Estimation of Generalized Mixtures and Its Application in Image Segmentation

Yves Delignon, Abdelwaheb Marzouki, and Wojciech Pieczynski

***Abstract*—We introduce in this work the notion of a generalized mixture and propose some methods for estimating it, along with applications to unsupervised statistical image segmentation. A distribution mixture is said to be "generalized" when the exact nature of components is not known, but each belongs to a finite known set of families of distributions. For instance, we can consider a mixture of three distributions, each being exponential or Gaussian. The problem of estimating such a mixture contains thus a new difficulty: We have to label each of three components (there are eight possibilities). We show that the classical mixture estimation algorithms—expectation-maximization (EM), stochastic EM (SEM), and iterative conditional estimation (ICE)—can be adapted to such situations once as we dispose of a method of recognition of each component separately. That is, when we know that a sample proceeds from one family of the set considered, we have a decision rule for what family it belongs to. Considering the Pearson system, which is a set of eight families, the decision rule above is defined by the use of "skewness" and "kurtosis." The different algorithms so obtained are then applied to the problem of unsupervised Bayesian image segmentation. We propose the adaptive versions of SEM, EM, and ICE in the case of "blind," i.e., "pixel by pixel," segmentation. "Global" segmentation methods require modeling by hidden random Markov fields, and we propose adaptations of two traditional parameter estimation algorithms: Gibbsian EM (GEM) and ICE allowing the estimation of generalized mixtures corresponding to Pearson's system. The efficiency of different methods is compared via numerical studies, and the results of unsupervised segmentation of three real radar images by different methods are presented.**

***Index Terms*—Bayesian segmentation, generalized mixture estimation, hidden Markov fields, mixture estimation, unsupervised segmentation.**

I. Introduction

OUR WORK addresses the mixture estimation problem with applications to unsupervised statistical image segmentation. In the case of independent observations, some iterative mixture estimation algorithms giving generally satisfying results have been proposed. The expectation-maximization (EM) [5], [9], [31], which allows, in some circumstances, to reach the maximum likelihood, is the pioneer one. Some variants, such as stochastic EM (SEM) [24], [26], which tend to facilitate calculations or improve the EM's performances, have since been proposed. An alternative method, called iterative conditional estimation (ICE) [3], [4], [26]–[28], is based on the conditional expectation instead of the maximum likelihood, and still allows estimate mixtures. All these methods allow one to treat the case where the nature of the components of a given mixture is known. The aim of our work is to introduce a more general model, called "generalized mixture," and propose some methods deriving from EM, SEM, or ICE for its estimation. A generalized mixture is a mixture of m components $f_1, \cdots, f_m$ where the nature of each f_i is not known exactly; however, this nature belongs to a given finite set $F = \{F_1, \cdots, F_q\}$ of natures. For instance, if we consider a mixture of two densities f_1, f_2 each of them being exponential or Gaussian, we have $F_1 = \{\text{exponential laws}\}, F_2 = \{\text{Gaussian laws}\}$. There are four possibilities of "classical" mixtures (both f_1, f_2 exponential, both f_1, f_2 Gaussian, f_1 exponential and f_2 Gaussian, f_1 Gaussian and f_2 exponential) and we do not know in what case we are. The problem of the estimating such a generalized mixture becomes twofold: First, we have to decide to which family of F each of the densities f_1, f_2 belongs; second, what are the parameters defining them.

The generalized mixture estimators we propose below are then applied to the statistical unsupervised image segmentation problem. Among numerous methods of image segmentation, the family of statistical ones turns out to be of exceptional efficiency in some situations [1]–[8], [10]–[29], [31]–[36]. The use of such methods requires modeling by random fields: For S (the set of pixels) we consider two sets of random variables $X = (X_s)_{s\in S}, Y = (Y)_{s\in S}$ called "random fields". Each X_s takes its values in a finite set of classes $\Omega = \{\omega_1, \cdots, \omega_m\}$ and each Y_s takes its values in R. The problem of segmentation is then that of estimating the unobserved realization $X = x$ of the field X from the observed realization $Y = y$ of the field Y, where $y = (y_s)_{s\in S}$ is the digital image to be segmented. The problem is then solved by the use of a Bayesian strategy, which is the "best" in the sense of some criterion. If we want to use a given Bayesian strategy s_B, we need to know some parameters defining the distribution of (X, Y). The latter distribution is generally defined by P_X, the distribution of X and the family $P_Y^{X=x}$ of the distributions of Y conditional to X. Let us denote by α all parameters concerning P_X and by β all parameters concerning the family $P_Y^{X=x}$ we need. Making the strategy s_B unsupervised amounts to proposing a way of estimating α and β from $Y = y$, the only data available. The parameter β is

Manuscript received September 26, 1995; revised January 30, 1997. The associate editor coordinating the reivew of this manuscript and approving it for publication was Prof. William E. Higgins.

Y. Delignon and A. Marzouki are with the Département Electronique, École Nouvelle d'Ingénieurs en Communication, 59650 Villeneuve d'Ascq, France.

W. Pieczynski is with the Département Signal et Image, Institut National des Télécommunications, 91000 Evry, France (e-mail: wojciech.pieczynski@int-evry.fr).

Publisher Item Identifier S 1057-7149(97)07026-7.

generally of the form $\beta = (\beta_1, \cdots, \beta_m)$, where β_i defines the distribution of Y_s conditional to $X_s = \omega$. If these distributions are Gaussian, which is the most frequently considered case, each β_i is of the form $\beta_i = (\mu_i, \sigma_i^2)$ with μ_i being the mean and σ_i^2 being the variance. The previous parameter estimation problem is then the Gaussian mixture estimation problem. In real situations, the nature of the grey-level distribution can vary in time. For instance, the nature of the radar grey-level distribution of the sea surface depends on its state [8], the latter depending on the weather. Thus, if we want to segment a radar image where sea is one of the classes and we wish to dispose of an algorithm insensitive to weather conditions, we must consider the problem of estimating a generalized mixture.

The organization of the paper is as follows. In the next section, we address the generalized mixture estimation problem without reference to the image segmentation problem. Such a mixture is defined and a method of its estimation based on the SEM is proposed.

Section III contains a description of Pearson's system, which is a set of eight families of distributions, and different methods for estimating generalized mixtures whose components belong to this set are proposed. In fact, it is shown that the classical methods EM, SEM, or ICE can be generalized resulting in generalized EM, SEM, ICE (denoted by GEM, GSEM, GICE, respectively).

In Section IV, we address the problem of unsupervised image segmentation, treating "local" and "global" methods. In the first case, GEM, GSEM, and GICE can be applied directly and we show that the use of their adaptive versions is of interest. The second case, where the segmentation is performed by the maximum posterior mode (MPM) [21], requires modeling by hidden Markov random fields. Different parameter estimation methods have been proposed; let us mention Gibbsian EM [5], the algorithms of Zhang *et al.* [37], [38], stochastic gradient [35], the algorithm of Lakshmanan *et al.* [20], the algorithm of Devijver [16], and ICE. We consider two of them (Gibbsian EM and ICE) and show that they can be generalized in order to deal with the generalized mixtures estimation problem we are interested in.

Section V contains results of some simulations, and segmentations of three real radar images are presented.

Conclusions are in the sixth section.

II. Generalized Mixture Estimation

The "classical" mixture estimation problem can be treated with methods like EM, SEM, or ICE. In this section, we will limit our presentation to GSEM. Furthermore, for the sake of simplicity, we shall consider the case of two classes and two families of distributions; its generalization is immediate and does not pose any problem. Let us note that the results of this section can be applied to any problem outside image segmentation.

