such a fuzzy Markov random field is then classically degraded by Gaussian noise and several segmentation methods can be considered [30].

The aim of our work is to make these methods unsupervised. Thus the main problem is to estimate all parameters of the model from Y, the noisy version of X. We propose the using the iterative conditional estimation method (ICE), which seems well adapted to the model considered. In fact, the principle of ICE does not refer to the likelihood, a notion which is difficult to handle in the context of our study. Let us note that when local unsupervised segmentation is concerned, ICE and expectation-maximization (EM [13, 33]) give comparable results [28]. However, in the context of local fuzzy unsupervised segmentation, there exist some situations in which ICE is preferable to EM [8].

Let us remark that fuzzy statistical classification using hidden fuzzy Markovian random fields has already been proposed in [23]. The difference with our approach is that in the model proposed in [23] a positive probability of having a hard pixel cannot be obtained, unless one makes a certain parameter tend to infinity. This can be seen as a drawback in situations where there clearly exist a positive probability of having a hard class and a positive probability of having a fuzzy class at a given pixel.

The organization of the paper is as follows:

In the next section we review modeling by hidden fuzzy Markov fields, as recently proposed in [30]. Section Three is devoted to the ICE, a recent general method of estimation in the case of hidden data. In Section Four we briefly recall the principle behind Bayesian segmentation and describe its mechanism in the context of the hidden fuzzy Markov fields model. Two segmentation methods, which in connection with ICE become unsupervised, are described. Computer simulation results for synthetic images and a fuzzy segmentation of a real image of clouds are presented in Section Five and Section Six contains the conclusion.

2. HIDDEN FUZZY MARKOV FIELDS

We briefly present below the recently proposed model of hidden fuzzy Markov fields (HFMF) and specify how it generalizes classical hidden hard Markov fields (HHMF). The two models are very similar and very different at the same time. They are very close in that the densities defining their distributions are of the same form; thus different computations are nearly the same. The are very different in that these densities are with respect to two different measures. Thus our general presentation is rather brief, although, we better develop different calculations in the case of a particular model used for simulation in Section 5.

2.1. *Distribution of X*

Let us consider the classical case of two classes $\Omega = \{0,1\}$, to be called hard in what follows. If X is a Markovian field with respect to a neighborhood V, its distribution is given by

$$P[X = x] = h(x) = ce^{-U(x)}, \tag{1}$$

where U, called energy, is a sum of functions defined on cliques, a clique being either a singleton or a set of pixels that are neighbors with respect to V. Let us consider the stationary case, i.e., the case where the functions defying U depend only on the shape of cliques and do not depend on their position in the set of pixels. Thus, if C is a clique of a given shape and $n = \mathrm{Card}(C)$, the associated function φ_C is a function from $\Omega^n = \{0,1\}^n$ into R. On the other hand, for each pixel s, the distribution of X_s is a distribution on $\Omega = \{0,1\}$. The problem is to generalize this model in such a way that for each pixel s, the distribution of X_s is a distribution on $\Omega_f = [0,1]$ given by a density with respect to the measure

$$\nu = \delta_0 + \delta_1 + \mu, \tag{2}$$

where δ_0, δ_1 are the Dirac measures on $\{0,1\}$, and μ is the Lebesgue measure on $[0,1]$.

In what follows N will designate the number of pixels.

Let us consider the function defined on $\Omega_f^N = [0,1]^N$ by $h_f(x) = ce^{-U_f(x)}$. The function U_f is of the same shape as the function U with the following difference: for C a clique with $n = \mathrm{Card}(C)$, the associated $\varphi_{f,C}$ is a function defined on $\Omega_f^n = [0,1]^n$ instead of $\Omega^n = \{0,1\}^n$. If we consider that $h_f(x) = ce^{-U_f(x)}$ is the density of P_X with respect to the measure ν^N, i.e.,

$$P_X = h_f \nu^N, \tag{3}$$

then it is possible to show, exactly as in the hard case, that X is Markovian with respect to V. Furthermore, the distribution of X_s is a distribution on $\Omega_f = [0,1]$ given by a density with respect to the measure $\nu = \delta_0 + \delta_1 + \mu$.

When N increases, the measure ν^N becomes very difficult to handle. However, when we deal with Markovian fields, all we need is to be able to compute the distribution of each X_s conditional to $(X_t)_{t\in V_s, t\neq s}$. As we will see in Section 5, the computation of these distributions is only slightly more complicated than in the hard case.

Hard Markov fields can be seen as particular cases of fuzzy Markov fields in the following sense: the distribution of each hard Markov field is a limit, as some parameter tends to infinity, of distributions of a family of fuzzy Markov fields. To be more precise, let us consider a hard Markov field given by the family of functions φ_C and let us define $\varphi_{f,C}^{\lambda}$ by

$$\varphi_{f,C}^{\lambda}(x_C) = \begin{cases} \varphi_C(x_C) & \text{if } x_C \text{ hard} \\ \lambda & \text{if } x_C \text{ fuzzy,} \end{cases} \tag{4}$$

where x_C hard means that all components $x_s, s \in C$ are hard and x_C fuzzy means that at least one of these components is fuzzy.

Let us consider two sets $E_f \subset [0,1]^N$, $E_h \subset [0,1]^N$ (which are fuzzy and hard images respectively),

$$[x = (x_1, \ldots, x_N) \in E_f] \Leftrightarrow [\exists 1 \leq i \leq N \text{ such that } x_i \in]0,1[] \tag{5}$$

$$[x = (x_1, \ldots, x_N) \in E_h] \Leftrightarrow [\forall 1 \leq i \leq N, x_i \in \{0,1\}]. \tag{6}$$

We shall insist that an image is fuzzy when one at least pixel is fuzzy (it can contain hard pixels).

We give below a result which slightly generalizes two propositions presented in [29].

PROPOSITION. *There exists a positive constant A such that*

$$P_X[E_h] \geq 1 - Ae^{-\lambda} \tag{7}$$

and as a consequence $\lim_{\lambda\to+\infty} P_X[E_h] = 1$ *and* $\lim_{\lambda\to+\infty} P_X[E_f] = 0$.

Proof. Let $x \in E_f$. The density of (3) is written

$$h_f^{\lambda}(x) = c(\lambda)e^{-U_f^{\lambda}(x)}. \tag{8}$$

Let us show that $c(\lambda)$ is bounded by a constant not depending on λ. For each $x_0 \in E_h$ the number $h_f^{\lambda}(x_0)$ is the probability of x_0; thus $h_f^{\lambda}(x_0) \leq 1$. Furthermore, λ does not intervene in the energy defining $h_f^{\lambda}(x_0)$; thus $h_f^{\lambda}(x_0) = c(\lambda)e^{-U(x_0)} \leq 1$. Putting $M = e^{U(x_0)}$, we have

$$c(\lambda) \leq M. \tag{9}$$

With C the set of cliques, let us consider, for each $C' \subset C$ and $y \in E_h$, the partial sums $\sum_{c\in C'} \varphi_c(y_c)$. As these sums are of finite number, there exists $a \in R$ such that $\sum_{c\in C'} \varphi_c(y_c) \geq a$ for all $C' \subset C$ and $y \in E_h$. Let us return to $x \in E_f$ and $\sum_{c\in C} \varphi^\lambda_{f,c}(x_c)$. Let $C' \subset C$ be the subset of cliques on which all components of x_c are hard and let $C'' \subset C$ be the subset of cliques on which one at least component of x_c is fuzzy. As $\sum_{c\in C''} \varphi^\lambda_{f,c}(x_c) \geq \lambda$, by construction of $\varphi^\lambda_{f,c}(x_c)$ we can write

$$U^\lambda_f(x) = \sum_{c\in C} \varphi^\lambda_{f,c}(x_c) = \sum_{c\in C'} \varphi_c(x_c) + \sum_{c\in C''} \varphi^\lambda_{f,c}(x_c) \geq a + \lambda \tag{10}$$

and finally

$$h^\lambda_f(x) = c(\lambda)e^{-U^\lambda_f(x)} \leq Me^{-(a+\lambda)}. \tag{11}$$

As h^λ_f is the density of P_X with respect to ν^N we have $P_X[E_f] \leq \nu^N[E_f]Me^{-(a+\lambda)}$. On the other hand, $\nu^N[E_f] = \nu^N[[0,1]^N] - \nu^N[E_d] = 3^N - 2^N$. Finally, $P_X[E_f] \leq (3^N - 2^N)Me^{-(a+\lambda)}$, which completes the proof with $A = (3^N - 2^N)Me^{-a}$.

2.2. *Distribution of* (X, Y)

We have now to define the distribution of (X, Y). The distribution of X having been defined above, we need only define distributions of Y conditional on X. As is usual made in the hard case we will assume:

(i) The random variables (Y_s) are independent conditionally on X.

(ii) The distribution of each Y_s conditional on X is equal to its distribution conditional on X_s.

Distributions of Y conditional on X are then defined by distributions of Y_s conditional on X_s. Assuming that distributions of (X_s, Y_s) are independent of s and denoting by $N(m, \sigma^2)$ the normal distribution of mean m and variance σ^2, we will take for the distribution of Y_s conditional on $X_s = x_s \in [0,1]$

$$N((1 - x_s)m_0 + x_s m_1, (1 - x_s)\sigma_0^2 + x_s\sigma_1^2), \tag{12}$$

where $m_0, m_1, \sigma_0^2, \sigma_1^2$ are given parameters. Thus $m_0, m_1, \sigma_0^2, \sigma_1^2$ define all distributions of Y conditional on X.

