Estimation of generalized mixtures and its application in image segmentation

Yves Delignon, Abdelwaheb Marzouki, Wojciech Pieczynski

To cite this version:
Yves Delignon, Abdelwaheb Marzouki, Wojciech Pieczynski. Estimation of generalized mixtures and its application in image segmentation. IEEE Transactions on Image Processing, Institute of Electrical and Electronics Engineers, 1997, 6 (10), pp.1364-1375. 10.1109/83.624951. hal-00681507

HAL Id: hal-00681507
https://hal-imt.archives-ouvertes.fr/hal-00681507
Submitted on 21 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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 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 above 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: Gibbsonian 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 f1,...,fm where the nature of each fi is not known exactly; however, this nature belongs to a given finite set F = {F1,...,Fq} of natures. For instance, if we consider a mixture of two densities f1, f2 each of them being exponential or Gaussian, we have F1 = {exponential laws}, F2 = {Gaussian laws}. There are four possibilities of “classical” mixtures (both f1, f2 exponential, both f1, f2 Gaussian, f1 exponential and f2 Gaussian, f1 Gaussian and f2 exponential) and we do not know in which 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 f1, f2 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]–[12], [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 = (Xs)s∈S, Y = (Ys)s∈S called “random fields”. Each Xs takes its values in a finite set of classes Ω = {ω1,...,ωm} and each Ys 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 = (ys)s∈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 Sβ, we need to know some parameters defining the distribution of (X, Y). The latter distribution is generally defined by PX, the distribution of X and the family PβY|x of the distributions of Y conditional to X. Let us denote by α all parameters concerning PX and by β all parameters concerning the family PβY|x we need. Making the strategy Sβ unsupervised amounts to proposing a way of estimating α and β from Y = y, the only data available. The parameter β is
generally of the form $\beta = (\beta_1, \cdots, \beta_n)$, where $\beta_i$ defines the distribution of $Y_i$ conditional to $X_i = \omega$. If these distributions are Gaussian, which is the most frequently considered case, each $\beta_i$ is of the form $\beta_i = (\mu_i, \sigma_i^2)$ with $\mu_i$ being the mean and $\sigma_i^2$ 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 $X_i = \omega_i, \omega_2$ 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 $\theta = (\pi_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. SEM is an iterative procedure that runs as follows.

1) Initialization: let $\theta^0 = (\pi_1^0, \mu_1^0, \mu_2^0, \sigma_1^2, \sigma_2^2)$ be an initial guess of $\theta = (\pi_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.

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 $1 \leq i \leq n$, the distribution of $X_i$ conditioned on $Y_i = y_i$. If we denote by $f_1^k, f_2^k$ the $\theta^k$ based densities $f_1, f_2$, this distribution is given by

$$
\pi_{1,i}^k = \frac{\pi_1^k f_1^k(y_i)}{\pi_1^k f_1^k(y_i) + \pi_2^k f_2^k(y_i)},
$$

$$
\pi_{2,i}^k = 1 - \pi_{1,i}^k.
$$

b) Sample, for each $1 \leq i \leq n$, 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

$$
[i \in Q_1] \Leftrightarrow [x_i^k = \omega_1] \quad \text{and} \quad [i \in Q_2] \Leftrightarrow [x_i^k = \omega_2].
$$

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

$$
\pi_1^{k+1} = \frac{\text{Card}(Q_1)}{n}, \quad \mu_1^{k+1} = \frac{\sum_{i \in Q_1} y_i}{\text{Card}(Q_1)},
$$

$$
\mu_2^{k+1} = \frac{\sum_{i \in Q_2} y_i}{\text{Card}(Q_2)}, \quad \sigma_1^{k+1} = \frac{\sum_{i \in Q_1} (y_i - \mu_1^{k+1})^2}{\text{Card}(Q_1)}, \quad \sigma_2^{k+1} = \frac{\sum_{i \in Q_2} (y_i - \mu_2^{k+1})^2}{\text{Card}(Q_2)}.
$$

3) Stop when the sequence $(\theta^k)$ 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 $F_1: R^n \rightarrow \Phi$, 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 “$\Phi$ 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 $P[X = \omega_1], P[X = \omega_2]$ 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 $(y_1, \ldots, y_n) \in R^n$ 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 $\Phi$ RECOGNITION