A. Classical Mixture Estimation and the SEM Algorithm

Let us suppose that the random variables $Z_i(X_i, Y_i)$, with $i \in N$, are independent and identically distributed (i.i.d.), each X_i taking its values in $\Omega = \{\omega_1, \omega_2\}$ and Y_i in R. The distributions of Y_i conditioned on $X_i = \omega_1, \omega_2$ are Gaussians $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ respectively. So, given $\pi_1 = P[X_i = \omega_1]$, the parameter defining the distribution of (X, Y) is $\theta = (\pi_1, \mu_1, \mu_2, \sigma_1, \sigma_2)$. SEM is an iterative procedure that runs as follows.

1) Initialization: let $\theta^0 = (\pi_1^0, \mu_1^0, \mu_2^0, \sigma_1^0, \sigma_2^0)$ be an initial guess of $\theta = (\pi_1, \mu_1, \mu_2, \sigma_1, \sigma_2)$.
2) Calculation of $\theta^{k+1} = (\pi_1^{k+1}, \mu_1^{k+1}, \mu_2^{k+1}, \sigma_1^{k+1}, \sigma_2^{k+1})$ from $\theta^k = (\pi_1^k, \mu_1^k, \mu_2^k, \sigma_1^k, \sigma_2^k)$ and $Y = y = (y_1, \cdots, y_n)$, as follows.
   a) Compute, for each $1 \le i \le n$, the distribution of X_i conditioned on $Y_i = y_i$. If we denote by f_1^k, f_2^k the θ^k based densities f_1, f_2, this distribution is given by
   $$\pi_{1,i}^k = \frac{\pi_1^k f_1^k(y_i)}{\pi_1^k f_1^k(y_i) + \pi_2^k f_2^k(y_i)} \quad (1)$$
   $$\pi_{2,i}^k = 1 - \pi_{1,i}^k. \quad (2)$$
   b) Sample, for each $1 \le i \le n$, a realization in $\Omega = \{\omega_1, \omega_2\}$ according to the distribution above and consider $x^k = (x_1^k, \cdots, x_n^k)$ the "artificial" sample of X so obtained.
   c) Consider Q_1, Q_2 the partition of $x^k = (x_1^k, \cdots, x_n^k)$ defined by
   $$[i \in Q_1] \Leftrightarrow [x_i^k = \omega_1] \quad \text{and} \quad [i \in Q_2] \Leftrightarrow [x_i^k = \omega_2]. \quad (3)$$
   d) Calculate $\theta^{k+1} = (\pi_1^{k+1}, \mu_1^{k+1}, \mu_2^{k+1}, \sigma_1^{k+1}, \sigma_2^{k+1})$ by
   $$\pi_1^{k+1} = \frac{Card(Q_1)}{n} \qquad \mu_1^{k+1} = \frac{\sum_{i \in Q_1} y_i}{Card(Q_1)}$$
   $$\mu_2^{k+1} = \frac{\sum_{i \in Q_2} y_i}{Card(Q_2)} \quad (4)$$
   $$(\sigma_1^{k+1})^2 = \frac{\sum_{i \in Q_1} (y_i - \mu_1^{k+1})^2}{Card(Q_1)}$$
   $$(\sigma_2^{k+1})^2 = \frac{\sum_{i \in Q_2} (y_i - \mu_2^{k+1})^2}{Card(Q_2)}. \quad (5)$$
3) Stop when the sequence (θ^k) stabilizes.

B. Generalized Mixture Estimation

Let us consider $\Phi = \{F_1, F_2\}$ a set of two families of distributions, Z a real random variable whose distribution belongs either to F_1 or to F_2, and $z = (z_1, \cdots, z_n) \in R^n$ a sample of realizations of Z. Let us temporarily assume that we dispose of a decision rule $\hat{F}$: $R^n \to \Phi$, which allows us to decide from z in what set between F_1 and F_2 the distribution of Z lies. Such a decision rule, still called "Φ recognition," will be made more explicit in what follows.

In order to simplify things, we expose the generalized mixture estimation algorithm in the case of two classes and two possible families, but the generalization to any number of classes and any number of possible families is quite straightforward. Thus, we consider two random variables X, Y, where X takes its values in $\Omega = \{\omega_1, \omega_2\}$ and Y in R. The distribution of X is given by $\pi_1 = P[X = \omega_1], \pi_2 = P[X = \omega_2]$ and the distributions of Y conditional to $X = \omega_1, \omega_2$ are given by densities f_1, f_2, respectively. Let $\Phi = \{F_1, F_2\}$, with F_1 the Gaussian family and F_2 the exponential one. We assume that f_1 is Gaussian ($f_1 \in F_1$) or exponential ($f_1 \in F_2$), and likewise for f_2. Thus, we have four possibilities for "classical" mixture (both f_1, f_2 Gaussian, both f_1, f_2 exponential, f_1 Gaussian and f_2 exponential, f_1 exponential and f_2 Gaussian) and we do not know in what case we are. We observe a sample $(y_1, \cdots, y_n) \in R^n$ of realizations of Y, and the problem is to

1) estimate priors;
2) choose between the four cases above;
3) estimate the parameters of the densities chosen.

The GSEM we propose runs as follows.

1) Initialization.
2) At each iteration
   a) sample as in the case of the SEM;
   b) apply, on Q_1 and Q_2, the rule $\hat{F}$ determining the families that f_1 and f_2 belong to;
   c) use Q_1 and Q_2 for estimating parameters (mean and variance if the family is Gaussian, mean if the family is exponential), in the same way that with SEM.

Thus, the GSEM will be defined once we propose a decision rule $\hat{F}$.

In this paper, we will consider a $\hat{F}$ well suited to the Pearson family described in the next section; however, other possibilities exist [14].

III. System of Pearson and Φ Recognition

A. System of Pearson

In this section, we specify the family Φ we will use in the unsupervised radar image segmentation and a decision rule $\hat{F}$. Our statement about Pearson's system we will use is rather short, and further details can be found in [17].

A distribution density f on R belongs to Pearson's system if it satisfies

$$\frac{1}{f(y)}\frac{df(y)}{dy} = -\frac{y+a}{c_0 + c_1 y + c_2 y^2}. \tag{6}$$

The variation of the parameters a, c_0, c_1, c_2 provides distributions of different shape and, for each shape, defines the parameters fixing a given distribution. Let Y be a real random variable whose distribution belongs to Pearson's system. For $q = 1, 2, 3, 4$ let us consider the moments of Y defined by

$$\mu_1 = E[Y] \tag{7}$$

$$\mu_q = E[(Y - E(Y))^q] \quad \text{and} \quad q \geq 2 \tag{8}$$

and two parameters γ_1, γ_2 defined by

$$\gamma_1 = \frac{(\mu_3)^2}{(\mu_2)^3} \quad \gamma_2 = \frac{\mu_4}{(\mu_2)^2}. \tag{9}$$

$\sqrt{\gamma_1}$ is called "skewness" and γ_2 "kurtosis."

On the one hand, the coefficients a, c_0, c_1, c_2 are related to $\mu_1, \mu_2, \gamma_1, \gamma_2$ by (10)–(13), shown at the bottom of the page.

On the other hand, given $\lambda = \gamma_1(\gamma_2+3)^2/4(4\gamma_2 - 3\gamma_1)(2\gamma_2 - 3\gamma_1)(2\gamma_2 - 3\gamma_1 - 6)$, the eight families of the set $\Phi = \{F_1, \cdots, F_8\}$, whose exact shape will be given in the next section, are defined by

$$\begin{aligned}
&[P_Y \in F_1] \Leftrightarrow [\lambda < 0]\\
&[P_Y \in F_2] \Leftrightarrow [\gamma_1 = 0 \quad \text{and} \quad \gamma_2 < 3]\\
&[P_Y \in F_3] \Leftrightarrow [2\gamma_2 - 3\gamma_1 - 6 = 0]\\
&[P_Y \in F_4] \Leftrightarrow [0 < \lambda < 1]\\
&[P_Y \in F_5] \Leftrightarrow [\lambda = 1]\\
&[P_Y \in F_6] \Leftrightarrow [\lambda > 1]\\
&[P_Y \in F_7] \Leftrightarrow [\gamma_1 = 0 \quad \text{and} \quad \gamma_2 > 3]\\
&[P_Y \in F_8] \Leftrightarrow [\gamma_1 = 0 \quad \text{and} \quad \gamma_2 = 3].
\end{aligned} \tag{14}$$

The eight families are illustrated in the Pearson's graph given in Fig. 1.