For $m_0, m_1, \sigma_0^2, \sigma_1^2$ fixed, let us denote by ψ_{x_s} the Gaussian density defined above. The density ψ of the distribution of (X, Y) with respect to $\nu^N \otimes \mu^N$ (ν being the measure on [0,1] defined by (2), μ the Lebesgue measure on R, and N the number of pixels) is then given by

$$\psi(x, y) = ke^{-U_f(x)} \prod_{s\in S} \psi_{x_s}(y_s). \tag{13}$$

2.3. *Distribution of X a Posteriori*

Putting $\prod_{s\in S} \psi_{x_s}(y_s) = e^{\sum_{s\in S} \mathrm{Log}\, \psi_{x_s}(y_s)} = e^{-V_x(y)}$ and

$$W_y(x) = U_f(x) + V_x(y),$$

(13) is written

$$\psi(x, y) = ke^{-W_y(x)}. \tag{14}$$

The density of the distribution of X a posteriori (i.e., conditional on $Y = y$) with respect to ν^N is thus given by

$$\psi^y(x) = \frac{ke^{-W_y(x)}}{\int_{[0,1]^N} ke^{-W_y(x)}\, d\nu^N(x)}, \tag{15}$$

which can be written

$$\psi^y(x) = k(y)e^{-W_y(x)} = k(y)e^{-(U_f(x)+V_x(y))}. \tag{16}$$

Thus the energy in (16) is of the same kind as that in (1). The additional term $V_x(y)$ is

$$V_x(y) = -\sum_{s\in S} \mathrm{Log}\, \psi_{x_s}(y_s), \tag{17}$$

where ψ_{x_s} is Gaussian with mean $(1 - x_s)m_0 + x_s m_1$ and variance $(1 - x_s)\sigma_0^2 + x_s\sigma_1^2$. As in the case of hard Markovian fields, the Markovian nature of the posterior distribution of X is thus preserved and one can use the Gibbs sampler in order to simulate its realizations.

The exact running of the Gibbs sampler is described in the case of a particular model used for experiments in Section 5. However, its principle remains valid for every hidden fuzzy Markovian field.

3. PARAMETER ESTIMATION USING ICE

3.1. *Definition*

Iterative conditional estimation is a general procedure for parameter estimation in the case of hidden data that we describe below in some detail.

The idea behind this procedure is the following: the complexity of the estimation problem is due to the absence of the observation of X. If X were observable, one could generally use some efficient parameter estimation procedure. In fact, if the estimation of θ from (X, Y) is impossible there is no sense in estimating it from the Y data alone. So let us suppose temporarily that X is observable and let us consider $\hat{\theta} = \hat{\theta}(X, Y)$, an estimator of the parameter θ.

As an estimator $\hat{\theta} = \hat{\theta}(X, Y)$ is a random variable whose construction does not use θ. In a general manner, if we want to approach a random variable Z by some function of a random variable W, the best approximation, where the squared error is concerned, is the conditional expectation. To be more precise, if we denote the conditional expectation by $E[Z/W]$ we have

$$E[Z - E[Z/W]]^2 = \min_{\varphi} E[Z - \varphi(W)]^2. \tag{18}$$

As we consider the problem of estimating θ using Y alone, we have to approach $\hat{\theta} = \hat{\theta}(X, Y)$ by a function of Y. ICE proposes using precisely the conditional expectation $E[\hat{\theta}(X, Y)/Y]$. The drawback is that this conditional expectation depends on θ, which leads to the following iterative procedure:

$$\begin{aligned} &\text{(i) initialization } \theta_0 \\ &\text{(ii) } \theta_{k+1} = E_{\theta_k}[\hat{\theta}/Y = y]. \end{aligned} \tag{19}$$

When $E_{\theta_k}[\hat{\theta}/Y = y]$ is not computable but samplings of X according to the distribution conditional on $Y = y$ are possible, one can use a stochastic approximation. In fact, the conditional expectation is the expectation according to the conditional distribution. Thus it can be approached, by virtue of the law of large numbers, by the empirical mean. After having sampled m realizations $x_1, \ldots, x_m$ of X according to its distribution conditioned on $Y = y$, we can consider

$$\theta_{k+1} = \frac{1}{m}[\hat{\theta}(x_1, y) + \cdots + \hat{\theta}(x_m, y)]. \tag{20}$$

As we will see, the calculation of $E[\hat{\theta}(X, Y)/Y]$ is not possible when considering the statistical fuzzy image segmentation based on the model described in the previous section. Thus we will have to use (20), which is feasible because of the possibility of simulating realizations of X according to its posterior distribution by Gibbs sampler.

Let us note that most of the recently proposed estimation methods are iterative: the next value of the parameter is computed from the current one and the data by the application of some criterion. In the case of the estimation–maximization (EM) algorithm, this criterion is the increase of the likelihood of the distribution of Y [13, 33].

3.2. *Parameter Estimation in FHMRF Using ICE*

Let us denote by α the set of parameters defining U_f which gives the distribution of X, and by $\beta = (\mu_0, \mu_1, \sigma_0, \sigma_1)$ parameters defining distributions of Y conditional on X. The parameter $\theta = (\alpha, \beta)$ defines the distribution of (X, Y) and the problem is to estimate it from Y. According to the previous section, we can use the ICE procedure once:

(i) we have an estimator $\hat{\theta} = \hat{\theta}(X, Y)$ of $\theta = (\alpha, \beta)$ from (X, Y)

(ii) we can perform, for each $\theta = (\alpha, \beta)$, simulations of realizations of X according to its posterior distribution.

Let us begin with the second point. According to the general hypotheses the distribution of X is a Gibbs distribution. The proof of this is exactly the same as in the hard case. It is than possible to use a fuzzy Gibbs sampler [30, 35], which is a simple adaptation to our model of the classical hard Gibbs sampler. The same procedure can be used to simulate realizations of X according to its posterior distribution.

In order to solve point (i) let us consider problems of estimating α and β separately. We take as estimator $\hat{\alpha} = \hat{\alpha}(X)$ of α (from X, the fuzzy noise-free field) the iterative procedure [1], which is an adaptation to the fuzzy case of the stochastic gradient algorithm [40],

$$\begin{aligned} &\text{(a) } \alpha_0, X = x_0 \text{ given} \\ &\text{(b) } \alpha_{n+1} = \alpha_n + \frac{c}{(n+1)}[U_f'(x_{n+1}) - U_f'(x_0)], \end{aligned} \tag{21}$$

where $U_f'(x)$ is the gradient of $U_f(x)$ with respect to α, and x_{n+1} is a realization of X simulated by the fuzzy Gibbs sampler, according to its prior distribution and using the current parameter α_n. In the hard case, a convenient choice of the parameter c ensures the convergence to the true values of the parameters [40].

We choose as estimator $\hat{\beta} = \hat{\beta}(X, Y)$ of β the empirical means and variances $(\hat{m}_i, \hat{\sigma}_i^2)$, which are defined from (X_{Q_i}, Y_{Q_i}) (with, for $i = 0, 1$, $Q_i = \{s \in S/X_s = i\}$) by

$$\hat{m}_i(X, Y) = \frac{\sum_{s \in Q_i} Y_s 1_{[X_s = i]}}{\sum_{s \in Q_i} 1_{[X_s = i]}} \tag{22}$$

$$\hat{\sigma}_i^2(X, Y) = \frac{\sum_{s \in Q_i} (Y_s - \hat{m}_i)^2 1_{[X_s = i]}}{\sum_{s \in Q_i} 1_{[X_s = i]}}. \tag{23}$$

Thus $\hat{\theta} = \hat{\theta}(X, Y)$ is defined by (21)–(23).

Finally, according to the ICE principle, the estimation of θ from Y alone is as follows:

(i) take $\theta^0 = (\alpha^0, \beta^0)$ as an initial value of θ

(ii) compute $\theta^{k+1} = (\alpha^{k+1}, \beta^{k+1})$ from $\theta^k = (\alpha^k, \beta^k)$ and $Y = y$ in the following way:

(1) Using the Gibbs sampler, simulate m realizations $x_1, x_2, \ldots, x_m$ of X according to the posterior distribution corresponding to $\theta^k = (\alpha^k, \beta^k)$ and $Y = y$.

(2) For each x_j estimate α by (21) (which requires r new simulations by the Gibbs sampler). Let $\alpha(x_1), \alpha(x_2), \ldots, \alpha(x_N)$ be the values so obtained.

(3) For each x_j estimate β (from x_j and $Y = y$) by formula (22) and (23). Let $\beta(x_1, y), \beta(x_2, y), \ldots, \beta(x_N, y)$ be the values so obtained.

(4) $\theta^{k+1} = (\alpha^{k+1}, \beta^{k+1})$ is given by

$$\alpha^{k+1} = \frac{1}{m}[\alpha(x_1) + \alpha(x_2) + \cdots + \alpha(x_m)] \tag{24}$$

$$\beta^{k+1} = \frac{1}{m}[\beta(x_1, y) + \beta(x_2, y) + \cdots + \beta(x_m, y)]. \tag{25}$$

(iii) if the sequence $\theta^k = (\alpha^k, \beta^k)$ approaches steady state, stop.

