Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation

Nathalie Giordana and Wojciech Pieczynski

Abstract—This paper attacks the problem of generalized multisensor mixture estimation. A distribution mixture is said to be generalized when the exact nature of components is not known, but each of them belongs to a finite known set of families of distributions. Estimating such a mixture entails a supplementary difficulty: One must label, for each class and each sensor, the exact nature of the corresponding distribution. Such generalized mixtures have been studied assuming that the components lie in the Pearson system. Adaptations of classical algorithms, such as Expectation-Maximization, Stochastic Expectation-Maximization, or Iterative Conditional Estimation, can then be used to estimate such mixtures in the context of independent identically distributed data and hidden Markov random fields. We propose a more general procedure with applications to estimating generalized multisensor hidden Markov chains. Our proposed method is applied to the problem of unsupervised image segmentation. The method proposed allows one to: (i) identify the conditional distribution for each class and each sensor, (ii) estimate the unknown parameters in this distribution, (iii) estimate priors, and (iv) estimate the "true" class image.

Index Terms—Multisensor data, mixture estimation, generalized mixture estimation, hidden Markov chain, Bayesian segmentation, unsupervised segmentation.

1 Introduction

Hidden Markov chains are a useful tool for tackling numerous concrete problems, for instance in speech processing [27], communications [15], and image processing [1], [8], [25]. These all fall in the framework of estimating some discrete phenomenon from observed noisy data. The noise is often modelled as Gaussian, but in many applications, such as radar, sonar, ultrasound, infrared, or magnetic resonance images, the noise is not necessarily Gaussian [15], [18]. Furthermore, for a given sensor and a given class, the nature of the noise distribution can vary with time. For example, the form of the gray level of the sea surface in radar images can vary with the weather [6]. Thus, it may be desirable to automatically determine the correct noise distribution for each class and each sensor at a given time. Early algorithms treating this problem were proposed in [6] and [26] and applied to unsupervised image segmentation. They combine mixture estimation algorithms such as Expectation-Maximization (EM) [7], [28], Stochastic Expectation-Maximization (SEM) [20], [22], or Iterative Conditional Estimation (ICE) [23], [24], with the recognition of the form of a distribution in the Pearson system, assuming that a given sample is generated from a unique distribution.

The present paper lies within the scope of this general problem. We first propose a multisensor generalized mixture estimation method based on ICE and valid in a general hidden data context. We then tailor our method to generalized multisensor hidden Markov chain estimation. The effectiveness of the proposed method is validated by simulations.

The algorithms are then applied to the problem of unsupervised image segmentation. In image segmentation, statistical methods are based on random field models: For the set of pixels S, we consider two sets of random variables $X = (X_s)_{s\in S}$, $Y = (Y_s)_{s\in S}$ called "random fields." Each X_s takes its values in a finite set of classes $\Omega = \{\omega_1, ..., \omega_k\}$ and each Y_s takes its values in $\mathbb{R}$. The segmentation problem is then to estimate 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 observed image. There are two families of methods: global methods, which use hidden Markov field models [2], [4], [5], [9], [11], [16], [19], [31], [32], and local methods, in which each pixel is classified from observations of a local neighborhood [13], [20], [21], [22], [30]. The efficiency of the global methods is striking in many cases, although local methods still remain of interest [3]. We proposed in [1] a third method, which uses hidden Markov chains instead of hidden Markov fields, with transformation of the two-dimensional set of pixels to a one-dimensional set using Hilbert-Peano scans [29]. The results are comparable to those obtained with hidden Markov field based methods, with the added benefit of being sufficiently versatile to treat spatio-temporal unsupervised segmentation problems. The use of Markov chains instead of Markov fields also affords computational advantages. Indeed, several relevant distributions can be calculated analytically in

• *The authors are with the Institut National des Télécommunications, Département Signal et Image, 9 rue Charles Fourier, 91000 EVRY - France. E-mail: Wojciech.Pieczynski@int-evry.fr.*

Manuscript received 13 Nov. 1995; revised 2 Jan. 1996. Recommended for acceptance by D.M. Titterington.
For information on obtaining reprints of this article, please send e-mail to: transpami@computer.org, and reference IEEECS Log Number P97005.

the case of the Markov chains, whereas iterative estimation procedures, like the Gibbs sampler, would be required using Markov fields. Parameter estimation and restoration algorithms [1], [10] are thus faster in the Markov chain case. Furthermore, the use of the EM algorithm is complicated in the hidden Markov field context and modifications may be required [4], [32].

The paper is organized as follows: In the next section we address the generalized mixture estimation problem in a broad setting, in order to present our ICE-based method. Section 3 describes the hidden Markov chain model. In Section 4 we present some particular methods and simulation results in the case of monosensor hidden Markov chains, with the multisensor case treated in Section 5. Applications to the problem of unsupervised statistical image segmentation are described in Section 6, while Section 7 contains the conclusions.

2 Generalized Mixture Estimation

Let us consider a finite set S and random variables $(X, Y) = ((X_s)_{s\in S}, (Y_s)_{s\in S})$. We consider first the monosensor case. Thus each X_s takes its values in Ω and each Y_s takes its values in $\mathbb{R}$. The distribution of X depends on a parameter α and is denoted by π_α. The random variables $(Y_s)_{s\in S}$ are independent conditionally on X and the distribution of each Y_s conditional on X is equal to its distribution conditional on X_s. Thus all distributions of Y conditional on X are determined by the k distributions of Y_s conditional on each of $\omega_1, ..., \omega_k$, respectively, specified by densities $f_1, ..., f_k$ with respect to the Lebesgue measure. The problem of mixture estimation is to find α and $f_1, ..., f_k$ from $Y = y$. In the "classical" mixture case the general forms of $f_1, ..., f_k$ are known and they depend on some parameter β which is to be estimated from $Y = y$. For instance, if $f_1, ..., f_k$ are Gaussian, β contains k means and k variances. In the "generalized" mixture case, by contrast, the general form of $f_1, ..., f_k$ is not known exactly; however, the form of each f_j is assumed to belong to a given finite set of forms. To be more precise, let $\Psi = \{F_1, ..., F_M\}$ be a set of families of distributions. For instance, F_1 may be Gaussian distributions, F_2 gamma distributions, and so on. Then each f_j belongs to one of the families $F_1, ..., F_M$, but we do not know which. The problem of determining of $f_1, ..., f_k$ is then two-fold: For each f_j find the family F_i to which f_j belongs, and then find the parameter which fixes f_j in F_i.