What is important is that moments $\mu_1, \cdots, \mu_4$ can be easily estimated from empirical moments, from which we deduce the estimated values of γ_1, γ_2 by (9). Finally, we estimate the family using (14). Once the family is estimated, values of a, c_0, c_1, c_2, given by (10)–(13) can be used to solve for parameters defining the corresponding densities (given

$$a = \frac{(\gamma_2+3)\sqrt{\gamma_1 \mu_2}}{10\gamma_2 - 12\gamma_1 - 18} - \mu_1 \tag{10}$$

$$c_0 = \frac{\mu_2(4\gamma_2 - 3\gamma_1) - \mu_1(\gamma_2+3)\sqrt{\gamma_1\mu_2} + (\mu_1)^2(2\gamma_2 - 3\gamma_1 - 6)}{10\gamma_2 - 12\gamma_1 - 18} \tag{11}$$

$$c_1 = \frac{(\gamma_2+3)\sqrt{\gamma_1\mu_2} - 2\mu_1(2\gamma_2 - 3\gamma_1 - 6)}{10\gamma_2 - 12\gamma_1 - 18} \tag{12}$$

$$c_2 = \frac{(2\gamma_2 - 3\gamma_1 - 6)}{10\gamma_2 - 12\gamma_1 - 18} \tag{13}$$

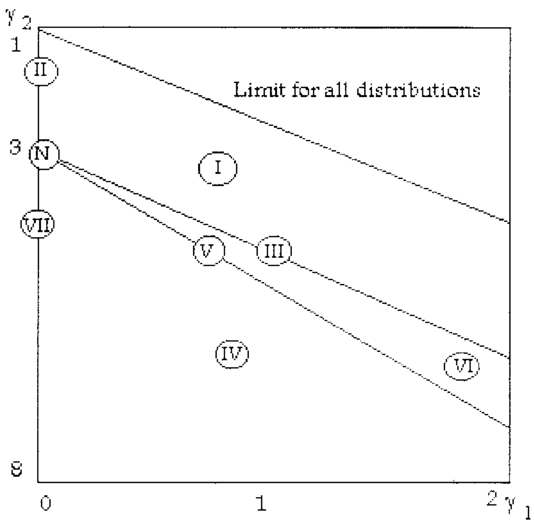

Fig. 1. The eight families of Pearson's system function of (γ_1, γ_2).

in Section III-B, where the shapes of the eight families are recalled).

Let us consider an i.i.d. sequence of real random variables $Y_1, \cdots, Y_n$ whose distribution belongs to Pearson's system. We now specify the estimator $\hat{F}$ used in step 2 of GSEM (see Section II-B).

1) Consider $Q_1, \cdots, Q_m$ the partition of x^k
2) For each class ω_i use Q_i in order to estimate $\mu_{j,i}$ by

$$\hat{\mu}_{1,i} = \frac{\sum_{s \in Q_i} y_s}{Card(Q_i)} \tag{15}$$

$$\hat{\mu}_{j,i} = \frac{\sum_{s \in Q_i} (y_s - \hat{\mu}_{1,i})^j}{Card(Q_i)} \quad \text{for} \quad j = 2, 3, 4. \tag{16}$$

3) For each class ω_i calculate $\hat{\gamma}_1^i, \hat{\gamma}_2^i$ from $\hat{\mu}_{1,i}, \cdots, \hat{\mu}_{4,i}$ according to (9).
4) For each class ω_i use $\hat{\gamma}_1^i, \hat{\gamma}_2^i$ and (14) to estimate which family among $F_1, \cdots, F_8$ the density f_i belongs to.
5) With the estimated family and the computed a, c_0, c_1, c_2 [(10)–(13)], estimate the parameters of the distribution. (For each $F_1, \cdots, F_8$ the exact relationship between density parameters and the computed a, c_0, c_1, c_2 is given in the next section.)

B. Shape of Pearson's System Densities

In this section, we specify the shape of the eight distribution families forming Pearson's system.

F_1 (Beta Distributions of the First Kind): Densities are given by

$$f_1(y) = \begin{cases} \dfrac{1}{B(p,q)} \dfrac{(y-b_1)^{p-1}(b_2-y)^{q-1}}{(b_2-b_1)^{p+q-1}}, & \text{for } y \in [b_1, b_2] \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

with

$$\Delta = \frac{c_1^2 - 4c_0c_2}{c_2^2} \quad b_1 = -\frac{1}{2}\left(\frac{c_1}{c_2} + \sqrt{\Delta}\right)$$
$$b_2 = -\frac{1}{2}\left(\frac{c_1}{c_2} - \sqrt{\Delta}\right) \tag{18}$$
$$p = \frac{a + b_1}{c_2(b_2 - b_1)} + 1 \quad q = -\frac{a + b_2}{c_2(b_2 - b_1)} + 1. \tag{19}$$

Parameters $p > 0, q > 0$ are called form parameters. f_1 can take five different forms according to p, q. To be more precise

1) for $p > 1, q > 1$ density f_1 is bell shaped;
2) for $0 < p < 1, q > 1$ density f_1 is L shaped with $\lim_{y \to a} f_1(y) = +\infty$;
3) for $p > 1, 0 < q < 1$ density f_1 is J shaped with $\lim_{y \to b} f_1(y) = +\infty$;
4) for $0 < p < 1, 0 < q < 1$ density f_1 is U shaped with $\lim_{y \to a} f_1(y) = \lim_{y \to b} f_1(y) = +\infty$;
5) for $p = q = 1$ density f_1 is uniform.

F_2 (Type II Distributions): These distributions are particular cases of F_1 obtained for $p = q$ in (17), as follows:

$$f_2(y) = \begin{cases} \dfrac{1}{B(p,q)} \dfrac{(y-b_1)^{p-1}(b_2-y)^{p-1}}{(b_2-b_1)^{2p-1}}, & \text{for } y \in [b_1, b_2] \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

with

$$b_1 = -a - \frac{1}{2}\sqrt{\Delta} \qquad b_2 = -a + \frac{1}{2}\sqrt{\Delta} \tag{21}$$
$$p = \frac{a + b_1}{c_2(b_2 - b_1)} + 1. \tag{22}$$

F_3 (Gamma Distributions): Densities are given by

$$f_3(y) = \begin{cases} \dfrac{1}{p\Gamma(q)} \left(\dfrac{y-r}{p}\right)^{q-1} e^{-(y-r)/p}, & \text{for } y \geq r \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

with

$$p = c_1 \quad q = \frac{1}{c_1}\left(\frac{c_0}{c_1} - a\right) + 1 \quad r = -\frac{c_0}{c_1}. \tag{24}$$