Remark 1. Let us note that the method ICE + stochastic gradient can be seen as a generalization to the noisy case of the stochastic gradient, which applies to the noise free case. In fact, when the noise vanishes, i.e., variances tend to zero, the m realizations $x_1, x_2, \ldots, x_m$ of X sampled using the Gipps sampler according to the posterior distribution corresponding to $\theta^k = (\alpha^k, \beta^k)$ and $Y = y$ (see (ii, 1) of the procedure above) are no longer stochastic realizations, but are all equal to x_0 (see (21, b)), i.e., the observed fuzzy image. Thus, if $m = 1$, one obtains exactly the Stochastic Gradient; i.e., the next value of the parameter is given by (21, b), and is the final value ($k = 1$). If m is superior to 1, there is very little difference: $\alpha(x_1), \alpha(x_2), \ldots, \alpha(x_m)$ are different estimates, with $x_1 = x_2 = \cdots = x_m = x_0$, of α by (21). As the procedure (21) stops when the sequence stabilizes, $\alpha(x_1), \alpha(x_2), \ldots, \alpha(x_m)$ change little and thus $\alpha^{k+1} = (1/m)[\alpha(x_1) + \alpha(x_2) + \cdots + \alpha(x_m)]$ differs little from each of them.

Thus ICE + stochastic gradient stays valid in the noise free case and should automatically "degenerate" into stochastic gradient. This is of interest in practical applications. For instance, if one wishes to classify different images using the estimated parameters, it is possible to assume that they are noisy, and if they are not, the ICE + stochastic gradient will automatically become the stochastic gradient.

4. UNSUPERVISED SEGMENTATION

Unsupervised segmentation is obtained by augmenting the ICE parameter estimation above by a method of segmentation. Probabilistic models allow the use of Bayesian methods of segmentation, whose general principle we briefly recall. One considers a function $L^*: [0,1]^N \times [0,1]^N \to R^+$, called a loss function. $L^*(x, x')$ models the severity of assuming that the real value is x', when it is x. The Bayesian strategy is then a strategy $\hat{s}_B$ which minimizes $E[L^*(X, \hat{s}(Y)]$, with respect to $\hat{s}$. Following the law of large numbers, $E[L^*(X, \hat{s}(Y)]$ can be seen as the average severity of the errors committed when using $\hat{s}$ on a long run. Thus $\hat{s}_B$ is a strategy which minimizes the latter mean, which is quite satisfying from an intuitive point of view. For instance, in the case of hard hidden Markov random fields, the MAP [16] is the Bayesian strategy corresponding to the loss function

$$L^*(x, x') = \begin{cases} 0 & \text{if } x = x' \\ 1 & \text{if } x \neq x' \end{cases}$$

and the MPM [25] is the Bayesian strategy corresponding to the loss function

$$L^*(x, x') = \sum_{s \in S} L(x_s, x'_s) \tag{26}$$

with

$$L(x_s, x'_s) = \begin{cases} 0 & \text{if } x_s = x'_s \\ 1 & \text{if } x_s \neq x'_s. \end{cases} \tag{27}$$

In this work we will consider L^* of the form given by (26). Thus L^* is defined once $L(x_s, x'_s)$ is defined, for each x_s, x' in [0,1]. In this case the Bayesian strategy is the strategy which minimizes $E[L(X_s, \hat{s}_s(Y)]$ for each pixel $s \in S$. In the hard case things are relatively simple: for k classes, L is given by a $k \times k$ matrix. In the fuzzy case concerning us, even in the case of two classes, things are more complicated. There are two different natures of pixels (hard and fuzzy) and L will depend on what we are looking for. For instance, if we strongly wish to detect the fuzzy class, $L(x_s, x'_s)$ will be sizeable on $]0,1[\times \{0,1\}$; if the detection of class 0 has grater importance that the detection of the class 1, we will have to consider $L(1, x'_s) \rangle L(0, x'_s)$ for $x'_s \in]0,1[$; and so on. Let us note that this presents a complication, but also a richness showing the flexibility of the Bayesian methods.

For a given L, the practical search for the Bayesian strategy $\hat{s}_B$ is the following. As

$$\begin{aligned} E[L^*(X, \hat{s}(Y))] &= E[E[L^*(X, \hat{s}(Y))/Y]] \\ &= \sum_{s \in S} E[E[L(X_s, \hat{s}_s(Y))/Y]], \end{aligned} \tag{28}$$

$\hat{s}_B$ will be the solution once $\hat{s}_B$ minimizes

$$E[L(X_s, \hat{s}_s(Y))/Y = y] \tag{29}$$

for each $Y = y$.

The calculus of (29) thus requires knowledge of the a posteriori distribution of each X_s. In the case concerning us, the latter distributions are given by densities $h_{s,y}$, where $Y = y$, with respect to ν. Let us assume temporarily that these distributions are known. (29) is then written

$$h_{s,y}(0)L(0, \hat{s}_s(y)) + h_{s,y}(1)L(1, \hat{s}_s(y)) + \int_0^1 h_{s,y}(t)L(t, \hat{s}_s(y))\, dt. \tag{30}$$

Let us consider

$$L(x_s, x'_s) = |x_s - x'|. \tag{31}$$

(30) then becomes

$$h_{s,y}(0)|\hat{s}_s(y)| + h_{s,y}(1)|1 - \hat{s}_s(y)| + \int_0^1 h_{s,y}(t)|t - \hat{s}_s(y)|\, dt. \tag{32}$$

Finally, segmentation is performed by attributing to each pixel $s \in S$ a number $\hat{s}_{B,s}(y) \in [0,1]$ which minimizes (32), the latter problem being solved numerically. In the following, this algorithm will be called ALG1.

The second segmentation algorithm that we will test in Section 5, and which was proposed in [7], is the following:

(i) choose from $\{0, 1, F\}$ (F for fuzzy) according to the classical Bayesian rule: the chosen element maximizes the probability $h_{s,y}(0)$, $h_{s,y}(1)$, $1 - h_{s,y}(0) - h_{s,y}(1)$.

(ii) if the chosen element is in $\{0,1\}$ stop, otherwise choose in $]0,1[$ the element which maximizes the restriction of $h_{s,y}$ to $]0,1[$.

This algorithm will be called ALG2.

As in the hard case, the densities $h_{s,y}$ cannot be calculated analytically and we have to estimate them in a previous step. This estimation is performed from realizations of X simulated according to its posterior distribution by a Gibbs sampler, in a way analogous to that followed in the hard MPM [25].

Finally, all parameters of the model described in Section 2 being known, the segmentation step itself becomes:

(i) Estimate $h_{s,y}$ for each $s \in S$

(ii) perform the segmentation with ALG1 or ALG2.

Let us notice that the method exposed in [7] is different in that we deal with blind, say pixel by pixel, fuzzy segmentation. However, once $h_{s,y}$ known, the problem of classifying s is strictly the same. Two other fuzzy Bayesian segmentations can be found in [8].

Remark 2. Let us specify how this method differs from the method of fuzzy segmentation briefly suggested in a recent paper by Zhang, Modestino, and Langan [42]. The authors propose a family of unsupervised hard segmentation methods applicable in complex situations; in particular, it is possible to take texture into account. They use the hidden Markov field model and propose estimating the parameters with an original variant of the iterative expectation–maximization (EM) algorithm. At each iteration EM gives the posterior marginals, which are probability measures $p_s = (p_s^1, \ldots, p_s^k)$ on the set of hard classes $\Omega = \{\omega_1, \ldots, \omega_k\}$. These probabilities are called "indicator vectors" and the authors briefly indicate that this can be considered a "soft" segmentation, as opposed to the "hard" segmentation obtained by their maximization. In particular, such a soft segmentation can be considered at the final iteration of EM. As the function $\omega_i \to p_s^i = P^{Y=y}[X_s = \omega_i]$ takes its values in [0,1], it can also be seen as a fuzzy set. Finally, the probability $p_s = (p_s^1, \ldots, p_s^k)$ considered at each pixel can be seen as a fuzzy segmentation. Thus fuzziness is obtained as an interpretation of a probability measure. This is different from the viewpoint of this paper, in that we consider fuzzy and probabilistic aspects of things simultaneously. A pixel can be hard or fuzzy, outside any probabilistic consideration. Wishing to use statistical processing, we then define a probability distribution on fuzzy classes, which is different from interpreting a probability measure on hard classes as their fuzziness. This conceptual difference results in concrete differences in the behavior of the methods. For instance, taking two classes $\Omega = \{\omega_1, \omega_2\}$ considered in this paper and a fuzzy image containing about 60% of fuzzy pixels (Im 8 of the next section), we can notice that the method we propose releases their approximate proportion (Table 3 of the next section), even in rather strongly noisy cases (Im 9 and Im 14 of the next section). The hard version of the Zhang, Modestino, and Langan algorithm applied to Im 9 and Im 14 would give no fuzzy pixels and, what is more, the soft version of this algorithm would give all pixels fuzzy. The latter is due to the fact that all realizations of a hard Markovian field have a strictly positive probability and, as a consequence, in the proposed fuzzy segmentation $p_s = (p_s^1, p_s^2)$ we have $p_s^1 \neq 0$ and $p_s^2 \neq 0$ for all pixels.

Thus the two methods are conceptually quite different. However, the indicator vector approach could be followed by some specific transformation and could conceivably give, in practice, results comparable to those obtained with the method we propose in some situations. For instance, some of the pixels obtained with $p_s = (p_s^1, p_s^2)$ above could be hardened by considering that a pixel is hard when its probability is superior to a given threshold.