We propose a general algorithm, called ICE-GEMI (GEMI for generalized mixture), to solve such problems, based on the following assumptions.

- A_1) An estimator $\hat{\alpha} = \hat{\alpha}(X)$ of α from X is available.
- A_2) One may simulate realizations of X according to its distribution conditional on Y.
- A_3) Each family F_j of $\Psi = \{F_1, ..., F_M\}$ is characterized by a parameter β^j, i.e., $F_j = \left\{f_{\beta^j}\right\}_{\beta^j \in B^j}$. In practice, B^j is a subset of $\mathbb{R}^{n_j}$ with n_j depending on F_j: For instance, $n_j = 2$ if F_j are Gaussian.
- A_4) M estimators $\hat{\beta}^1, ..., \hat{\beta}^M$ are available such that if a sample $z = (z_1, ..., z_r)$ is generated by a distribution f_{β^j} in F_j, then $\hat{\beta}^j = \hat{\beta}^j(z)$ estimates β^j.
- A_5) A decision rule D is available, such that, for any sample $z = (z_1, ..., z_r)$ and any $(f_1, ..., f_M) \in F_1 \times ... \times F_M$, the rule D associates with z the "best suited" density among $f_1, ..., f_M$, according to some criterion.

We call our method ICE-GEMI as it can be seen as a generalization of ICE, which in turn is a general method for estimating hidden models and, in particular, for estimating classical mixtures [23], [24].

The ICE-GEMI algorithm is an iterative method: at step q, let α^q and $f_1^q, ..., f_k^q$ be current prior parameters and current densities $f_1, ..., f_k$. The updating is as follows.

a) Simulate x^q, a realization of X, according to its distribution, conditional on $Y = y$ and based on α^q and $f_1^q, ..., f_k^q$.

b) Calculate $\alpha^{q+1} = E_q\left[\hat{\alpha}(X) \middle| Y = y\right]$, where $E_q[. \mid Y = y]$ denotes the conditional expectation given $\alpha = \alpha^q$ and $(f_1, ..., f_k) = \left(f_1^q, ..., f_k^q\right)$. If this calculation is impossible, calculate $\alpha^{q+1} = \hat{\alpha}\left(x^q\right)$.

c) For $i = 1, ..., k$, consider $S_i^q = \left\{s \in S \middle| x_s^q = \omega_i\right\}$. Let $y_i^q = \left(y_s\right)_{s\in S_i^q}$. For each $i = 1, ..., k$ estimate the M parameters $\beta_i^1 = \hat{\beta}^1\left(y_i^q\right), ..., \beta_i^M = \hat{\beta}^M\left(y_i^q\right)$.

d) For $i = 1, ..., k$, consider $D\left(y_i^q\right) \in \left\{f_{\beta_i^1}, ..., f_{\beta_i^M}\right\}$.

e) Update $f_1, ..., f_k$ by putting

$$\left(f_1^{q+1}, ..., f_k^{q+1}\right) = \left(D\left(y_1^q\right), ..., D\left(y_k^q\right)\right)$$

In the multisensor case each Y_s takes its values in $\mathbb{R}^m$. Thus $Y_s = \left(Y_s^1, ..., Y_s^m\right)$. We shall assume

- A_6) Random variables $Y_s^1, ..., Y_s^m$ are independent conditionally on X_s.

This means that for a given class the observations in different sensors are independent. Thus each f_i of densities $f_1, ..., f_k$, which are densities on $\mathbb{R}^m$, is given by m densities $f_i^1, ..., f_i^m$ on $\mathbb{R}$:

$$f_i\left(y_s^1, ..., y_s^m\right) = f_i^1\left(y_s^1\right) \times ... \times f_i^m\left(y_s^m\right) \quad (1)$$

Note that in many practical situations (A_6) is probably too strong an assertion. Taking into account the stochastic dependence among sensors, however, is nontrivial in the context of our model and further study would be necessary to solve this problem.

The ICE-GEMI algorithm in the multisensor case differs

little from the ICE-GEMI algorithm in the monosensor case. Roughly speaking, all formulas of the model stay valid, except that $f_i(y_s)$ is replaced by $f_i\left(y_s^1, \ldots, y_s^m\right)$. Thus, steps (a) and (b) of the ICE-GEMI described above remain. Steps (c), (d), and (e) differ slightly, as follows. They become:

c) For $i = 1, \ldots, k$, consider $S_i^q = \left\{s \in S \middle| x_s^q = \omega_i\right\}$. Let $y_i^q = \left(y_s\right)_{s \in S_i^q} = \left(y_s^1, \ldots, y_s^m\right)_{s \in S_i^q}$ and $y_i^{q,r} = \left(y_s^r\right)_{s \in S_i^q}$. For each sensor $r = 1, \ldots, m$ and each class $i = 1, \ldots, k$, calculate M parameters

$$\beta_i^{1,r} = \hat{\beta}^1\left(y_i^r\right), \ldots, \beta_i^{M,r} = \hat{\beta}^M\left(y_i^r\right).$$

d) Apply (d), above, to each sensor, giving

$$\beta_i^{1,q+1}, \ldots, \beta_i^{m,q+1}.$$

e) For each class $i = 1, \ldots, k$, consider

$$f_i^{q+1} = f_{\beta_i^{1,q+1}} \times \ldots \times f_{\beta_i^{m,q+1}}$$

and update $(f_1, \ldots, f_k)$ with $f_1^{q+1}, \ldots, f_k^{q+1}$.

Once the sets $S_i^q = \left\{s \in S \middle| x_s^q = \omega_i\right\}$ are known, the workload for m sensors is m times that of the monosensor case, resulting in m monosensor updates followed by the m-sensor update using the product of monosensor functions so obtained. We should remark that this applies only to the noise parameter updates, and that the complete m-sensor estimation algorithm cannot be reduced to the use of the monosensor algorithm m times; the latter procedure would give, in particular, m different priors.