F_4 (Type IV Distributions): Densities are given by

$$f_4(y) = K[C_0 + c_2(y + C_1)^2]^{-(1/2c_2)} \cdot \exp\left(-\frac{a - C_1}{\sqrt{C_0 c_2}} \operatorname{arctg}\left\{\sqrt{\frac{c_2}{C_0}}(y + C_1)\right\}\right) \tag{25}$$

with K such that $\int_R f_4(y)\, dy = 1$ and

$$C_0 = c_0 - \frac{c_1^2}{4c_2} \quad C_1 = \frac{c_1}{2c_2}.$$

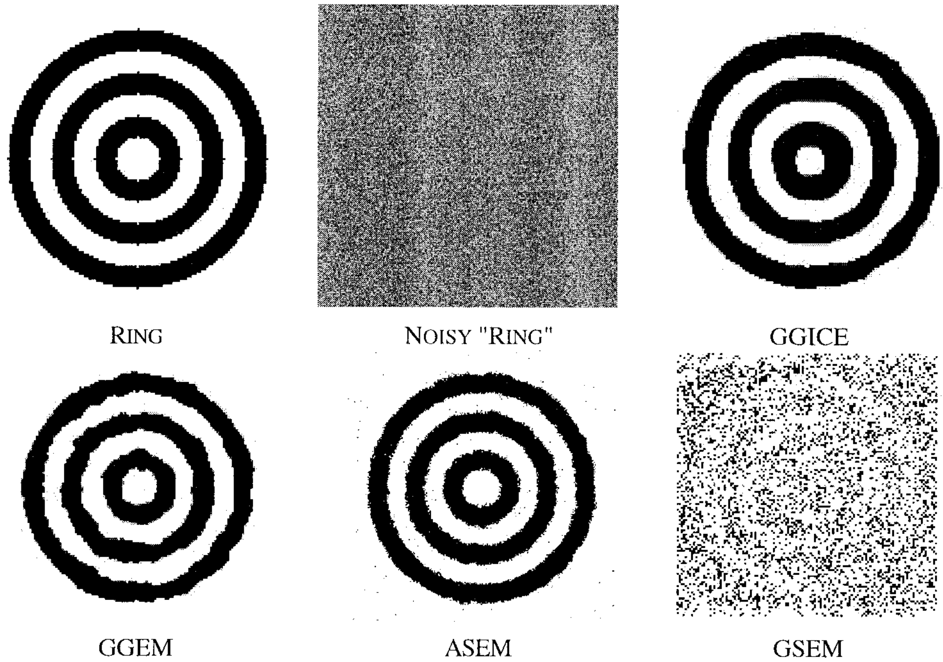

Fig. 2. Ring image, its noisy version, and results of unsupervised segmentations based on GGICE, GGEM, ASEM, and GSEM.

F_5 (Inverse Gamma Distributions): Densities are given by

$$f_5(y) = \begin{cases} \dfrac{p}{\Gamma(q)}(p(y-r))^{-q-1}e^{-(2/p(y-r))} & \text{for } y \geq r \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

with

$$p = \frac{c_2}{a - \dfrac{c_1}{2c_2}} \quad q = \frac{1}{c_2} - 1 \quad r = -\frac{c_1}{2c_2}.$$

F_6 (Beta Distributions of the Second Kind): Densities are given by

$$f_6(y) = \begin{cases} \dfrac{s^q}{B(p,q)}\dfrac{(y-r)^{p-1}}{(y-(r-s))^{p+q}}, & \text{for } y \geq r \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

with

$$p = -\frac{a - \dfrac{1}{2c_2}(c_1 - \sqrt{c_1^2 - 4c_0c_2})}{\sqrt{c_1^2 - 4c_0c_2}} + 1 \quad q = \frac{1}{c_2} - 1$$

$$r = -\frac{1}{2c_2}(c_1 - \sqrt{c_1^2 - 4c_0c_2}) \quad s = \sqrt{\frac{c_1^2 - 4c_0c_2}{c_2^2}}$$

where $s > 0$ is the scale parameter and p, q are the form parameters.

F_7 (Type VII Distributions): Densities are given by

$$f_7(y) = K[c_0 + c_2y^2]^{-(1/2c_2)} \cdot \exp\left(-\frac{a}{\sqrt{c_0c_2}}\operatorname{arctg}\left(\sqrt{\frac{c_2}{c_0}}y\right)\right) \tag{28}$$

with K such that $\int_R f_7(y)\,dy = 1$.

F_8 (Gaussian Distributions): Densities are given by

$$f_8(y) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-((y-\mu)^2/2\sigma^2)} \tag{29}$$

with $\mu = -a$ and $\sigma^2 = c_0$.

C. Generalized EM and ICE Algorithms

The EM and ICE algorithms are two other mixture estimation methods that can also be “generalized” to give the GEM and GICE. We briefly describe below their operation.

1) GEM: Let $\pi_{i,1}^k, \cdots, \pi_{i,n}^k$ be the distributions $P[X_1 = \omega_i/Y_1 = y_1], \cdots, P[X_n = \omega_i/Y_n = y_n]$ computed from the current parameter θ^k. Priors are reestimated by formula (30), which is the same as that in the EM algorithm, and the Φ recognition is the same as that the Φ recognition described at the end of Section III-A, with the difference that $\mu_{1,i}^{k+1}, \mu_{2,i}^{k+1}, \mu_{3,i}^{k+1}, \mu_{4,i}^{k+1}$, given for $j = 1, 2, 3, 4$ by formulas (31) and (32), are used instead of those given by formulas (15) and (16).

$$\pi_i^{k+1} = \frac{1}{n}(\pi_{i,1}^k + \cdots + \pi_{i,n}^k) \tag{30}$$

$$\mu_{1,i}^{k+1} = \frac{y_1\pi_{i,1}^k + \cdots + y_n\pi_{i,n}^k}{\pi_{i,1}^k + \cdots + \pi_{i,n}^k} \tag{31}$$

$$\mu_{j,i}^{k+1} = \frac{(y_1 - \mu_{1,i}^{k+1})^j\pi_{i,1}^k + \cdots + (y_n - \mu_{1,i}^{k+1})^j\pi_{i,n}^k}{\pi_{i,1}^k + \cdots + \pi_{i,n}^k}. \tag{32}$$

2) GICE: In the context of this paper, the GICE used is a “mixture” of GSEM and GEM. In fact, the reestimation of priors is the same as in GEM, and the family recognition and noise parameter reestimation is the same as in GSEM.

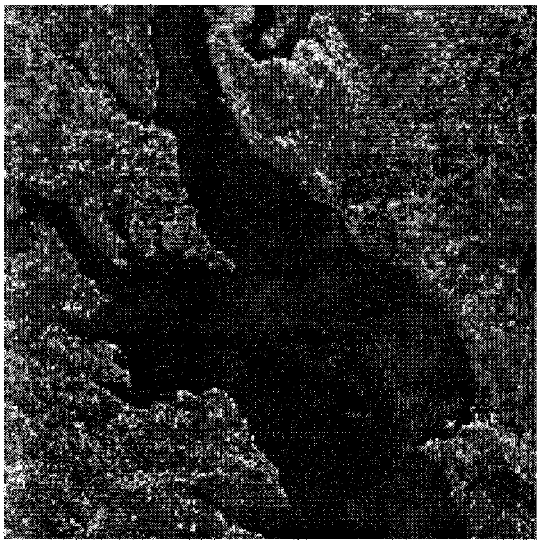

Fig. 3. Image 1: SEASAT image of the Brittany coast.

IV. Unsupervised Image Segmentation

In this section, we propose some applications of different generalized mixture estimators to the problem of unsupervised image segmentation. We shall consider two well known approaches: the "blind" approach and the "global" one. In the blind approach the generalized SEM, EM, and ICE algorithms above can be applied directly. In the global one we propose two adaptations of Gibbsian EM and ICE. For each method we specify here the reestimation formulas; the initialization of different algorithms is described in Section V.