A. System of Pearson

In this section, we specify the family $\Phi$ 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

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

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

$$\mu_1 = E[Y] $$

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

and two parameters $\gamma_1, \gamma_2$ defined by

$$\gamma_1 = \left(\frac{\mu_3}{\mu_2}\right)^2 $$

$$\gamma_2 = \left(\frac{\mu_4}{\mu_2}\right)^2. $$

$\sqrt{\gamma_1}$ is called “skewness” and $\gamma_2$ “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 $\lambda = \gamma_2(\gamma_2 + 3)^2/4(4\gamma_2 - 3\gamma_1)(2\gamma_2 - 3\gamma_1)(2\gamma_2 - 3\gamma_1 - 6)$, the eight families of the set $\Phi = \{F_1, \ldots, F_8\}$, whose exact shape will be given in the next section, are defined by

$$[P_Y \in F_1] \iff [\lambda < 0] $$

$$[P_Y \in F_2] \iff [\gamma_1 = 0 \text{ and } \gamma_2 < 3] $$

$$[P_Y \in F_3] \iff [2\gamma_2 - 3\gamma_1 - 6 = 0] $$

$$[P_Y \in F_4] \iff [0 < \lambda < 1] $$

$$[P_Y \in F_5] \iff [\lambda = 1] $$

$$[P_Y \in F_6] \iff [\lambda > 1] $$

$$[P_Y \in F_7] \iff [\gamma_1 = 0 \text{ and } \gamma_2 > 3] $$

$$[P_Y \in F_8] \iff [\gamma_1 = 0 \text{ and } \gamma_2 = 3]. $$

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

What is important is that moments $\mu_1, \ldots, \mu_4$ can be easily estimated from empirical moments, from which we deduce the estimated values of $\gamma_1, \gamma_2$ 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

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

$$c_0 = \mu_2(4\gamma_2 - 3\gamma_1) - \mu_1(\gamma_2 + 3)\sqrt{\gamma_1} \mu_2 + (\mu_1)^2 (2\gamma_2 - 3\gamma_1 - 6) $$

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

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

(10)

(11)

(12)

(13)
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 \( Y_1, \ldots, Y_n \) 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 \( Q_{1}, \ldots, Q_{n} \) the partition of \( x^n \)
2) For each class \( \omega_i \) use \( Q_i \) in order to estimate \( \mu_{j,i} \) by

\[
\hat{\mu}_{j,i} = \frac{\sum_{y_k \in Q_i} y_k}{\text{Card}(Q_i)} \quad (j = 1, 2, 3, 4) \quad (15)
\]

\[
\hat{\beta}_{j,i} = \frac{\sum_{y_k \in Q_i} (y_k - \hat{\mu}_{j,i})^2}{\text{Card}(Q_i)} \quad \text{for} \quad j = 2, 3, 4. \quad (16)
\]

3) For each class \( \omega_i \) calculate \( \hat{\gamma}_{1,i}, \hat{\gamma}_{2,i} \) from \( \hat{\mu}_{1,i}, \ldots, \hat{\mu}_{4,i} \) according to (9).
4) For each class \( \omega_i \) use \( \hat{\gamma}_{1,i}, \hat{\gamma}_{2,i} \) and (14) to estimate which family among \( F_1, \ldots, F_8 \) the density \( f_i \) belongs to.
5) With the estimated family and the computed \( \alpha, \alpha_0, c_1, c_2 \quad (10)-(13)) \), estimate the parameters of the distribution. (For each \( F_1, \ldots, F_8 \) the exact relationship between density parameters and the computed \( \alpha, \alpha_0, c_1, c_2 \) 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

\[
f_1(y) = \begin{cases} 
\frac{1}{B(p,q)} (y - b_1)^{p-1}(b_2 - y)^{p-1} / (b_2 - b_1)^{2p-2}, & \text{for} \; y \in [b_1, b_2] \\
0, & \text{otherwise} 
\end{cases} \quad (17)
\]

with

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

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:

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

with

\[
b_1 = -a - \frac{1}{2} \sqrt{\Delta} \\
b_2 = \frac{a + b_1}{c_2(b_2 - b_1)} + 1. 
\]

\[
p = \frac{a + b_1}{c_2(b_2 - b_1)} + 1. 
\]

\( F_3 \) (Gamma Distributions): Densities are given by

\[
f_3(y) = \begin{cases} 
\frac{1}{\Gamma(p)} (y - r)^{p-1} e^{-(y-r)/p}, & \text{for} \; y \geq r \\
0, & \text{otherwise} 
\end{cases} \quad (23)
\]

with

\[
p = c_1 \\
q = \frac{c_1}{c_2} - a + 1 \\
r = \frac{c_0}{c_1} \quad (24)
\]

\( F_4 \) (Type IV Distributions): Densities are given by

\[
f_4(y) = K[C_0 + c_2(y + C_1)^2]^{-(1/2c_2)} \exp \left( \frac{-a - C_1}{\sqrt{c_0c_2}} \arctan \left( \frac{\sqrt{c_0c_2}}{C_0 + C_1} \right) \right) \quad (25)
\]

with \( K \) such that \( f_R \int f_4(y) dy = 1 \) and

\[
C_0 = c_0 - \frac{c_1}{4c_2} \\
C_1 = \frac{c_1}{2c_2}.
\]
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

\[
f_5(y) = \begin{cases} \dfrac{p}{1(\gamma)} \left( p(y - r) \right)^{-\gamma - 1} e^{-p/(y-r)} & \text{for } y \geq r \\ 0 & \text{otherwise} \end{cases}
\]

with

\[
p = \dfrac{c_2}{a - c_1 \over 2c_2} \quad q = \dfrac{1}{c_2} - 1 \quad r = -\dfrac{c_1}{2c_2},
\]

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

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

with

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

\[
r = -\dfrac{1}{2c_2} (c_1 - \sqrt{c_1^2 - 4c_0c_2}) \quad s = \sqrt{\dfrac{c_1^2 - 4c_0c_2}{c_2}}
\]

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

\( F_7 \) (Type VII Distributions): Densities are given by

\[
f_7(y) = K [c_0 + c_2 y^2]^{-1/(2c_2)} \cdot \exp \left( -\frac{a}{\sqrt{c_2}} \arctan \left( \dfrac{c_2}{c_0} y \right) \right)
\]

with \( K \) such that \( \int_{-\infty}^{\infty} f_7(y) dy = 1. \)

\( F_8 \) (Gaussian Distributions): Densities are given by

\[
f_8(y) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-(y-\mu)^2/2\sigma^2}
\]

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 \( \pi_{k,1}, \ldots, \pi_{k,n} \) be the distributions \( P[X_1 = \omega_1, Y_1 = y_1], \ldots, P[X_n = \omega_1, Y_n = y_n] \) computed from the current parameter \( \theta_1 \). Priors are reestimated by formula (30), which is the same as that in the EM algorithm, and the \( \Phi \) recognition is the same as that the \( \Phi \) recognition described at the end of Section III-A, with the difference that \( \mu_{j,1}^{k+1}, \mu_{j,2}^{k+1}, \mu_{j,3}^{k+1}, \mu_{j,4}^{k+1} \), given for \( j = 1, 2, 3, 4 \) by formulas (31) and (32), are used instead of those given by formulas (15) and (16).

\[
\pi_{k+1} = \frac{1}{n} (\pi_{k,1}^{k+1} + \cdots + \pi_{k,n}^{k+1})
\]

\[
\mu_{j,1}^{k+1} = \frac{y_1 \pi_{k,1}^{k+1} + \cdots + y_n \pi_{k,n}^{k+1}}{\pi_{k,1}^{k+1} + \cdots + \pi_{k,n}^{k+1}}
\]

\[
\mu_{j,2}^{k+1} = \frac{(y_1 - \mu_{j,1}^{k+1}) \pi_{k,1}^{k+1} + \cdots + (y_n - \mu_{j,1}^{k+1}) \pi_{k,n}^{k+1}}{\pi_{k,1}^{k+1} + \cdots + \pi_{k,n}^{k+1}}
\]