3 Hidden Markov Chains

We present here a brief review of the hidden Markov chain model; for the sake of simplicity, we describe it in the monosensor case, the generalization to the multisensor case being immediate.

We shall assume that $(X_n)_{n \in \mathbb{N}^*}$ is a Markov chain, with each $X_n \in \{\omega_1, \ldots, \omega_k\}$, and with stationary transition probabilities. We assume that

$$c_{ij} = P[X_n = \omega_i, X_{n+1} = \omega_j] \tag{2}$$

does not depend on n. Thus the initial distribution is given by

$$\pi_i = P\left[X_1 = \omega_i\right] = \sum_{1 \le j \le k} c_{ij} \tag{3}$$

and the transition matrix $A = [a_{ij}]$ has entries

$$a_{ij} = P\left[X_{n+1} = \omega_j \middle| X_n = \omega_i\right] = \frac{c_{ij}}{\sum_{1 \le j \le k} c_{ij}} \tag{4}$$

The k^2 parameters $(c_{ij})_{1 \le i \le k, 1 \le j \le k}$ entirely define, by virtue of Kolmogorov's theorem, the distribution of X.

We assume that the conditional structure of the $(Y_1, \ldots, Y_n)$, given the $(X_1, \ldots, X_n)$, as described at the beginning of Section 2.

We will denote $(X_1, \ldots, X_n)$ by X, $(Y_1, \ldots, Y_n)$ by Y, and their realizations by $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

Let

$$\alpha_t(i) = P[X_t = \omega_i, Y_1 = y_1, \ldots, Y_t = y_t] \tag{5}$$

and

$$\beta_t(i) = P[Y_{t+1} = y_{t+1}, \ldots, Y_n = y_n \mid X_t = \omega_i] \tag{6}$$

the so-called forward and backward probabilities, which can be calculated by the following forward and backward recursions:

- Initialization:

$$\alpha_1(i) = \pi_i f_i(y_1) \text{ for } 1 \le i \le k;$$

- Induction:

$$\alpha_{t+1}(j) = \left(\sum_{i=1}^{k} \alpha_t(i) a_{ij}\right) f_j\left(y_{t+1}\right)$$
$$\text{for } 1 \le i \le k, \quad 1 \le t \le n-1 \tag{7}$$

- Initialization:

$$\beta_n(i) = 1 \text{ for } 1 \le i \le k$$

- Induction:

$$\beta_t(i) = \sum_{j=1}^{k} a_{ij} f_j\left(y_{t+1}\right) \beta_{t+1}(j) \tag{8}$$
$$\text{for } 1 \le i \le k, \quad 1 \le t \le n-1$$

Also, let

$$\Psi_t(i, j) = P\left[X_t = \omega_i, X_{t+1} = \omega_j \middle| Y = y\right] = \frac{\alpha_t(i) a_{ij} f_j\left(y_{t+1}\right) \beta_{t+1}(j)}{\sum_{l=1}^{k} \sum_{m=1}^{k} \alpha_t(l) a_{lm} f_m\left(y_{t+1}\right) \beta_{t+1}(m)} \tag{9}$$

Thus

$$\xi_t(i) = P\left[X_t = \omega_i \middle| Y = y\right] = \frac{\alpha_t(i) \beta_t(i)}{\sum_{l=1}^{k} \alpha_t(l) \beta_t(l)} \tag{10}$$

indicating that $\Psi_t(i, j)$ and $\xi_t(i)$ are computable using forward and backward procedures. For the multisensor hidden Markov chain model, all formulas stay valid, except that $f_i(y_s)$ is replaced by $f_i\left(y_s^1, \ldots, y_s^m\right)$.

4 Estimation of Generalized Hidden Markov Chains

4.1 ICE-GEMI Algorithm

We consider first the monosensor case, and verify the assumptions (A_1)-(A_5).

(A_1) The parameter α is given here by (2): $\alpha = (c_{ij})_{1 \le i,j \le k}$, and it is possible to use the empirical frequencies as estimators:

$$\hat{c}_{ij}(X) = \frac{1}{n-1} \sum_{t=1}^{n-1} 1_{\left[X_t = \omega_i, X_{t+1} = \omega_j\right]} \tag{11}$$

(A_2) It can be shown that, conditional on $Y = y$, X is a nonstationary Markov chain with transition matrix at time t given by

$$a_{ij}^t = \frac{\Psi_t(i, j)}{\xi_t(i)} \tag{12}$$

where $\Psi_t(i, j)$ and $\xi_t(i)$ are defined in (9) and (10). Thus X can be simulated.

(A_3) and (A_4): There are numerous standard families Ψ verifying these conditions.

(A_5) Let $y = (y_1, ..., y_n)$ be a sample and $(f_1, ..., f_M) \in F_1 \times \ldots \times F_M$. Let $\hat{G}_n$ be the empirical cumulative distribution function $\hat{G}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty\ \ y_i]}(t)$, and let $G_1, ..., G_M$ be the cumulative distribution functions associated with the distributions $f_1, ..., f_M$. Then one can base the choice of an f_j on any measure of distance between $\hat{G}_n$ and the G_j. For instance, let

$$K_j(y) = \sup_{t\in\mathbb{R}} \left|\hat{G}_n(t) - G_j(t)\right|$$

which is the Kolmogorov distance between $\hat{G}_n$ and G_j, and define D by

$$\left[D(y) = f_i\right] \Leftrightarrow \left[K_i(y) = \inf_{1\le j\le M} K_j(y)\right] \tag{13}$$

With (A_1)-(A_5) verified, we may apply the corresponding ICE-GEMI algorithm by following (a)-(e) of Section 2. Note that in step (b) the calculation of $\alpha^{q+1} = E_q\left[\hat{\alpha}(X)\middle|Y = y\right]$ is feasible. In fact, for each c_{ij} we have

$$c_{ij}^{q+1} = E_q\left[\hat{c}_{ij}(X)\middle|Y = y\right] = \frac{1}{n-1}\sum_{t=1}^{n-1} \Psi_t^q(i, j) \tag{14}$$

which results from (9) and (11).

The multisensor case is analogous, just replacing densities on $\mathbb{R}$ by densities on $\mathbb{R}^m$.