5. EXPERIMENTS

5.1. *Model Used*

We consider hidden fuzzy Markov fields with respect to eight nearest neighbors and we do not take into account cliques whose cardinal is superior to two. We have then five kinds of cliques: singletons, "horizontal" neighbors, "vertical" neighbors, "north–east" (which are the same as "south–west") neighbors, and "north–west" (which are the same as "south–east") neighbors. We have to define five kinds of φ which will define the energy of the fuzzy Markov field X.

Concerning singletons we will consider

$$\varphi(x_s) = \begin{cases} \eta_0 & \text{if } x_s = 0 \\ \eta_1 & \text{if } x_s = 1 \end{cases} \quad \text{and} \quad \varphi(x_s) = \lambda \text{ if } x_s \in \,]0,1[. \tag{33}$$

Concerning pairs of neighbors, we will consider two cases: both are hard or at least one is fuzzy. Taking horizontal neighbors we consider

For $(x_s, x_t) \in \{0,1\}^2$ (both pixels hard):

$$\varphi_{\text{hor}}(x_s, x_t) = \begin{cases} -\alpha^{\text{h}}_{\text{hor}} & \text{if } x_s = x_t \\ \alpha^{\text{h}}_{\text{hor}} & \text{if } x_s \neq x_t. \end{cases} \tag{34}$$

For $(x_s, x_t) \in [0,1]^2 - \{0,1\}^2$ (at least one pixel fuzzy):

$$\varphi_{\text{hor}}(x_s, x_t) = -\alpha^{\text{f}}_{\text{hor}}(1 - 2|x_s - x_t|), \tag{35}$$

and the same formula for vertical, north–east, and north–west neighbors with corresponding parameters $\alpha^{\text{h}}_{\text{ver}}$, $\alpha^{\text{f}}_{\text{ver}}$, $\alpha^{\text{h}}_{\text{en}}$, $\alpha^{\text{f}}_{\text{en}}$, $\alpha^{\text{h}}_{\text{wn}}$, $\alpha^{\text{f}}_{\text{wn}}$.

Finally, the distribution of X is defined by

$$\alpha = (\eta_0, \eta_1, \lambda, \alpha^{\text{h}}_{\text{hor}}, \alpha^{\text{f}}_{\text{hor}}, \alpha^{\text{h}}_{\text{ver}}, \alpha^{\text{f}}_{\text{ver}}, \alpha^{\text{h}}_{\text{en}}, \alpha^{\text{f}}_{\text{en}}, \alpha^{\text{h}}_{\text{wn}}, \alpha^{\text{f}}_{\text{wn}}). \tag{36}$$

Let us detail the parameter estimation and segmentation steps in the particular model above. We basically have to specify how the Gibbs sampler runs. In fact, it allows sampling of X according to its prior and posterior distributions which is sufficient, according to the general description of ICE and MPM in Sections 3 and 4, implement them and perform an unsupervised segmentation.

First, let us indicate the sampling of X according to its prior distribution, which is needed in ICE (see (ii, 2) in the description of ICE, Section 3.2). As in the hard case, we have to specify the sampling of X_s according to its distribution conditional on $(X_t)_{t \in V}$, where V is the set of eight nearest neighbors of s.

Thus let us consider $x_\nu = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ as in Fig. 1 (we omit s because of stationarity). Let

x_8	x_1	x_2
x_7	s	x_3
x_6	x_5	x_4

FIGURE 1

1	0	0
1	s	x_3
1	x_5	x_4

FIGURE 2

$$\begin{aligned} g(x, x_\nu) = {} & \varphi(x) + \varphi_{\text{hor}}(x, x_3) + \varphi_{\text{hor}}(x, x_7) \\ & + \varphi_{\text{ver}}(x, x_1) + \varphi_{\text{ver}}(x, x_5) + \varphi_{\text{en}}(x, x_2) \\ & + \varphi_{\text{en}}(x, x_6) + \varphi_{\text{wn}}(x, x_4) + \varphi_{\text{wn}}(x, x_8). \end{aligned} \tag{37}$$

The distribution of X_s conditional on $X_V = x_V$ is then defined by the density with respect to ν

$$h^{x_\nu}(x) = \frac{e^{-g(x, x_V)}}{\int_0^1 e^{-g(x, x_V)}\, d\nu(x)} \tag{38}$$

or

$$h^{x_\nu}(x) = \frac{e^{-g(x, x_V)}}{e^{-g(0, x_V)} + e^{-g(1, x_V)} \int_0^1 e^{-g(x, x_V)}\, dx}. \tag{39}$$

For instance, if $x_\nu = (0, 0, x_3, x_4, x_5, 1, 1, 1)$ with x_3, x_4, x_5 in $]0,1[$, as in Fig. 2, $g(x, x_\nu)$ is

$$\begin{aligned} g(0, x_V) = {} & \eta_0 - \alpha^{\text{f}}_{\text{hor}}(1 - 2x_3) + \alpha^{\text{h}}_{\text{hor}} - \alpha^{\text{h}}_{\text{ver}} \\ & - \alpha^{\text{f}}_{\text{ver}}(1 - 2x_5) + \alpha^{\text{h}}_{\text{wn}} - \alpha^{\text{f}}_{\text{wn}}(1 - 2x_4) \end{aligned} \tag{40}$$

$$\begin{aligned} g(1, x_V) = {} & \eta_1 + \alpha^{\text{f}}_{\text{hor}}(1 - x_3) - \alpha^{\text{h}}_{\text{hor}} + \alpha^{\text{f}}_{\text{ver}}(1 - x_5) \\ & + \alpha^{\text{h}}_{\text{ver}} + \alpha^{\text{f}}_{\text{wn}}(1 - x_4) - \alpha^{\text{h}}_{\text{wn}} \end{aligned} \tag{41}$$

$$\begin{aligned} g(x, x_V) = {} & \lambda - \alpha^{\text{f}}_{\text{hor}}(1 - 2|x - x_3|) + \alpha^{\text{f}}_{\text{hor}}(1 - x) \\ & - \alpha^{\text{f}}_{\text{ver}}(1 - 2x) - \alpha^{\text{f}}_{\text{ver}}(1 - 2|x - x_5|) \\ & - \alpha^{\text{f}}_{\text{en}}(1 - 2x) + \alpha^{\text{f}}_{\text{en}}(1 - x) + \alpha^{\text{f}}_{\text{wn}}(1 - x) \\ & - \alpha^{\text{f}}_{\text{wn}}(1 - 2|x - x_4|). \end{aligned} \tag{42}$$

The integral in (39) is calculated by discretization. Thus one disposes of $h^{x_V}(0) = P[X_s = 0 / X_V = x_V]$, $h^{x_V}(1) = P[X_s = 1 / X_V = x_V]$, and for $x \in \,]0,1[$, of $h^{x_V}(x)$, which is a density with respect to the Lebesgue measure. The sampling of X_s is then performed as follows:

(i) sample in $\{0, 1, F\}$ according to the distribution $h^{x_V}(0)$, $h^{x_V}(1)$, $1 - h^{x_V}(0) - h^{x_V}(1)$. If $X_s = 0$ or $X_s = 1$ stop. If $X_s = F$:

(ii) sample in $]0,1[$ according to the density $h^{x_V}(x)$. The latter sampling is performed by discretizing h^{x_V}.