A. Blind Approach

The "blind" approach consists of estimating the realization of each X_s from Y_s. This is the simplest one and, generally, the least efficient. However, its "adaptive" version can be very competitive in some situations [26]. Let π_1, π_2 be priors and f_1, f_2 be densities of the distribution of Y_s conditional to X_s. The blind Bayesian strategy is

$$s_B(y_s) = \begin{cases} \omega_1, & \text{if } \pi_1 f_1(y_s) \geq \pi_2 f_2(y_s) \\ \omega_2, & \text{if } \pi_1 f_1(y_s) \leq \pi_2 f_2(y_s). \end{cases} \quad (33)$$

This strategy is made unsupervised by the direct use of the GSEM algorithm described above: *One chooses a sequence of pixels $s_1, s_2, \cdots, s_n$ and considers that y_i is the value of the grey level at pixel s_i*. In an "adaptive" version of the "blind" approach, one considers that priors depend on the position of the pixel in S. The blind adaptive Bayesian strategy is the same as s_B above with π_1^s, π_2^s instead of π_1, π_2. The GSEM algorithm is modified as follows. Let $x_1, x_2, \cdots, x_n$ be the sequence obtained by sampling at a given iteration. In GSEM the priors π_1, π_2 are reestimated by the frequencies computed using all the sample points; in "adaptive" GSEM one considers, for each s_i, a window W_i centred at s_i and $\pi_1^{s_i}, \pi_2^{s_i}$ are reestimated by frequencies computed from $(x_j)_{j \in W_i}$. Let us note that in "adaptive" GSEM the sequence of pixels $s_1, s_2, \cdots, s_n$ has to cover S. In the following, the generalized adaptive SEM, EM, and ICE will be denoted by GASEM, GAEM, and GAICE.

B. Global Approach

1) Markovian Model and Global Segmentation: In the global approach, each X_s is estimated from $Y = (Y_s)_{s \in S}$. The field $X = (X_s)_{s \in S}$ is a Markov random field and we will consider Ising's model, which is the simplest one. In order to simplify notations we will limit our presentation to the case of two classes; however, the generalization to any other number of classes poses no particular problem.

The distribution of X is given by

$$P_X[x] = ce^{-U_\alpha(x)} \quad (34)$$

with

$$U_\alpha(x) = \sum_{s,t \text{ neighbors}} \varphi_\alpha(x_s, x_t) \quad \text{and}$$
$$\varphi_\alpha(x_s, x_t) = \begin{cases} -\alpha, & \text{if } x_s = x_t \\ \alpha & \text{,if } x_s \neq x_t. \end{cases} \quad (35)$$

Thus, P_X is defined by α. The random variables $(Y_s)_{s \in S}$ will be assumed independent conditionally to X, and furthermore, the distribution of each Y_s conditional to X will be assumed equal to its distribution conditional on X_s. Under these hypothesis all distributions of Y conditional to X are defined by the two distributions of Y_s conditional to $X_s = \omega_1, \omega_2$ respectively. Let us denote by f_1, f_2 the densities of these distributions and assume that they belong to Pearson's system. They are thus given by parameters $\beta_1 = (a^1, c_0^1, c_1^1, c_2^1)$ and $\beta_2 = (a^2, c_0^2, c_1^2, c_2^2)$, respectively.

Finally, all distributions of Y conditional to X are defined by $\beta = (\beta_1, \beta_2)$ and thus $\theta = (\alpha, \beta)$ defines the distribution of (X, Y).

The possibility of simulating realizations of X according to its posterior, i.e. conditional to Y, distribution constitutes the main interest of this model.

2) Generalized Global ICE (GGICE): According to the ICE principle, let us suppose that X is observable. We have then to propose $\hat{\theta} = \hat{\theta}(X, Y) = (\hat{\alpha}(X, Y), \hat{\beta}(X, Y))$.

There exist numerous estimators $\hat{\alpha} = \hat{\alpha}(X)$ of the parameter α from X, such as the coding method [2], the least squares error method [12], or the maximum likelihood estimate [35]. As our model is very simple, we can use an empirical frequency based estimator. In fact, there exists a simple link between α and probabilities "$X_s = \omega_1$ knowing that the neighborhood of s contains ω_1 r times," where r can take 0, 1, 2, 3, 4 as values. For instance, if we take $r = 2$, we select in the image $X = x$ a sample of neighborhoods of s containing two ω_1 and two ω_2. The probability "$X_s = \omega_1$ knowing that the neighborhood of s contains two ω_1" is estimated by the proportion of the sample giving $X_s = \omega_1$. On the other hand this probability is given by

$$\frac{e^{-2\alpha}}{e^{-2\alpha} + e^{2\alpha}} \quad (36)$$

which gives an estimated value of α.

We take for $\hat{\beta}(X, Y)$ the same estimator as in the case of independent mixture, Section III-A.

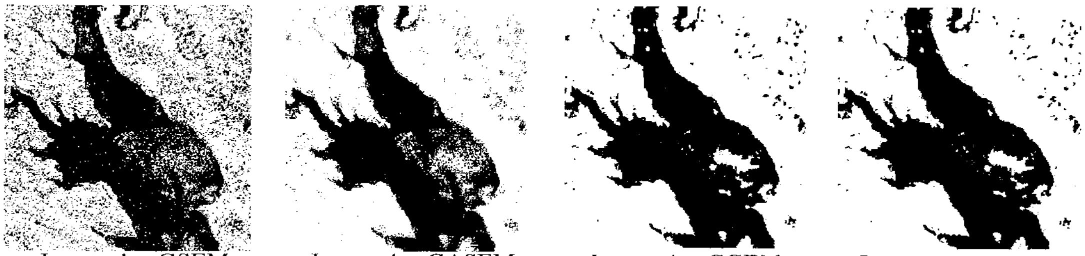

Fig. 4. Unsupervised segmentation results of Image 1, Fig. 3.

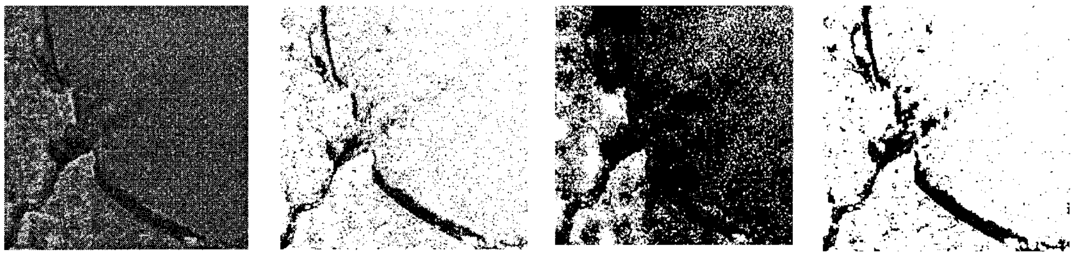

Fig. 5. Image 2 and its unsupervised segmentations.

Finally, the GGICE runs as follows.

1) Sample x^{k+1} according to $P_{X,\theta^k}^{Y=y}$.
2) Compute $\alpha^{k+1} = \hat{\alpha}(x^{k+1})$.
3) Consider $Q_1 = \{s \in S/x_s^{k+1} = \omega_1\}$ and $Q_2 = \{s \in S/x_s^{k+1} = \omega_2\}$ and apply 2–5 of the end of Section III-A.

3) Generalized Gibbsian EM (GGEM)

The difference between GGICE above and GGEM is situated at the noise parameter reestimation level. We have two noise distributions conditional to the two classes and we are interested in estimating the four first moments of each of them. In the case of GGICE, these two problems are treated separately by considering the partition on $Q_1 = \{s \in S/x_s^{k+1} = \omega_1\}$ and $Q_2 = \{s \in S/x_s^{k+1} = \omega_2\}$ of the set of pixels S. In the case of GGEM each of them is treated by the use of the whole set S. Let us put

$$p_{1,s}^{y,k} = P_{\theta_k}[X_s = \omega_1/Y = y]. \tag{37}$$

The first four moments of the noise corresponding to the first class $\mu_{1,1}^{k+1}, \mu_{1,2}^{k+1}, \mu_{1,3}^{k+1}, \mu_{1,4}^{k+1}$ are given by

$$\mu_{1,1}^{k+1} = \frac{\sum_{s\in S} y_s p_{1,s}^{y,k}}{\sum_{s\in S} p_{1,s}^{y,k}} \tag{38}$$

and

$$\mu_{1,i}^{k+1} = \frac{\sum_{s\in S} (y_s - \mu_{1,1}^{k+1})^i P_{1,s}^{y,k}}{\sum_{s\in S} p_{1,s}^{y,k}} \tag{39}$$

for $i = 2, 3, 4$.