In the sequel we denote by ICE-KOLM the ICE-GEMI based on the minimization of the Kolmogorov distance above.

4.2 ICE-PEAR Algorithm

The algorithms proposed in our earlier work in [6], [26] use the Pearson family, which contains eight subfamilies. The first four moments of Pearson a distribution determine both the subfamily to which it belongs and the values of the parameters [17]. Methods in [26] were based on the SEM algorithm, which is a stochastic variant of the EM algorithm that can be used to estimate classical mixtures [20]. ICE-based procedures, which we call ICE-PEAR here, were later proposed in [6], with applications to independently identically distributed samples and hidden Markov random fields. For the model considered in this paper, ICE-PEAR is slightly different. It resembles ICE-GEMI save for steps (c) and (d) which become the following:

(c*) For $i = 1, ..., k$, consider $S_i^q = \left\{s \in S\middle|x_s^q = \omega_i\right\}$. Calculate the four first moments from $y^q = \left(y_s\right)_{s\in S_i^q}$ and decide in which family $F_{j(i)} \in \{F_1, ..., F_8\}$ the distribution lies. Calculate $\beta_i^{j(i)}$ from the four first moments and set $\beta_i^{q+1} = \beta_i^{j(i)}$.

4.3 ICE-KOLM, ICE-PEAR, and ICE-GAUSS: Numerical Results

We present some numerical results of unsupervised Maximum Posterior Mode (MPM) restoration using our methods. We compare their efficiency, in cases involving non-Gaussian noise, with a classical method that assumes that the noise is Gaussian. The classical method used is ICE-based MPM, so that, when the noise is Gaussian, both approaches are the same theoretically. We also compare ICE-KOLM and ICE-PEAR.

The procedure for the MPM restoration is as follows:

i) for each $t \in \{1, ..., n\}$ and $i \in \{1, ..., k\}$ calculate $\xi_t(i)$ (see (10), Section 3);
ii) for each $t \in \{1, ..., n\}$ let ω^t be a class maximizing $\xi_t(i)$ with respect to $i \in \{1, ..., k\}$;
iii) take $\hat{x} = \left(\omega^1, \ldots, \omega^n\right) = MPM(y)$.

Although we have chosen the MPM algorithm, the so-called maximum a posteriori (MAP) algorithm could also be applied. MAP is based on the principle

$$\hat{x} = MAP(y) \Leftrightarrow \hat{x} \arg\max_x P\left[X = x\middle|Y = y\right] \tag{15}$$

With the analytic solution given by the Viterbi algorithm [10]. The hidden Markov chain model enjoys a computational advantage over the hidden Markov field model when MPM is used at the restoration step, and this is even more pronounced when using MAP. With the hidden Markov field model, one must resort to simulated annealing algorithm [11], which can be very time consuming. Note that Iterated Conditional Mode (ICM [2]) can be used as a fast approximation to MAP, but then the convergence is not ensured and the restoration result depends strongly on the initialization.

To initialize the three methods we take $c_{ij}^0 = \frac{1}{2k}$ if $i = j$, and $c_{ij}^0 = \frac{1}{2k(k-1)}$ if $i \ne j$ for $i, j \in \{1, \ldots, k\}$. We calculate the empirical mean μ and the empirical variance σ^2 of the sample, and all noise distributions are assumed Gaussian with variances equal to σ^2, and means distributed around μ and distanced $\frac{\sigma}{2}$ from each other.

We take $\Psi = \{F_1, F_2, F_3\}$, where F_1 denotes beta distributions of the first kind [14], F_2 denotes gamma distributions [14], and F_3 denotes Gaussian distributions, along with the Markov chain defined by $(c_{11}, c_{12}, c_{21}, c_{22}) = (0.49, 0.01, 0.01, 0.49)$. The common marginal distribution is $(\pi_1, \pi_2) = (0.5, 0.5)$ and the transition matrix is $A = \begin{pmatrix} 0{,}98 & 0{,}02 \\ 0{,}02 & 0{,}98 \end{pmatrix}$.

Observe that the initialization we use gives $\left(c_{11}^0, c_{12}^0, c_{21}^0, c_{22}^0\right) = (0.25, 0.25, 0.25, 0.25)$, which is far from the true values. However, the results below show that this

TABLE 1
GENERALIZED MIXTURE ESTIMATION
AND UNSUPERVISED RESTORATION

Noise 2	Class 1	Class 2	τ_{MPM}
Real	B(7.3, 0.200)	Γ(8.10, 100)	2.14 %
ICE-GAUS	N(141.768)	N(179.814)	4.8 %
ICE-PEAR	B(7.3; 3.8, 46, 357)	N(184.843)	2.7 %
ICE-KOLM	B(6.7, 3.4, 18,208)	Γ(10, 8, 100)	2.5 %
τ_{blind}			31.3 %
NOISE 3	CLASS 1	CLASS 2	TMPM
Real	Γ(2,10,100)	Γ(4,10,100)	3.4 %
ICE-GAUS	N(117,96)	N(141,397)	15.6 %
ICE-PEAR	B(1.8, 5.5, 134, 200)	N(155, 490)	11.3 %
ICE-KOLM	Γ(1.9, 10.2, 100)	Γ(3.9, 10.2, 99)	7.2 %
τ_{blind}			27.5

N, *Γ*, B: *Gaussian, gamma, and beta of the first kind distributions respectively.* τ_{MPM}*: the error rate of the MPM restoration. Real: real distributions. ICE-GAUS: estimation with ICE of a "classical" Gaussian mixture. ICE-PEAR and ICE-KOLM: generalized mixture estimation algorithms.* τ_{blind}*: theoretical error rates of blind restoration of the two noises considered.*