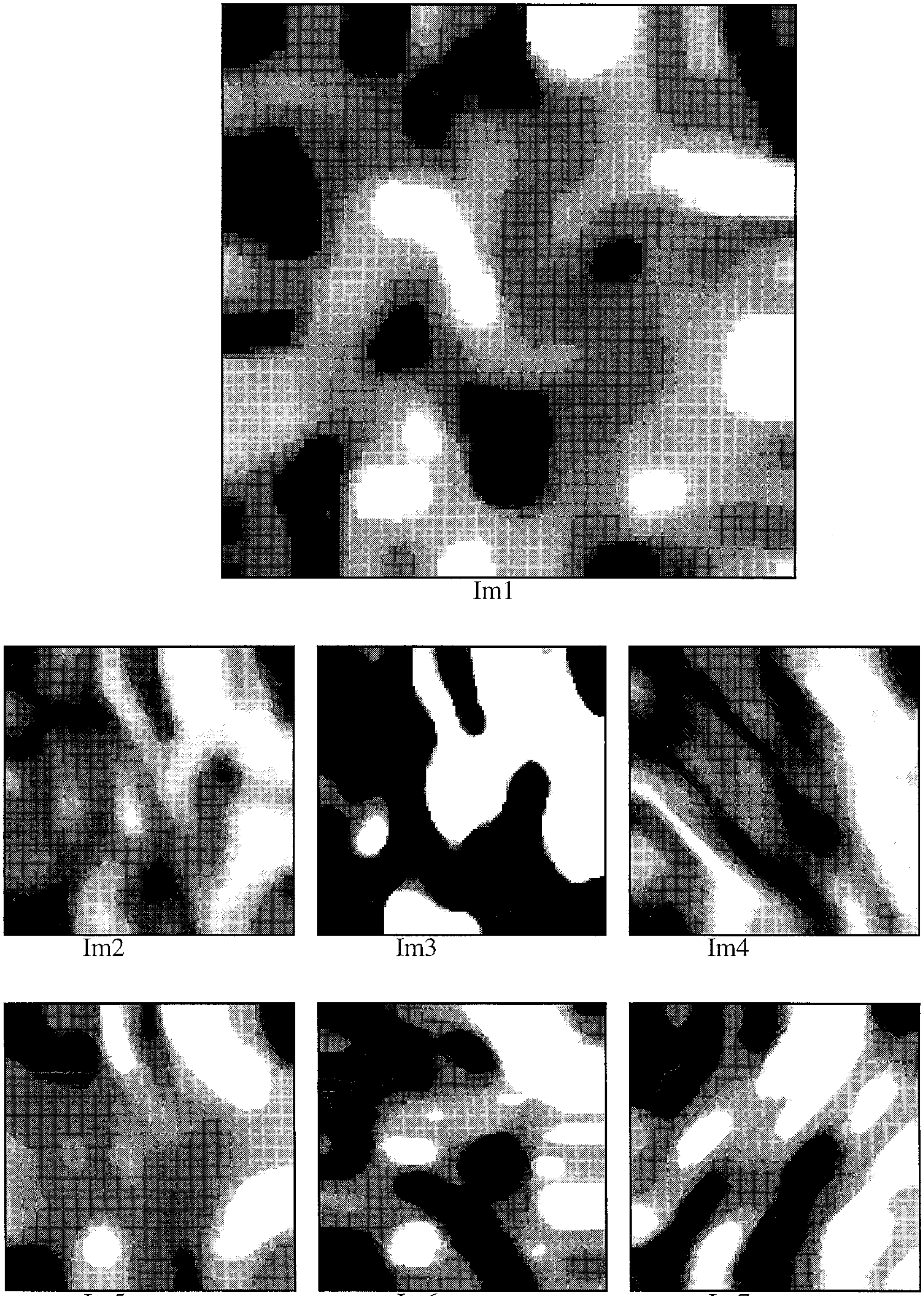


FIG. 3. Realizations of fuzzy Markov fields.

When the posterior distribution of X is considered the principle of sampling by the Gibbs sampler remains. Formula (38), (39) stay valid with g replaced by

$$g_s^y(x, x_V) = g(x, x_V) + f_x(y_s), \qquad (43)$$

where $f_x(y_s)$ is the density of the normal distribution $N[(1-x)m_0 + xm_1, (1-x)\sigma_0^2 + x\sigma_1^2]$.

Remark 3. When the proposed method is applied to real situations two questions arise; (i) how to determine the form of energy and, in particular, the number of parameters α in (36), and (ii) what can be said about the statistical behavior of the parameter sequence produced by ICE. Concerning the first point, the form of functions defined on cliques (33)–(35) has been determined experimentally; other forms have been tested that give visually less satisfying results [35]. However, this problem remains open and in each particular situation the question arises. As the whole procedure is rather time consuming, it seems difficult to consider more parameters that those defined by (36). On the other hand, some of them may be superfluous in simple situations. Concerning the behavior of the sequence produced by ICE, nothing can be said in the context of its general definition. The study of its convergence is difficult even in the case of a simple hard mixture and the only theoretical result we can put forth is its equivalence to the EM algorithm for particular parameterizations of some particular models [8]. However, ICE is better suited to the model considered in this paper because, contrary to EM, its principle is not based on likelihood, which is difficult to interpret.

5.2. *Realizations of Fuzzy Markov Random Field*

We present in Fig. 3 seven realizations of fuzzy Markov random field. The first one, presented as Im 1, is an image of size 128 × 128 and 16 grey levels have been used in expressing of the fuzzy membership. Its aim is to show that the intuitive feeling of the fuzzy reality can be rendered by the model used. In fact, one can clearly see the hard classes (black and white) and the fuzzy classes. That is likely made possible by the simultaneous use of Dirac and Lebesgue measures.

Parameters used in simulations are given in Table 1. The images Im2, Im3, Im5 show that different proportions of fuzzy pixels can be obtained; in particular, Im3 is nearly a realization of a hard field. Im4 and Im7 show that spatial anisotropy can be taken into account by the model.

Several other realizations of fuzzy Markov fields can be seen in [30].

5.3. *Unsupervised Fuzzy Segmentation*

We present in this section the results of parameter estimation and unsupervised segmentation of two synthetic images. The visual impressions are presented in Fig. 4 and in Fig. 5. The first image, Im 8, contains a strong proportion of fuzzy pixels and Im9 and Im14 are quite noisy. In fact, it is practically impossible to see anything in Im9 and Im14. Basically, the aim of the study concerning Im8 is to answer three questions: are the real parameter based methods efficient in such noisy cases? How does the parameter estimation step degrade the efficiency of the segmentations? Is the correct proportion of the fuzzy pixels retained? The second purpose of this section is to verify that the unsupervised fuzzy segmentation method we propose stays valid for hard class images. Roughly speaking, the question is: Does the algorithm retain hard pictures? It is not possible to answer this question a priori; in fact, we have seen in Section 2.1 that hard fields are not strictly particular cases of fuzzy fields, but can only be obtained when some parameter tends to infinity.

The error rates are defined by

$$\tau = \frac{1}{N} \sum_{s \in S} |x_s - \hat{s}_s(y)| \tag{44}$$

with N the number of pixels.

According to the results contained in Table 2, we note good noise parameter estimation in both MD and VD cases. Priors parameters concerning the cliques "singletons" (η_0, η_1, λ) are well estimated, and priors parameters concerning the cliques "pairs" (α^{h}, α^{f}) are rather poorly estimated. However, the degradation of the segmentation results, when using the estimated parameters instead of the real ones, seems acceptable. This shows good robustness of the segmentation methods with respect to α^{h}, α^{f}.

Remark 4. In the VD case the unsupervised ALG2 gives better results that the real parameters based one, which can appear as surprising. Such situations are not

TABLE 1
Parameters Used in Simulation of Im1–Im7

	η_0	η_1	λ	$\alpha^{\mathrm{h}}_{\mathrm{hor}}$	$\alpha^{\mathrm{h}}_{\mathrm{ver}}$	$\alpha^{\mathrm{h}}_{\mathrm{wn}}$	$\alpha^{\mathrm{h}}_{\mathrm{en}}$	$\alpha^{\mathrm{f}}_{\mathrm{hor}}$	$\alpha^{\mathrm{f}}_{\mathrm{ver}}$	$\alpha^{\mathrm{f}}_{\mathrm{wn}}$	$\alpha^{\mathrm{f}}_{\mathrm{en}}$	F (%)
Im1	0	0	0	15	15	0	0	16	16	0	0	59
Im2	0	0	0	8	8	8	8	8.7	8.7	8.7	8.7	74
Im3	0	0	0	8	8	8	8	8.3	8.3	8.3	8.3	03
Im4	0	0	0	3	3	3	20	3.5	3.5	21	3.5	57
Im5	−0.7	−0.7	1.4	8	8	8	8	8	8	8	8	70
Im6	−0.6	−0.6	1.2	20	3	3	3	20	3	3	3	40
Im7	−0.6	−0.6	1.2	3	3	20	3	3	3	20	3	40

Note. F (%): rate of fuzzy pixels.

TABLE 2
Real and Estimated Parameters of Image 7 Corrupted with MD and VD Noise, Respectively

	Priors parameters					Noise parameters				Error rates	
	η_0	η_1	λ	α^h	α^f	m_0	σ_0	m_1	σ_1	ALG1	ALG2
Real	0	0	0	8	9	1	1	3	1	9.44	9.67
Esti	−0.31	−0.30	0.61	1.1	1.2	1.04	0.98	2.89	1.10	13.04	12.87
Real	0	0	0	8	9	1	1	1	2	16.18	23.41
Esti	−0.25	−0.28	0.53	1.0	1.2	1.03	0.95	0.99	1.89	18.76	20.74

Note. ALG1, ALG2: error rates of ALG1, ALG2 based on real or estimated parameters. Number of iterations in ICE (see (24), (25)): 10. Value of the constant in the stochastic gradient (see (21)): $c = \text{Card}(S)^{-1}$. Number of iterations in both ALG1 and ALG2: 50.

impossible from the theoretical point of view, once the criterion used in order to measure the similarity between images is not adapted to the loss function used. Here we use the criterion defined by (44), which is adapted to the loss function defined by (31) and (26), the latter function defining ALG1. To be more precise, the rate defined with (44) is adapted to ALG1 in the following sense. On the one hand, this rate tends to $E[|X_s - \hat{s}_s(y)|/Y = y]$ when N tends to infinity. On the other hand, the real parameters based on ALG1 is exactly the method which minimizes $E[|X_s - \hat{s}_s(y)|/Y = y]$. Thus, for N large enough, the unsupervised ALG 1, which finally is an algorithm other than ALG1, must give worse results than ALG1. This is verified by the numerical results obtained.

The most striking visual impression concerning the results in Fig. 4 is the great efficiency of the real-parameters-based algorithm ALG1. The segmented image is very smooth in both MD and VD cases (Im 10 and Im 15) and seems more real that the real image. In the unsupervised case the difference between ALG1 and ALG2 is less apparent. According to Table 2, ALG 1 is more efficient in the VD case and ALG2 takes the upper hand in the MD case. The first case gives Im 16 for ALG1 and Im 18 for ALG2. In fact, visually Im 16 seems closer to the real image than Im 18. This could also be due to the fact that the proportion of the fuzzy pixels is better retained in Im 16 than in Im 18. In fact, according to Table 3, the real proportion is about 61%, and this proportion is about 56% in Im 16 and 49% in Im 18. Visual comparison of the results concerning the MD case is more difficult. The segmentation of the Im 8 + MD with ALG 1, resulting in Im 11, appears closer to Im 8. Indeed, the grey level variations in Im 13 are sometimes abrupt, which does not occur in the Im 8.