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.
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 Gibssian 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 \( \pi_1, \pi_2 \) be priors and \( f_1, f_2 \) be densities of the distribution of \( Y_s \) conditional to \( X_s \). The blind Bayesian strategy is

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

This strategy is made unsupervised by the direct use of the GSEM algorithm described above: One chooses a sequence of pixels \( s_1, s_2, \ldots, s_n \) and considers that \( y_s \) is the value of the grey level at pixel \( s_s \). 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 \( \pi_1, \pi_2 \) instead of \( \pi_1, \pi_2 \). The GSEM algorithm is modified as follows. Let \( x_1, x_2, \ldots, x_n \) be the sequence obtained by sampling at a given iteration. In GSEM the priors \( \pi_1, \pi_2 \) are reestimated by the frequencies computed using all the sample points; in “adaptive” GSEM one considers, for each \( s_s \), a window \( W_s \) centred at \( s_s \) and \( \pi_1^{W_s}, \pi_2^{W_s} \) are reestimated by frequencies computed from \( (y_{j})_{j \in W_s} \). Let us note that in “adaptive” GSEM the sequence of pixels \( s_1, s_2, \ldots, s_n \) has to cover \( S \). In the following, the generalized adaptive SEM, EM, and ICE will be denoted by GAEM, 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

\[
P_X(x) = c e^{-U_\alpha(x)}
\]

with

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

Thus, \( P_X \) is defined as \( \alpha \). 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 \( \beta_1 = (a^1, c_1^1, c_2^1, c_3^1) \) and \( \beta_2 = (a^2, c_0^2, c_1^2, c_2^2, c_3^2) \), 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 \( \alpha \) 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 \( \alpha \) and probabilities “\( X_s = \omega_1 \) knowing that the neighborhood of \( s \) contains \( \omega_1 \) \( r \) 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 \( \omega_1 \) and two \( \omega_2 \). The probability “\( X_s = \omega_1 \) knowing that the neighborhood of \( s \) contains two \( \omega_1 \)” is estimated by the proportion of the sample giving \( X_s = \omega_1 \). On the other hand this probability is given by

\[
\frac{e^{-2\alpha}}{e^{-2\alpha} + e^{2\alpha}}
\]

which gives an estimated value of \( \alpha \).

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

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

3) Generalized Gibbssian EM (GGEM)

The difference between GICE 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 $Q_1 = \{s \in S/x^{k+1} = \omega_1\}$ and $Q_2 = \{s \in S/x^{k+1} = \omega_2\}$ 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

$$P_{\theta|y} = P_{\theta|y}[X_s = \omega_1/Y = y].$$

(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

$$\mu_{1,1}^{k+1} = \sum_{s \in S} \frac{y_s P_{\theta|y}^{s}}{\sum_{s \in S} P_{\theta|y}^{s}}$$

(38)

and

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

(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 Gibbssian 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 $\mu_1^0 = H^{-1}(1/2)$, $\mu_2^0 = H^{-1}(3/2)$ 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 $\theta^0$. 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 $F_0$ 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; RH: REAL VALUES OF PARAMETERS; $\mu$, $\sigma^2$, $\gamma_1$, $\gamma_2$: MEAN AND VARIANCE; $\gamma_1$, $\gamma_2$: SKEWNESS AND KURTOSIS; TYPE: FAMILY IN PEARSON’S SYSTEM; ERROR: SEGMENTATION ERROR RATE; ITE: NUMBER OF ITERATIONS OF ALGORITHMS. $\alpha$: PARAMETER OF THE MARKOV DISTRIBUTION OF $X_i$. GGICE: GLOBAL GENERALIZED ICE; GGEM: GENERALIZED GIBBSIAN EM  

<table>
<thead>
<tr>
<th>CLASS 1</th>
<th>CLASS 2</th>
<th>LOCAL METHODS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PAR</td>
<td>$\mu$</td>
<td>$\sigma^2$</td>
<td>$\gamma_1$</td>
<td>$\gamma_2$</td>
</tr>
<tr>
<td>TH</td>
<td>130</td>
<td>500</td>
<td>0.28</td>
<td>3.51</td>
</tr>
<tr>
<td>INIT</td>
<td>121</td>
<td>361</td>
<td>0.5</td>
<td>8</td>
</tr>
<tr>
<td>GSEM</td>
<td>130</td>
<td>501</td>
<td>0.32</td>
<td>3.6</td>
</tr>
<tr>
<td>GEM</td>
<td>120</td>
<td>432</td>
<td>0.05</td>
<td>3.69</td>
</tr>
<tr>
<td>GICE</td>
<td>130</td>
<td>491</td>
<td>0.29</td>
<td>3.6</td>
</tr>
</tbody>
</table>