TABLE 2
RECOGNITION OF THE TWO SENSOR GENERALIZED MIXTURES
AND THE TWO SENSOR ERROR RATES OF THE MPM

	Sensor 1		Sensor 2		
	Class 1	Class 2	Class 1	Class 2	τ_{MPM}
Real	Γ(4, 10, 100)	N(180, 400)	Γ(8, 10, 100)	B(7, 3, 0, 200)	0.46 %
ICE-GAUS	N(140, 379)	N(180, 405)	N(140, 764)	N(179, 855)	3.14 %
ICE-KOLM	Γ(3.5, 12, 98)	N(180, 400)	Γ(10.2, 8, 95)	B(6.3, 2.3, 36, 198)	1.45 %
Real	Γ(2, 10, 100)	Γ(4, 10, 100)	*B*(9, 2, 0, 255)	B(7, 3, 0, 255)	1.9 %
ICE-GAUS	N(120, 158)	N(140, 408)	N(178, 128)	N(209, 709)	8.8 %
ICE-KOLM	Γ(2.4, 9.5, 98)	Γ(4, 9.8, 100)	*B*(4.3, 2.8, 18, 288)	*B*(2.8, 1.5, 125, 258)	7.3 %

Based on real parameters or estimates with ICE-GAUS and ICE-KOLM.

poor initialization has little effect on the efficiency of the different methods. The chain is corrupted by two noise sequences (Noise 2, Noise 3), whose distributions are presented in Fig. 1. The noises Noise 1 and Noise 4 are used in Section 6.2 and Section 6.4. Their influence is measured by the theoretical blind error rates τ_{blind}, which are theoretical Bayes error rates when using the true parameters and estimating each x_t from y_t, i.e., without reference to any context.

In order to generate different noises used we apply methods described in [17]. Table 1 and other results presented in [12] lead to the following remarks:

1) The noises used are rather strong. Comparing τ_{blind} to different τ_{MPM}s obtained with real parameters, one can note the advantage in using the Markovian model considered.
2) ICE-KOLM is quite efficient in recognizing the nature of the mixture components, while ICE-PEAR can encounter some difficulties.
3) Once the correct components are determined, the noise parameters appear correctly estimated.
4) The restoration results obtained from generalized mixture estimation are always better than those obtained from Gaussian mixture estimation.
5) The error rates of MPM based on the ICE-KOLM estimates are often very close to the error rates of MPM based on the true parameters, which attests to the stability of the whole procedure.

5 MULTISENSOR GENERALIZED HIDDEN MARKOV CHAINS

The restoration of multisensor hidden Markov chains is of interest in many situations. For example, in image processing, sensors can be of different natures, such as radar, infrared, optical, or ultrasound, and thus the distributions corresponding to a given class can have different natures. Also, the nature of the noise can vary with the class, with the sensor for a given class and, for a given class and sensor, with time. We briefly present in this section some numerical results comparing ICE-KOLM and ICE-GAUS in the multisensor case. Table 2 and additional results presented in [12] lead to the following remarks about two-sensor experiments:

1) Again, results obtained from generalized mixture estimation are better than those obtained from Gaussian mixture estimation.

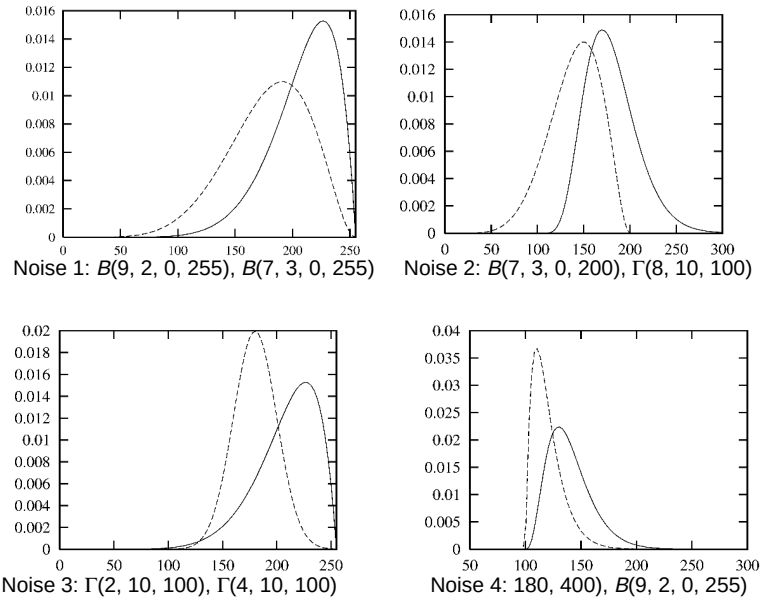


Fig. 1. Noise distributions used in simulations.

2) The error rates of MPM based on the ICE-KOLM estimates are still close to the error rates of MPM based on the true parameters; however, this is less striking than in the monosensor case. Stability of the whole procedure is retained. When there are more than two sensors, the problem of choosing a maximum number of "useful" sensors could arise.

6 Unsupervised Generalized Image Segmentation

We present in this section, some applications of generalized mixture estimation, with ICE-KOLM and ICE-PEAR, to unsupervised image segmentation. As we use the Markov hidden chain model, all results of the previous sections apply almost immediately. To transform a set of pixels into a support set for a hidden Markov chain, we proposed in [1] the use of the Hilbert-Peano scan and showed that results obtained, in the case of Gaussian noise, were better than those obtained using classical raster scans. The model is intuitively less satisfying than the hidden Markov field model, but several simulations performed in the Gaussian case reveal that it is competitive [1].

6.1 Monosensor Image Segmentation

Let S be the set of pixels, and s_1, ..., s_n be pixels ordered according to the Hilbert-Peano curve, the first three stages of whose construction in S is presented in Fig. 2, starting with a four pixel image, and at each step multiplying the image by four. Continuation of this sequence creates the Hilbert-Peano scan on an image of size $2^n \times 2^n$, for any n. Other Hilbert-Peano scans can be defined on images of any size [29].

Thus $X = (X_s)_{s \in S}$ is the random field of classes and $Y = (Y_s)_{s \in S}$ is the observed field, i.e., $Y = y$ is the digital image to be segmented. Considering

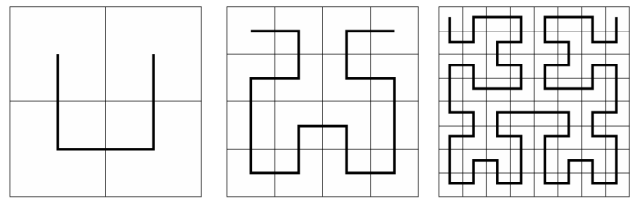


Fig. 2. Construction of the Hilbert-Peano scan.

$$(X_1, \ldots, X_n) = (X_{s_1}, \ldots, X_{s_n})$$

as a Markov chain and making the usual assumptions about

$$(Y_1, \ldots, Y_n) = (Y_{s_1}, \ldots, Y_{s_n})$$

we see that problem of image segmentation is now one of generalized hidden Markov chain restoration.