We have seen in Section 2 that the fuzzy hidden Markovian model becomes a hard hidden Markovian model when a certain parameter tends to infinity, which means that, strictly speaking, a hard hidden Markovian model can not be seen as a particular case of the fuzzy one. Thus a question arises: When the true nature of the class field is hard, is the parameter estimation procedure efficient enough to make the estimated fuzzy model close enough to a hard one to obtain a hard segmentation? Thus we apply our fuzzy unsupervised method to a hard image (Im 19) corrupted with MD noise (Im 20) and VD noise (Im 21). Estimates and error ratios are given in Table 4 and Im 22, Im 23 represent the fuzzy unsupervised segmentation results. Images 19–23 are presented in Fig. 5.

According to the results presented in Table 4, we note that the algorithm does not confuse the fuzzy aspect of classes with the noise. Let us point out that the noise level is comparable to those studied using hard unsupervised hidden Markov based segmentation methods. As the results of the segmentations are fairly hard fields, we can say that our fuzzy model can be very close to a hard model

TABLE 3
Rates of Fuzzy and Hard Pixels in the Real Image and in the Real Parameter or Estimated Parameter Based Segmentations with ALG1 and ALG2

		Real parameter segmentation				Estimated parameter segmentation			
		Noise 1		Noise 2		Noise 1		Noise 2	
	Real Image	ALG1	ALG2	ALG1	ALG2	ALG1	ALG2	ALG1	ALG2
0 (%)	27.15	28.7	29.2	0	0.3	37.0	38.1	19.65	24.46
1 (%)	11.45	12.4	12.4	7.1	8.1	22.6	22.6	24.54	26.54
F (%)	61.40	58.9	58.4	92.9	91.6	40.4	39.3	55.81	49.00

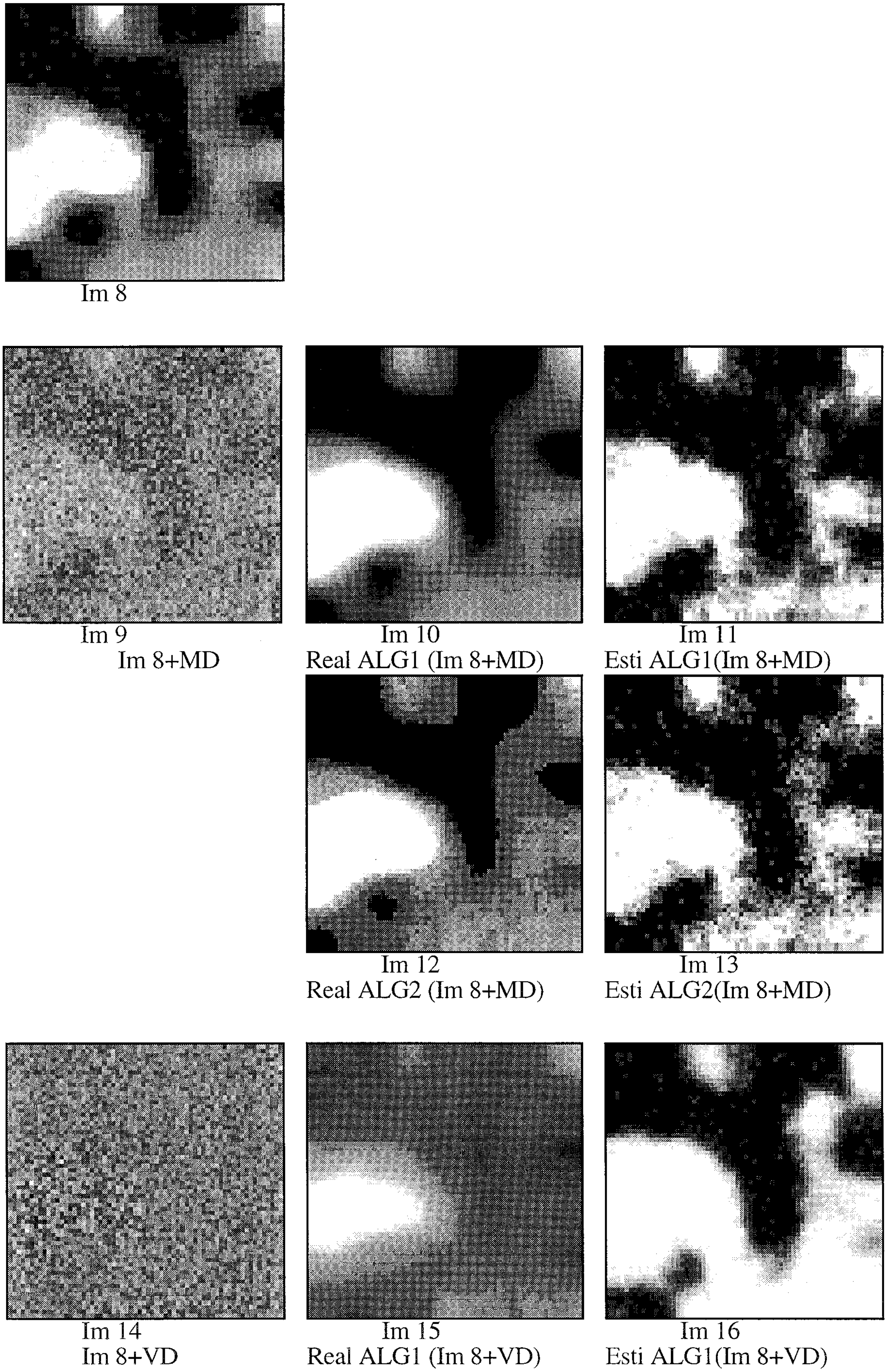
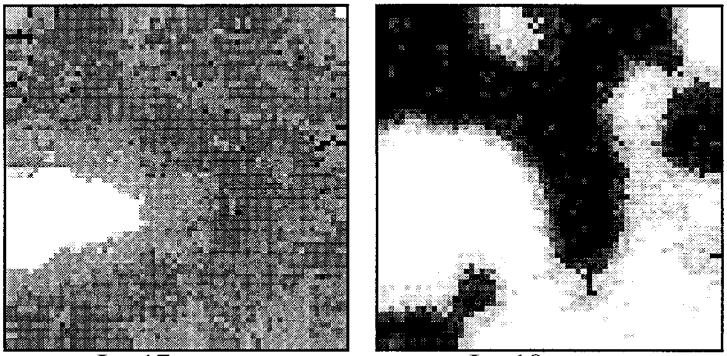


FIG. 4. Supervised and unsupervised segmentations of noisy fuzzy Markov fields.



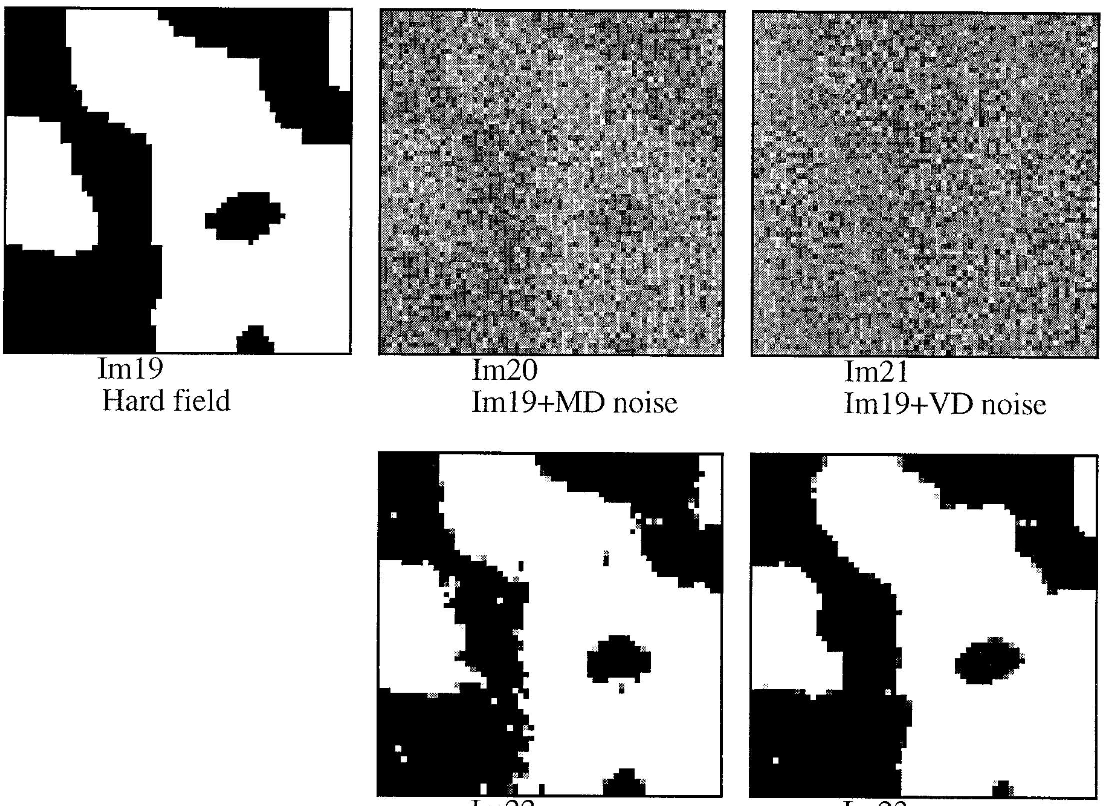
Im 17
Real ALG2 (Im 8+VD)

Im 18
Esti ALG2(Im 8+VD)

FIGURE 4—*Continued*

provided the parameters are correctly chosen. Finally, if we know nothing about fuzzy or hard nature of the class field, we take no risks by applying our fuzzy method. However, we do not claim that the fuzzy method proposed is better than the hard ones; on the contrary, according to the theory, it is equal or worse. Thus, if the fuzzy method gives a hard result, it would be undoubtedly wise to make a segmentation again by applying a hard hidden Markov model based method.