Use analogous formulas for the second class.

V. Experiments

We present in this section some results of numerical applications. Let us note that in the global case the segmentation is performed by the maximizer of posterior marginals (MPM) and, in the local case, it is performed by the rule (33). Thus, unsupervised segmentation algorithms considered in this paper mainly differ by their parameter estimation step: We will note them by the parameter estimation method used. For instance, GEM will denote the local segmentation (33) based on parameters estimated with generalized EM, GGEM will denote the global MPM segmentation based on parameters estimated with generalized Gibbsian EM, and so on. The first section is devoted to synthetic images and in the second one we deal with three real radar images.

The initialization of GEM, GSEM, and GICE is as follows. We assume that we have a mixture of two Gaussian distributions. With H denoting the cumulated histogram we take $\mu_1^0 = H^{-1}(\frac{1}{2}), \mu_2^0 = H^{-1}(\frac{2}{3})$ and $(\sigma_1^0)^2 = (\sigma_2^0)^2 = (\mu_2^0 - \mu_1^0)^2/4$.

In order to initialize GGEM and GGICE, we use the segmentation obtained by the blind unsupervised method, which gives α^0. The noise parameters are initialized by the final parameters obtained in the parameter estimation step of the blind unsupervised method used.

A. Experiments on a Synthetic Image

Let us consider a binary image "ring" given in Fig. 2. White is class 1 and black class 2. The class 1 is corrupted by a beta noise of the first kind (family F_1 in Pearson's system) and the class 2 is noised by a beta noise of the second kind (family F_6 in Pearson's system). The parameters defining the noise distributions, their estimates with different methods, and the segmentation error rates are given in Table I. The noisy version of the ring image and some segmentation results are presented in Fig. 2. We have taken the same means and variances on

TABLE I

PAR: PARAMETERS; TH: REAL VALUES OF PARAMETERS; μ, σ^2: MEAN AND VARIANCE; γ_1, γ_2: SKEWNESS AND KURTOSIS; TYPE: FAMILY IN PEARSON'S SYSTEM; ERROR: SEGMENTATION ERROR RATE; ITE: NUMBER OF ITERATIONS OF ALGORITHMS. α: PARAMETER OF THE MARKOV DISTRIBUTION OF X. GGICE: GLOBAL GENERALIZED ICE, GGEM: GENERALIZED GIBBSIAN EM

	CLASS 1						CLASS 2							
	LOCAL METHODS													
PAR	μ	σ^2	γ_1	γ_2	π	TYPE	μ	σ^2	γ_1	γ_2	π	TYPE	ERR	ITE
TH	130	500	0.28	3.51	0.67	6	130	500	0.03	3.02	0.33	1	32	
INIT	121	361			0.5	8	140	361			0.5	8		
GSEM	130	501	0.32	3.6	0.61	6	130	497	0.01	2.95	0.39	1	33	1000
GEM	120	432	0.05	3.69	0.5	4	140	347	0.11	3.76	0.50	4	40	100
GICE	130	491	0.29	3.6	0.67	4	129	516	0.02	2.92	0.33	1	33	1000
GLOBAL METHODS														
PAR	μ	σ^2	γ_1	γ_2	α	TYPE	μ	σ^2	γ_1	γ_2	α	TYPE	ERR	ITE
INIT	130	501	0.32	3.6	-0.49	1	130	497	0.01	2.95	-0.49	1		
GGEM	130	500	0.29	3.55	-0.7	6	130	499	0.03	2.94	-0.7	1	4.7	300
GGICE	130	499	0.27	3.52	-0.9	6	130	501	0.01	3.01	-0.9	1	4.1	300

TABLE II

PARAMETERS ESTIMATED FROM IMAGE 1, FIG. 3

	CLASS 1						CLASS 2						
	LOCAL METHODS												
PAR	μ	σ^2	γ_1	γ_2	π	TYPE	μ	σ^2	γ_1	γ_2	π	TYPE	ITE
INIT	45	1296			0.5	8	81	1296			0.5	8	
GSEM	34	125	0.21	3.19	0.25	1	81	1460	0.91	4.25	0.75	1	200
GEM	34	101	0.30	3.34	0.27	1	83	1444	0.90	4.27	0.73	1	200
GICE	35	109	0.30	3.30	0.30	1	84	1423	0.91	4.30	0.70	1	200
GLOBAL METHODS													
PAR	μ	σ^2	γ_1	γ_2	α	TYPE	μ	σ^2	γ_1	γ_2	α	TYPE	ITE
INIT	34	125	0.21	3.19	-0.49	1	81	1460	0.91	4.25	-0.49	1	
GGEM	35	125	0.43	3.56	-0.80	1	89	1311	1.11	4.52	-0.80	1	30
GGICE	35	127	0.48	3.69	-0.80	1	88	1317	1.10	4.50	-0.80	1	30

purpose: The human eye is essentially sensitive to the two first moments and, in fact, it is difficult to see anything in the noisy version of the ring image.

According to Table I, the behavior of the GSEM and the GICE is quite satisfactory when results obtained with GEM are clearly worse. In particular, GEM does not find the right families (F_1 and F_6). Furthermore, the GSEM- and GICE-based segmentation error rates are very close to the theoretical one. On the other hand, the behavior of both the GGICE and GGEM methods is very good. This is undoubtedly due to a good initialization with GSEM; however, the estimates of skewness and kurtosis are still improved. We do not dispose of the theoretical segmentation error rate, as the ring image is not a realization of a Markov field. However, the error rates obtained seem quite satisfactory. As a conclusion, we may say that the new difficulty of noise nature recognition is correctly treated by the methods proposed, and the final segmentation quality is not affected significantly. We also present in Fig. 2 the result of segmentation with generalized adaptive SEM (GASEM) whose quality is nearly comparable with the quality of global methods. The result obtained with GSEM is very poor compared to the results of global methods: This is not surprising, and is due to the segmentation method and not to the parameter estimation step.

B. *Segmentation of Real Images*

We present in this section some examples of unsupervised segmentation of three real radar images. The first one, given in Fig. 3, does not seem particularly noisy and adaptive local segmentation seems to be competitive compared to global methods. The second one, given in Fig. 5, is more difficult to segment, and the third one, given in Fig. 7, is very noisy.

From the results of Table II, we draw the following remarks.

1) Starting from Gaussian distributions (type 8) GSEM, GEM and GICE all find beta distributions of the first kind (kind 1) for both classes. Furthermore, all parameters get stabilised in these three methods at approximately the same values which can be relatively far from the initialized values. From this we may conjecture, on one hand, that real distributions are best represented by beta distributions of the first kind and, on the other hand, that the parameters are correctly estimated.