6.2 Synthetic Images

Consider the family $\Psi = \{F_1, F_2, F_3\}$ of Section 4, and two synthetic images: "Letter B" and "Ring." Each is corrupted by Noise 3 and Noise 4 of Section 4.3.

As mentioned above (Section 4.3, Table 1), the noise perturbations are quite sizeable: for priors equal to 0.5, the blind classification error rates are 31.3 percent and 27.5 percent, respectively. Images, their noisy versions with Noise 4, and segmentation results are presented in Fig. 3. Tables 3 and 4 display the families estimated by ICE-KOLM and ICE-PEAR and the error rates given by MPM based on the true distributions, ICE-GAUS-MPM, ICE-PEAR-MPM, and ICE-KOLM-MPM. We also present the results obtained with an unsupervised hidden Markov-field based algorithm, called ICE-FIELD, in which the parameter-estimation step is done by ICE and the segmentation step with MPM. Note that we use the Ising model, which is the simplest one, and more complex models could produce better results.

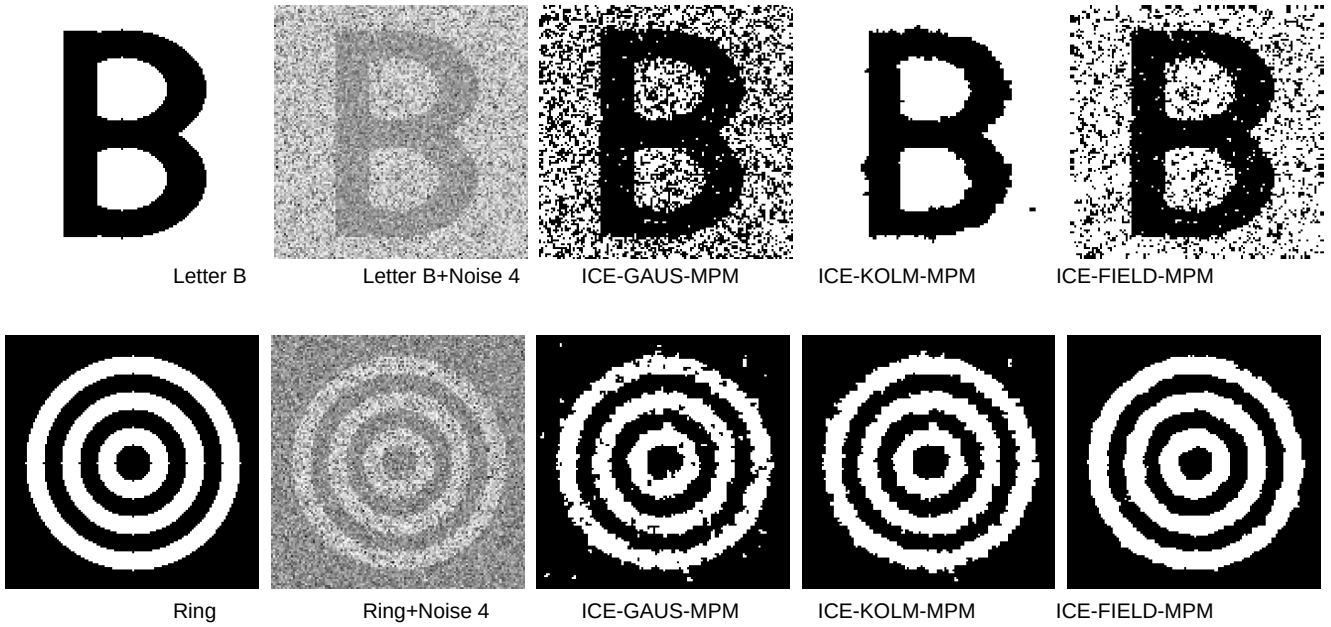


Fig. 3. Segmentations of "Letter B" and "Ring."

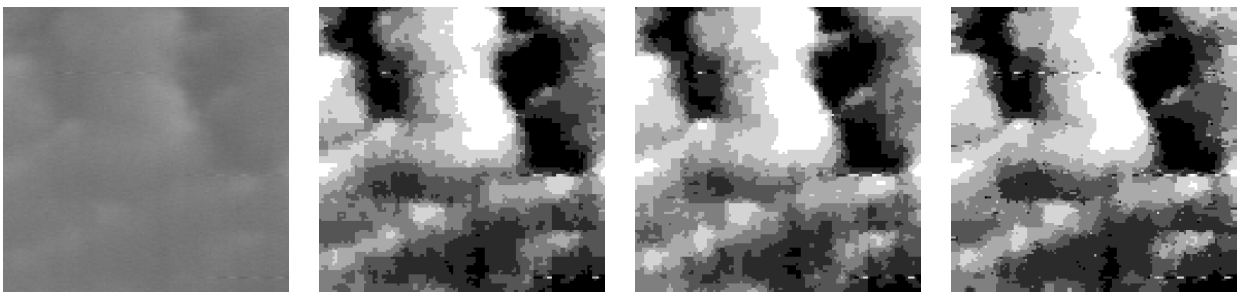


Fig. 4. Segmentations of clouds (seven classes). ICE-KOLM detects one beta distribution and six normal distributions.

Note that ICE-KOLM is always more efficient than ICE-PEAR. This advantage is generally slight, but can also be pronounced, as in the Letter B+Noise 3 case. On the other hand, both ICE-KOLM and ICE-PEAR perform better than ICE-GAUS. In some cases, such as Letter B+Noise 4 and Ring+Noise 3, ICE-GAUS gives very poor results and the advantage of the generalized mixture based methods is quite apparent. Otherwise, these two simple examples show that we have to be very careful in comparing our method with the classical methods based on hidden Markov fields with Gaussian noises. The former clearly gives better results in the case of Letter B, but the second takes the upper hand in the case of Ring. As the noise is the same in both cases, this reveals that the structure of images, which determines the prior distribution, plays an important role.