Remark 5. We do not address in this work the important problem of estimating the number of classes. Some methods for this purpose have been proposed by others, particularly Won and Derin who present in [39] a very



Im19
Hard field

Im20
Im19+MD noise

Im21
Im19+VD noise

Im22
Esti ALG1(Im20)

Im23
Esti ALG1(Im21)

FIG. 5. Unsupervised fuzzy segmentation of a hard noisy Markov field.

TABLE 4
Real and Estimated Parameters of the Hard Image 18 Corrupted with MD and VD Noise, Respectively

	Priors parameters					Noise parameters				E R		
	η_0	η_1	λ	α^h	α^f	m_0	σ_0	m_1	σ_1	ALG1	H C1	H C2
Real	X	X	X	2	X	1	1	2	1		0.41	0.59
Esti	−0.58	−0.59	1.17	−0.51	0.08	0.89	0.95	2.07	0.98	3.9%	0.39	0.60
Real	X	X	X	2	X	1	1	1	4		0.41	0.59
Esti	−0.48	−0.48	0.96	−0.57	−0.03	0.98	1.00	1.06	4.22	3.4%	0.407	0.584

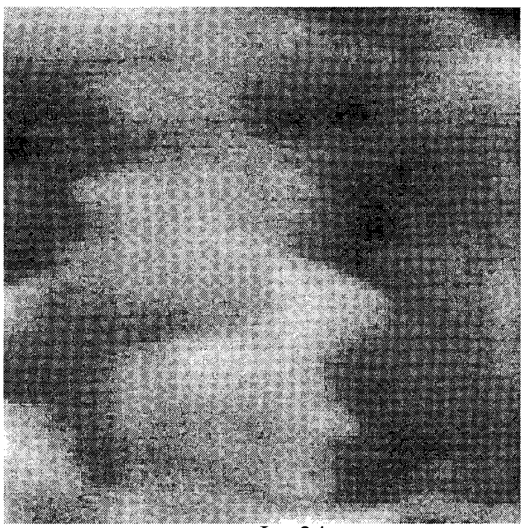
Note. E R: error rate of ALG 1 based on estimated parameters. H C1, H C2: rates of hard pixels in real and segmented images. Number of iterations in ICE (see (24), (25)): 10. Value of the constant in the stochastic gradient (see (21)): $c = \text{Card}(S)^{-1}$. Number of iteration in both ALG 1: 100. X: does not exist.

general unsupervised hard segmentation algorithm. In particular, their method estimates the number of classes in both noisy and textured cases. Such a method could perhaps be applied in a fuzzy context: indeed, the fuzzy noisy class presents some texture, and thus it should be identified as a third class.

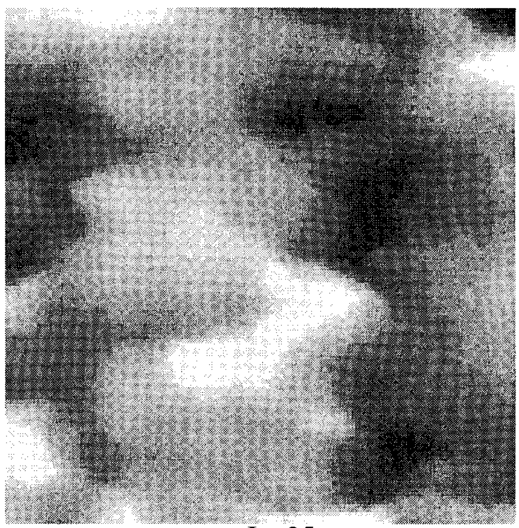
5.4. *Real Image Segmentation*

Several realizations of the fuzzy Markov field look like clouds, and thus we propose in this section a segmentation result of a real clouds image. The sky and the opaque cloud can thus be considered as hard classes, and the spots where the sky can be seen through clouds can be considered as fuzzy class. When it is fine weather, one can consider that images are noise-free, at least when they are optical images. However, according to the remarks in Section 3.2, we can consider that they are noisy. We use the model of Section 5.1; thus, the parameters to be estimated are $\alpha = (\eta_0, \eta_1, \lambda, \alpha^{\text{h}}_{\text{hor}}, \alpha^{\text{f}}_{\text{hor}}, \alpha^{\text{h}}_{\text{ver}}, \alpha^{\text{f}}_{\text{ver}}, \alpha^{\text{h}}_{\text{en}}, \alpha^{\text{f}}_{\text{en}}, \alpha^{\text{h}}_{\text{wn}}, \alpha^{\text{f}}_{\text{wn}})$ and $\beta = (\mu_0, \mu_1, \sigma_0, \sigma_1)$. The segmentation step is performed with ALG1. The real image and the segmentation result are presented in Fig. 6.

We can see that the estimates of the standard deviations are small compared to the difference between the estimates



Im 24
Real image "Clouds"



Im 25
Unsupervised fuzzy segmentation of "Clouds" with ALG1.

Parameters estimated with ICE+Stochastic Gradient (250 iterations). One iteration for ICE approximation ($m = 1$ in (24), (25)). Estimates of priors: η_0=-0.30, η_1=-0.25, λ=0.55, α^h_{hor}=0.61, α^h_{ver}=0.57, α^h_{wn}=0.51, α^h_{en}=0.50, α^f_{hor}=3.30, α^f_{ver}=2.39, α^f_{wn}=2.61, α^f_{en}=2.67. Estimates of noise parameters: m_0=22.13, m_1=203.93, σ_0=8.85, σ_1=13.25.

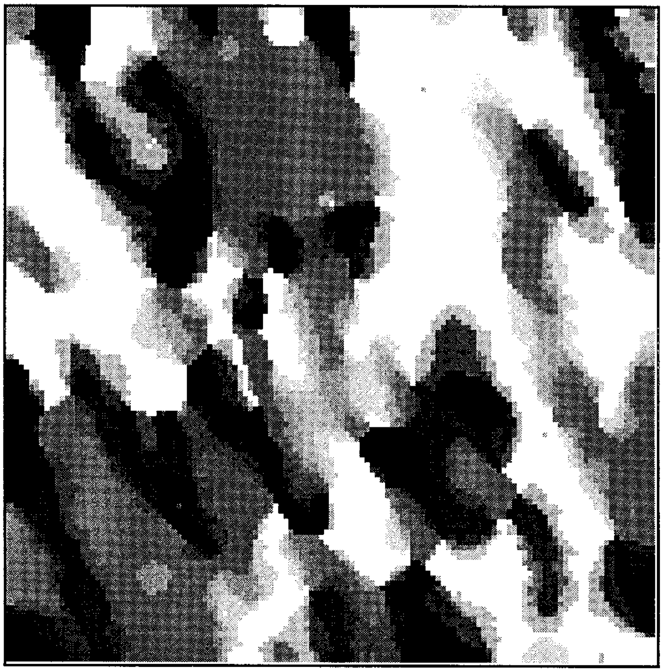
FIG. 6. Fuzzy segmentation of a real image.

of the means. Thus the algorithm ICE + ALG1 sees the real images as a fuzzy image with a little noise. However, the noise is present and the fact that the estimates of the standard deviations are different attest to the fact that the classes sky and clouds produce different noises.

Let us specify one possible application of such segmentations of clouds. An important problem in meteorology is that of automatically classifying clouds. One could imagine that different kinds of images produce different parameters. As the parameter estimation is automated, it should be possible to perform an automated classification from the estimates obtained.

5.5. *Generalization to k Classes*

Let us briefly specify how the case of more than two classes can be handled. The general form of the pdf of X given in Section 2 is stored, with the difference that for a clique C, with $n = \text{Card}(C)$, the function $\varphi_{f,C}$ associated with C is a function defined on $\Omega_f^n = ([0,1]^k)^n$. All functions $\varphi_{f,C}$ define then the density given by (1), Section 2, which is a density with respect to some measure $\nu_k^{\otimes N}$, where N is the number of pixels. For the case $k = 2$, the measure ν_k is the ν defined by (2), Section 2. Thus we have to define a measure ν_k on $[0,1]^k$. Different measures can be defined using Dirac and Lebesgue measures and each of them will define, for the same functions $\varphi_{f,C}$, a fuzzy Markov field. For instance, considering three classes ($k = 3$), one possibility is to consider ν_3 as the sum of 8 Dirac measures on vertices of $[0,1]^3$, 12 Lebesgue measures on the segments connecting these vertices, and the Lebesgue measure on $[0,1]^3$. Choosing such a measure, we relax the unit hypothesis, made through this paper, according to which the sum of fuzzy values is one. Note that such situations cannot be studied by the use of some probability measure. We present in Fig. 7 an example of realization of a fuzzy Markov field with three classes. The unit hypothesis is kept and we assume that each pixel cannot belong to more than two fuzzy classes: thus ν_3 is the sum of three Dirac measures on the vertices (1,0,0), (0,1,0), (0,0,1) and three Lebesgue measures on the segments connecting these vertices.