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

| GLOBAL METHODS |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

<table>
<thead>
<tr>
<th>PAR</th>
<th>$\mu$</th>
<th>$\sigma^2$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\pi$</th>
<th>TYPE</th>
<th>$\mu$</th>
<th>$\sigma^2$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\pi$</th>
<th>TYPE</th>
<th>ERR</th>
<th>ITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INIT</td>
<td>130</td>
<td>501</td>
<td>0.32</td>
<td>3.6</td>
<td>-0.49</td>
<td>1</td>
<td>130</td>
<td>497</td>
<td>0.01</td>
<td>2.95</td>
<td>-0.49</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGEM</td>
<td>130</td>
<td>500</td>
<td>0.29</td>
<td>3.55</td>
<td>0.7</td>
<td>6</td>
<td>130</td>
<td>499</td>
<td>0.03</td>
<td>2.94</td>
<td>-0.7</td>
<td>1</td>
<td>4.7</td>
<td>300</td>
</tr>
<tr>
<td>GGICE</td>
<td>130</td>
<td>499</td>
<td>0.27</td>
<td>3.52</td>
<td>-0.9</td>
<td>6</td>
<td>130</td>
<td>501</td>
<td>0.01</td>
<td>3.01</td>
<td>-0.9</td>
<td>1</td>
<td>4.1</td>
<td>300</td>
</tr>
</tbody>
</table>

TABLE II  
PARAMETERS ESTIMATED FROM IMAGE 1, FIG. 3  

<table>
<thead>
<tr>
<th>CLASS 1</th>
<th>CLASS 2</th>
<th>LOCAL METHODS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PAR</td>
<td>$\mu$</td>
<td>$\sigma^2$</td>
<td>$\gamma_1$</td>
<td>$\gamma_2$</td>
</tr>
<tr>
<td>INIT</td>
<td>45</td>
<td>1296</td>
<td>0.5</td>
<td>8</td>
</tr>
<tr>
<td>GSEM</td>
<td>34</td>
<td>125</td>
<td>0.21</td>
<td>3.19</td>
</tr>
<tr>
<td>GEM</td>
<td>34</td>
<td>101</td>
<td>0.30</td>
<td>3.34</td>
</tr>
<tr>
<td>GICE</td>
<td>35</td>
<td>109</td>
<td>0.30</td>
<td>3.30</td>
</tr>
</tbody>
</table>

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

| GLOBAL METHODS |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

<table>
<thead>
<tr>
<th>PAR</th>
<th>$\mu$</th>
<th>$\sigma^2$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\pi$</th>
<th>TYPE</th>
<th>$\mu$</th>
<th>$\sigma^2$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\pi$</th>
<th>TYPE</th>
<th>ERR</th>
<th>ITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INIT</td>
<td>34</td>
<td>125</td>
<td>0.21</td>
<td>3.19</td>
<td>-0.49</td>
<td>1</td>
<td>81</td>
<td>1460</td>
<td>0.91</td>
<td>4.25</td>
<td>-0.49</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGEM</td>
<td>35</td>
<td>125</td>
<td>0.43</td>
<td>3.56</td>
<td>-0.80</td>
<td>1</td>
<td>89</td>
<td>1311</td>
<td>1.11</td>
<td>4.52</td>
<td>-0.80</td>
<td>1</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>GGICE</td>
<td>35</td>
<td>127</td>
<td>0.48</td>
<td>3.69</td>
<td>-0.80</td>
<td>1</td>
<td>88</td>
<td>1317</td>
<td>1.10</td>
<td>4.50</td>
<td>-0.80</td>
<td>1</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

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.

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.
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 $\sigma^2_2$ 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).
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 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


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