6.3 Real Images

We now present the results of different segmentations of three real images. Our purpose is two-fold; on the one hand we show that ICE-KOLM yields better results than ICE-GAUS, and on the other hand, we compare the efficiency of ICE-KOLM with that of ICE-FIELD. We do not address the important problem of estimating the number of classes in this paper, although we refer to [20] for a procedure, proposed using estimating classical mixtures with SEM, that can be quite useful for fixed images. According to the different visual impressions, we can say that no clear general tendency appears and the hierarchy of efficiencies of the methods considered is subject to each particular image. Concerning the image Clouds, Fig. 4, the efficiency of the three methods seems comparable. As nearly all distributions detected by ICE-KOLM are normal, the equivalence between ICE-KOLM and ICE-GAUS is not surprising. The fact that the results obtained with both ICE-KOLM and ICE-GAUS are comparable to results obtained with ICE-FIELD indicates that they are competitive and should be used in images of this kind because of their speed. In "San Francisco" image, Fig. 5, ICE-GAUS gives clearly better results than ICE-FIELD, which reflects how hidden Markov chains can be suitable even outside the generalized mixture considerations. This could be due to the fact, as noted in [3], that the hidden Markov field based methods can encounter difficulties in detecting very fine details and the hidden Markov chain based methods seem to be better suited to such situations. Furthermore, ICE-KOLM-MPM is clearly more effective than ICE-GAUS-MPM. This is undoubtedly due to the fact that ICE-KOLM detects beta and gamma distributions and leads us to two conclusions. First, noise of different forms associated with different classes can exist in

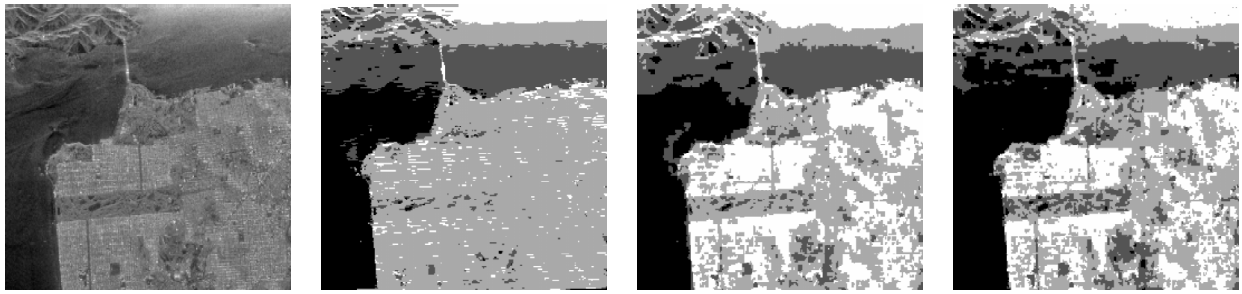


Image "San Francisco" ICE-FIELD-MPM ICE-GAUS-MPM ICE-KOLM-MPM

Fig. 5. Segmentations of San Francisco (four classes). ICE-KOLM detects one normal, one beta, and two gamma distributions.

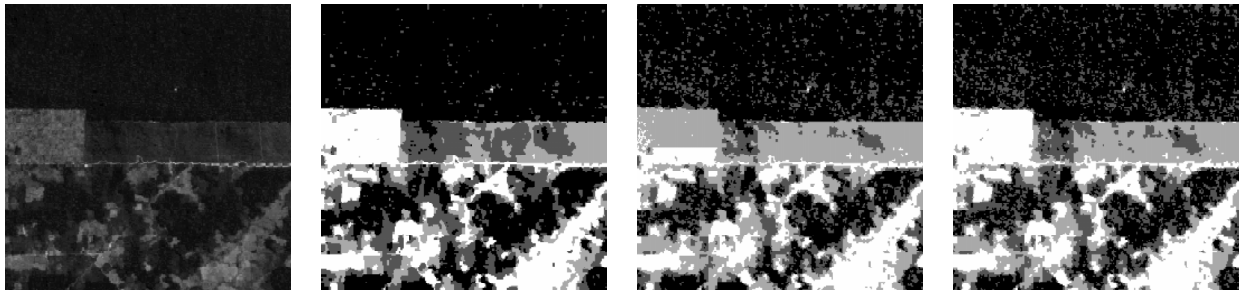


Image "Amazonia" ICE-FIELD-MPM ICE-GAUS-MPM ICE-KOLM-MPM

Fig. 6. Segmentations of Amazonia (four classes). ICE-KOLM detects one normal, one beta, and two gamma distributions.

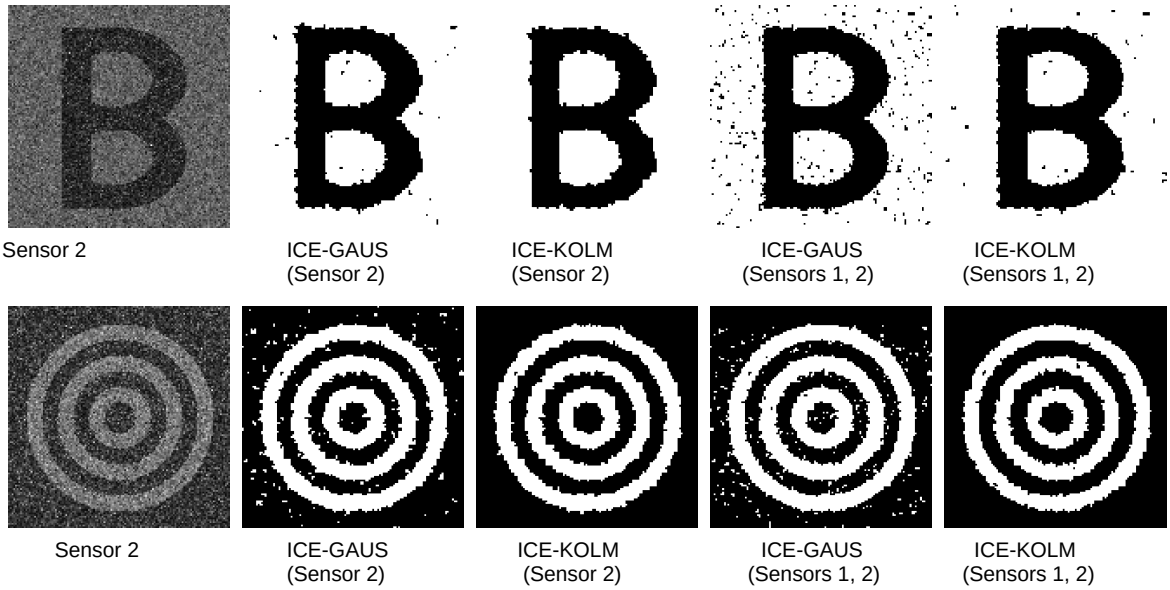


Fig. 7. Unsupervised monosensor and two-sensor MPM segmentation.

a same image. Second, their detection with ICE-KOLM can improve the final unsupervised MPM segmentation results.