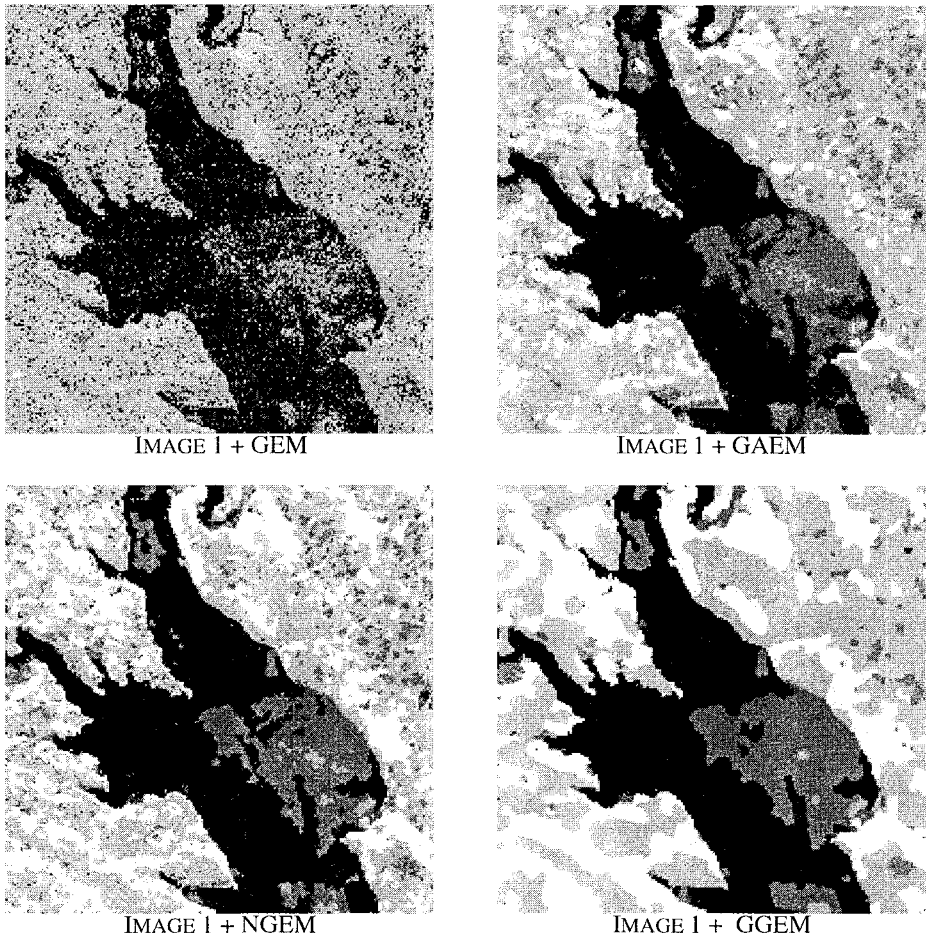

Fig. 6. Segmentation of the Image 1 into four classes by generalized EM (GEM), generalized adaptive EM (GAEM), normal Gibbsian EM (NGEM), and generalized Gibbsian EM (GGEM).

TABLE III
PARAMETERS ESTIMATED FROM IMAGE 2, FIG. 5

	CLASS 1						CLASS 2						
	LOCAL METHODS												
PAR	μ	σ^2	β_1	β_2	π	TYPE	μ	σ^2	β_1	β_2	π	TYPE	ITE
INIT	56	484			0.5	8	78	484			0.5	8	
GSEM	32	109	0.10	3.01	0.15	4	76	684	2.34	7.42	0.85	6	100
GEM	53	495	0.27	3.12	0.44	1	82	760	2.59	7.28	0.56	6	100
GICE	31	102	0.07	3.10	0.14	1	76	686	2.30	7.40	0.86	6	100
GLOBAL METHODS													
PAR	μ	σ^2	β_1	β_2	α	TYPE	μ	σ^2	β_1	β_2	α	TYPE	ITE
INIT	66	686	0.66	5.36	-0.35	4	85	1216	1.10	4.90	-0.35	6	
GGEM	31	118	0.24	3.21	-0.5	1	75	704	2.14	7.19	-0.5	6	30
GGICE	32	123	0.23	3.18	-0.6	1	75	701	2.17	7.23	-0.5	6	30

2) Global methods keep beta distributions of the first kind given with the initialization by GSEM and thus we can imagine that these distributions are well suited to the image considered.

As in the case of the synthetic ring image, the GSEM-based local segmentation, the only one represented on Fig. 4, gives visually slightly better results that the GEM- and GICE-based ones. Parameters are perhaps better estimated but no clear explanation appears when analyzing the results in Table II, apart from the fact that σ_1^2 is close to values estimated by GGEM and GGICE. The use of adaptive GSEM improves the segmentation quality, which approaches the quality of global segmentation methods. The efficiencies of the latter ones appear quite satisfying (see Fig. 4).

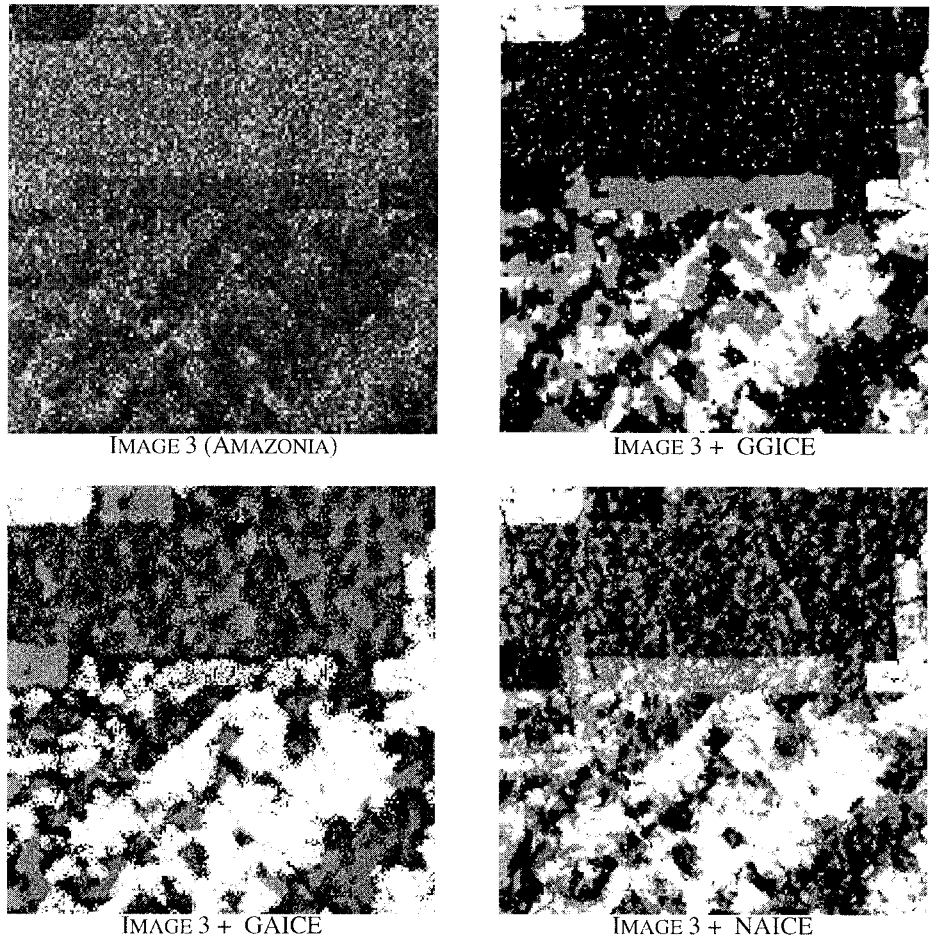

Fig. 7. Image 3 (Amazonia) and its segmentation into three classes by generalized global ICE (GGICE), generalized adaptive ICE (GAICE), and classical adaptive ICE, which uses Gaussian densities (NAICE).

Concerning Image 2, all methods apart GSEM find beta distributions of the first kind (type 1) for the first class distribution and beta of second kind (type 6) for the second class distribution; thus, we can reasonably assume that they are well suited, among distributions of Pearson's system, to real distributions. Image 2 is rather noisy and the difference between global and local methods appears clearly. We note that global methods provide visually better results that the local ones, and, among the latter, the adaptive manner of parameter estimating provides some improvement.

Let us briefly examine how the different methods work in the case of more than two classes.

We present in Fig. 6 the segmentations of the Image 1 into four classes by GEM, GAEM, GGEM, and NGEM respectively. NGEM means "normal Gibbsian," or "normal global" EM, in that no generalized mixture problem is considered and all noise densities are assumed Gaussian. Thus, note that GEM is generalized and local, and NGEM is traditional and global. According to Fig. 6, we note that GEM meanly indicates the presence of two classes and, as in the case of two classes segmentation, the results obtained by GAEM are visually close to the results, which means that the use of generalized mixture estimation instead of the classical Gaussian mixture estimation can have strong influence. Although their comparison is difficult in the absence of the truth of the ground, we may conjecture, as the Gaussian case is a particular case of the generalized one, that the results obtained with GGEM are better. As a curiosity, we note that the results obtained by GAEM look like the results obtained by NGEM.