Im 26

FIG. 7. Fuzzy Markov field with three hard classes.

6. CONCLUSION

In this paper, we presented an unsupervised statistical fuzzy image segmentation algorithm. The method is supported by a recent model of hidden fuzzy Markov fields [30], the original feature of which was the simultaneous introduction of Dirac and Lebesgue measures at the class field level. The aim of such a fuzzy Markov random class field was to allow the simultaneous existence of hard pixels and fuzzy pixels, according to the intuitive feeling that in real images such situations can occur. Simulations show that hard and fuzzy pixels can be obtained simultaneously and their proportions vary with some Markovian energy parameters. Furthermore, the fuzzy Markov random class field model appears as a generalization of classical Markov models in the sense that they are obtained when some parameter tends to infinity. The noisy versions of the fuzzy class field are then obtained in the same way as in the classical case. One can envisage an unsupervised segmentation of such a fuzzy hidden Markov random class field once given a segmentation method and a parameter estimation method. Bayesian methods of segmentation can be defined, using different loss functions, and we have proposed one of them. The essential novelty of this paper is that we solve the parameter estimation problem by using iterative conditional estimation [29, 31]. The principle of ICE requires that we be able to simulate the class field according to the posterior distribution, on the one hand, and that we be able to estimate the parameters from both noisy and noise-free class fields, on the other hand. As a fuzzy version of the Gibbs sampler can be used, the first point is solved. The second point is treated by adapting the stochastic gradient algorithm of Younes [40] to the fuzzy noise-free field and by using empirical moments for estimating the noise parameters.

Simulation studies on synthetic images show that the proposed unsupervised fuzzy segmentation algorithm does not confuse the fuzzy aspect of the classes with the noise. This means that, on the one hand, when the original synthetic image has a given proportion of fuzzy pixels, a great deal of this proportion is found again after the unsupervised segmentation. On the other hand, when the original synthetic image is hard, i.e., without fuzzy pixels, the segmentation result stays hard.

Results presented basically concern the case of two fuzzy classes, but it is possible to extend our results to any number of fuzzy classes.

The segmentation of a real image of clouds gives visually satisfying results and some possibility of applying the proposed method to the automated clouds classification problem has been indicated.

REFERENCES

1. L. Amoura, F. Salzenstein, and W. Pieczynski, Estimation des champs de Markov flous. [In preparation]
2. J. Besag, On the statistical analysis of dirty pictures, *J. Roy. Statist. Soc. Ser. B* **48,** 1986, 259–302.
3. J. C. Bezdek, *Pattern Recognition and Fuzzy Objective Function Algorithm,* Plenum, New York, 1981.
4. B. Bharathi Devi and V. V. S. Sarma, Estimation of fuzzy memberships from histograms, *Inform. Sci.* **35,** 1985, 43–59.
5. J. M. Boucher and P. Lena, Unsupervised Bayesian classification, application to the forest of Paimpont (Brittany), *Photo Interprétation* **32,** 1994/4, 1995/1, 1995, pp. 79–81.
6. B. Braathen, W. Pieczynski, and P. Masson, Global and local methods of unsupervised Bayesian segmentation of images, *Machine Graphics Vision* **2,** No. 1, 1993, 39–52.
7. H. Caillol, A. Hillon, and W. Pieczynski, Fuzzy random fields and unsupervised image segmentation, *IEEE Trans. Geosci. Remote Sensing* **GE-31,** No. 4, 1993, 801–810.
8. H. Caillol, W. Pieczynski, and A. Hillon, Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation, *IEEE Trans. Image Process.* **IP-6,** No. 3, 1997, 425–440.
9. B. Chalmond, An iterative Gibbsian technique for reconstruction of *m*-ary images, *Pattern Recognition* **22,** No. 6, 1989, 747–761.
10. R. Chellapa and R. L. Kashyap, Digital image restoration using spatial interaction model, *IEEE Trans. ASSP,* **ASSP-30,** No. 3, 1982, 461–472.
11. R. Chellapa and A. Jain, Eds., *Markov Random Fields,* Academic Press, San Diego, 1993.
12. H. Derin and H. Elliot, Modeling and segmentation of noisy and textured images using Gibbs random fields, *IEEE Trans. Pattern Anal. Mach. Intelligence* **PAMI-9,** 1987, 39–55.
13. M. M. Dempster, N. M. Laird, and D. B. Rubain, Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc. Series B* **39,** 1977, 1–38.
14. R. C. Dubes and A. K. Jain, Random field models in image analysis, *J. Appl. Statist.* **16,** No. 2, 1989, 131–164.
15. I. Gath and A. B. Geva, Unsupervised optimal fuzzy clustering, *IEEE Trans. Pattern Anal. Mach. Intelligence* **PAMI-11,** No. 7, 1989, 773–781.
16. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intelligence* **PAMI-6,** No. 6, 1984, 721–741.
17. T. Gu and B. Dubuisson, Similarity of classes and fuzzy clustering, *Fuzzy Sets Systems* **34,** 1990, 213–221.
18. R. Haralick and J. Hyonam, A context classifier, *IEEE Trans. Geosci. Remote Sensing* **GE-24,** 1986, 997–1007.
19. A. Hillion, Les approches statistiques pour la reconnaissance des images de télédétection, in *Atti della XXXVI Riunione Scientifica,* Vol. 1, pp. 287–297, SIS, 1992.
20. T. L. Hunstberger, C. L. Jacobs, and R. L. Cannon, Iterative fuzzy image segmentation, *Pattern Recognition* **18,** No. 2, 1985, 131–138.
21. R. L. Kashyap and R. Chellapa, Estimation and choice of neighbors in spatial-interaction models of images, *IEEE Trans. Inform. Theory* **29,** 1983, 60–72.
22. P. A. Kelly, H. Derin, and K. D. Hartt, Adaptive segmentation of speckled images using a hierarchical random field model, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-36,** No. 10, 1988, 1628–1641.
23. J. T. Kent and K. V. Mardia, Spatial classification using fuzzy membership models, *IEEE Trans. Pattern Anal. Mach. Intelligence* **PAMI-10,** No. 5, 1988, 659–671.
24. S. Lakshmanan and H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing, *IEEE Trans. Pattern Anal. Mach. Intelligence* **PAMI-11,** No. 8, 1989, 799–813.
25. J. L. Marroquin, S. Mittle, and T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *J. Amer. Statist. Assoc.* **82,** 1987, 76–89.
26. P. Masson and W. Pieczynski, SEM algorithm and unsupervised statistical segmentation of satellite images, *IEEE Trans. Geosci. Remote Sensing* **GE-34,** No. 3, 1993, 618–633.
27. W. Pedrycz, Fuzzy sets in pattern recognition: Methodology and methods, *Pattern Recognition* **23,** No. 1/2, 1990, 121–146.
28. A. Peng and W. Pieczynski, Adaptive mixture estimation and unsupervised local Bayesian image segmentation, *Graph. Models Image Process.* **57,** No. 5, 1995, 389–399.
29. W. Pieczynski, Statistical image segmentation, *Machine Graphics Vision* **1,** No. 1/2, 1992, 261–268.
30. W. Pieczynski and J. M. Cahen, Champs de Markov cachés flous et segmentation d'images, *Rev. Statist. Appl.* **42,** No. 2, 1994, 13–31.
31. W. Pieczynski, Champs de Markov cachés et estimation conditionelle itérative, *Traitement Signal* **11,** No. 2, 1994, 141–153.
32. 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.* **16,** No. 2, 1989, 267–281.
33. R. A. Redner and H. F. Walker, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Rev.* **26,** 1984, 195–239.
34. A. Roesenfeld, Ed., *Image Modeling,* Academic Press, San Diego, 1981.
35. F. Salzenstein, *Modèles markoviens flous et segmentation statistique non supervisée d'images,* Ph.D. Thesis, Université de Rennes I, 1996.
36. A. Veijanen, A simulation-based estimator for hidden Markov random fields, *IEEE Trans. Pattern Anal. Mach. Intelligence* **PAMI-13,** No. 8, 1991, 825–830.
37. F. Wang, Fuzzy supervised classification of remote sensing images, *IEEE Trans. Geosci. Remote Sensing,* **GE-28,** No. 2, 1990, 194–201.
38. C. S. Won, Convergence of unsupervised image segmentation algorithms, *Proc. SPIE* **2568,** 1995, 209–220.
39. C. S. Won and H. Derin, Unsupervised segmentation of noisy and textured images using Markov random fields, *CVGIP: Graph. Models Image Process.* **54,** 1992, 308–328.
40. L. Younes, Estimation and annealing for Gibbsian fields, *Ann. Inst. Henri Poincaré* **24,** No. 2, 1988, 269–294.
41. L. Younes, Parametric inference for imperfectly observed Gibbsian fields, *Probab. Theory Relat. Fields* **82,** 1989, 625–645.
42. J. Zhang, J. W. Modestino, and D. A. Langan, Maximum likelihood parameter estimation for unsupervised stochastic model-based image segmentation, *IEEE Trans. Image Process.* **IP-3,** No. 4, 1994, 404–420.