As a second example, we present in Fig. 7 some results of segmentation into three classes of an ERS 1 image of a forest area of Amazonia. ICE is the basic parameter estimation method used, and we compare GGICE, GAICE, and NAICE. As above, "N" means that only Gaussian densities are used, which means that NAICE is the traditional AICE. Image 3 is very noisy and comparison between the results of the these segmentations is difficult in absence of the ground truth. Only we can say is that GGICE produces a result that seems visually the most consistent.

VI. CONCLUSIONS

We have proposed in this work some new solutions to the problem of generalized mixture estimation, with applications to unsupervised statistical image segmentation. A distribution mixture is said to be "generalized" when the exact nature of components is not known, but each of them belongs to a given finite set of families of distributions. The methods proposed allows one to

1) identify the conditional distribution for each class;
2) estimate the unknown parameters in this distribution;

3) estimate priors;
4) estimate the "true" class image.

Assuming that each of unknown noise probability distribution is in the Pearson system, our methods are based on merging two approaches of classical problems. On the one hand, we use classical mixture estimation methods like EM [5], [9], [31], SEM [24], [26], or ICE [3], [4], [26]–[28]. On the other hand, we use the fact that if we know that the sample considered proceeds from one family in the Pearson system, we dispose of decision rule, based on "skewness" and "kurtosis," for which family it belongs to. Different algorithms proposed are then applied to the problem of unsupervised Bayesian image segmentation in a "local" and "global" way. The results of numerical studies of a synthetic image and some real ones, and other results presented in [22], show the interest of the generalized mixture estimation in the unsupervised image segmentation context. In particular, the mixture components are, in general, correctly estimated.

As possibilities for future work, let us mention the possibility of testing the methods proposed in many problems outside the image segmentation context, like handwriting recognition, speech recognition, or any other statistical problem requiring a mixture recognition. Furthermore, it would undoubtedly be of purpose studying other generalized mixture estimation approaches, based on different decision rules and allowing one to leave the Pearson system.

Acknowledgment

The image of Fig. 3 was provided by CIRAD SA under the project GDR ISIS. The authors thank A. Bégué for her collaboration.

References

[1] R. Azencott, Ed., *Simulated Annealing: Parallelization Techniques*. New York: Wiley, 1992.
[2] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc. B,* vol. 48, pp. 259–302, 1986.
[3] B. Braathen, W. Pieczynski, and P. Masson, "Global and local methods of unsuperivsed Bayesian segmentation of images," *Machine Graph. Vis.*, vol. 2, pp. 39–52, 1993.
[4] H. Caillol, W. Pieczynski, and A. Hillon, "Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation," *IEEE Trans. Image Processing*, vol. 6, pp. 425–440, 1997.
[5] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognit.*, vol. 22, pp. 747–761, 1989.
[6] R. Chellapa and R. L. Kashyap, "Digital image restoration using spatial interaction model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 461–472, 1982.
[7] R. Chellapa and A. Jain, Eds., *Markov Random Fields*. New York: Academic,1993.
[8] Y. Delignon, "Etude statistique d'images radar de la surface de la mer," Ph.D. dissertation, Université de Rennes I, Rennes, France, 1993.
[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. B*, vol. 39, pp. 1–38, 1977.
[10] 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, 1987.
[11] P. A. Devijver, "Hidden Markov mesh random field models in image analysis," in *Advances in Applied Statistics, Statistics and Images: 1*. Abingdon, U.K.: Carfax, 1993, pp. 187–227.
[12] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Stat.*, vol. 16, 1989.
[13] S. Geman and G. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell*, vol. PAMI-6, pp. 721–741, 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, 1997.
[15] X. Guyon, "Champs aléatoires sur un réseau," in *Collection Techniques Stochastiques.* Paris, France: Masson, 1993.
[16] R. Haralick and J. Hyonam, "A context classifier," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-24, pp. 997–1007, 1986.
[17] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions,* vol. 1. New York: Wiley, 1970.
[18] R. L. Kashyap and R. Chellapa, "Estimation and choice of neighbors in spatial-interaction models of images," *IEEE Trans. Inform. Theory*, vol. 29, pp. 60–72, 1983.
[19] P. A. Kelly, H. Derin, and K. D. Hartt, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 36, pp. 1628–1641, 1988.
[20] S. Lakshmanan and H. Herin, "Simultaneous parameter estimation and segmentation of Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 799–813, 1989.
[21] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Stat. Assoc.*, vol. 82, pp. 76–89, 1987.
[22] A. Marzouki, "Segmentation statistique d'images radar," Ph.D. dissertation, Université de Lille I, Lille, France, Nov. 1996.
[23] A. Marzouki, Y. Delignon, and W. Pieczynski, "Adaptive segmentation of SAR images," in *Proc. OCEAN '94*, Brest, France.
[24] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised segmentation of satellite images, *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 618–633, 1993.
[25] E. Mohn, N. Hjort, and G. Storvic, "A simulation study of some contextual clasification methods for remotely sensed data," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-25, pp. 796–804, 1987.
[26] A. Peng and W. Pieczynski, "Adaptive mixture estimation and unsupervised local Bayesian image segmentation," *Graph. Models Image Processing*, vol. 57, pp. 389–399, 1995.
[27] W. Pieczynski, "Mixture of distributions, Markov random fields and unsupervised Bayesian segmentation of images," Tech. Rep. no. 122, L.S.T.A., Université Paris VI, Paris, France, 1990.
[28] ———, "Statistical image segmentation," *Machine Graph. Vis.*, vol. 1, pp. 261–268, 1992.
[29] 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. Stat.*, vol. 16, pp. 267–282, 1989.
[30] H. C. Quelle, J.-M. Boucher, and W. Pieczynski, "Adaptive parameter estimation and unsupervised image segmentation," *Machine Graph. Vis.,* vol. 5, pp. 613–631, 1996.
[31] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, pp. 195–239, 1984.
[32] A. Rosenfeld, Ed., *Image Modeling*. New York: Academic, 1981.
[33] J. Tilton, S. Vardeman, and P. Swain, "Estimation of context for statistical classification of multispectral image data," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-20, pp. 445–452, 1982.
[34] A. Veijanene, "A simulation-based estimator for hidden Markov random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 825–830, 1991.
[35] L. Younes, "Parametric inference for imperfectly observed Gibbsian fields," *Prob. Theory Related Fields*, vol. 82, pp. 625–645, 1989.
[36] J. Zerubia and R. Chellapa, "Mean field annealing using compound Gauss-Markov random fields for edge detection and image estimation," *IEEE Trans. Neural Networks*, vol. 8, pp. 703–709, 1993.
[37] J. Zhang, "The mean field theory in EM procedures for blind Markov random field image restoration," *IEEE Trans. Image Processing*, vol. 2, pp. 27–40, 1993.
[38] 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, 1994.

Yves Delignon received the Ph.D. degree in signal processing from the Université de Rennes 1, Rennes, France, in 1993.

He is currently Associate Professor at the Ecole Nouvelle d'Ingénieurs en Communication, Lille, France. His research interests include statistical modeling, unsupervised segmentation and pattern recognition in case of multisource data.

Abdelwaheb Marzouki received the Ph.D. degree in signal processing from the Université des Sciences et Technologies de Lille, Lille, France, in 1996. His Ph.D. was prepared at the École Nouvelle d'Ingénieurs en Communication, Lille, and the Institut National des Télécommunications, Evry, France.

He is mainly involved with unsupervised statistical segmentation of multisource images.

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

He is currently Professor at the Institut National des Télécommunications, Evry, France. His research interest include mathematical statistics, statistical image processing, and theory of evidence.