Comparison of computer times is not very reliable because our computer programs are not optimized and, what is more, the time used by every method is subject to different numbers of iterations subjectively chosen. Working with Ultra Spark, Enterprise 2 and in the case of "San Francisco," which is of 256×256 size, the respective times for ICE-GAUS-MPM, ICE-FIELD-MPM, and ICE-KOLM-MPM are about one minute, 15 minutes, and 50 minutes. This shows that ICE-GAUS-MPM is very interesting with respect to ICE-FIELD-MPM and that ICE-KOLM-MPM is rather time expensive. However, in contrast to ICE-GAUS-MPM and ICE-FIELD-MPM, the time of ICE-KOLM-MPM could undoubtedly be reduced. In particular, we use the whole subsets S_i^q, which maybe is a little superfluous and the use, at each iteration, of subsets of S_i^q could speed up the whole procedure significantly. In the case of the "Amazonia" image, Fig. 6, the result

TABLE 3
FAMILIES ESTIMATED BY ICE-KOLM AND ICE-PEAR

	Letter B					
	Noise 3			Noise 4		
	Class 1	Class 2	τ_{MPM}	Class 1	Class 2	τ_{MPM}
Real Distributions	Gamma	Gamma	8.25 %	Beta	Normal	2.90 %
ICE-GAUS	Normal	Normal	15.84 %	Normal	Normal	26.30 %
ICE-PEAR	Beta	Normal	11.70 %	Beta	Normal	2.90 %
ICE-KOLM	Beta	Gamma	5.35 %	Beta	Normal	2.53 %
ICE-FIELD	Normal	Normal	10.66 %	Normal	Normal	15.24 %

Error rates given by the real distributions-based MPM, ICE-GAUS-MPM, ICE-PEAR-MPM, and ICE-KOLM-MPM in the case of the "letter B."

TABLE 4
FAMILIES ESTIMATED BY ICE-KOLM AND ICE-PEAR

	Ring					
	Noise 3			Noise 4		
	Class 1	Class 2	τ_{MPM}	Class 1	Class 2	τ_{MPM}
Real Distributions	Gamma	Gamma	13.08 %	Beta	Normal	4.80 %
ICE-GAUS	Normal	Normal	30.75 %	Normal	Normal	6.10 %
ICE-PEAR	Beta	Normal	19.91 %	Beta	Normal	5.04%
ICE-KOLM	Gamma	Gamma	17.90 %	Beta	Normal	5.00 %
ICE-FIELD	Normal	Normal	10.90 %	Normal	Normal	3.40 %

Error rates given by the real distributions-based MPM, ICE-GAUS-MPM, ICE-PEAR-MPM, and ICE-KOLM-MPM in the case of the "Ring."

TABLE 5
DENSITY RECOGNITION WITH ICE-KOLM

	Letter B						Sensors
	Sensor 1			Sensor 2			1, 2
	Class 1	Class 2	τ^1_{MPM}	Class 1	Class 2	τ^2_{MPM}	τ^{12}_{MPM}
Real Distributions	Gamma	Gamma	8.25 %	Normal	Gamma	1.20 %	0.89 %
ICE-GAUS	Normal	Normal	15.90 %	Normal	Normal	2.30 %	3.70 %
ICE-KOLM	Gamma	Gamma	11.70 %	Normal	Gamma	2.10 %	1.70 %
	Ring						
Real Distributions	Gamma	Gamma	13.08 %	Normal	Gamma	3.77 %	3.97 %
ICE-GAUS	Normal	Normal	30.75 %	Normal	Normal	4.73 %	4.28 %
ICE-KOLM	Gamma	Gamma	17.90 %	Normal	Gamma	3.57 %	3.90 %

Real distribution-based segmentations, ICE-GAUS-based unsupervised segmentations, and ICE-KOLM-based unsupervised segmentations.

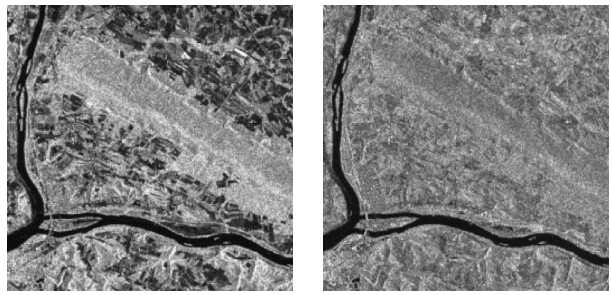
obtained with ICE-FIELD-MPM seems visually better than that obtained with ICE-GAUS-MPM, which suggests that the Markov field model is more appropriate than the Markov chain model. Otherwise, ICE-KOLM detects three different forms of noise distributions, which actually makes ICE-KOLM-MPM more effective than ICE-GAUS-MPM and shows again the advantage of generalized mixture estimation. The comparison between ICE-FIELD-MPM and ICE-KOLM-MPM is difficult; the latter seems to restore fine details better, but it is difficult to see in the real image whether such details exist.

6.4 Synthetic Multisensor Images Segmentation

Multisensor ICE-GAUS, ICE-KOLM based segmentations of noisy Letter B and the Ring are presented in Table 5. These results allow conclusions analogous to those in Section 5. ICE-KOLM recognizes the correct families, and the efficiency of the corresponding unsupervised segmentation is close to that of the method based on the true parameters. Visual results of segmentations are presented in Fig. 7.

6.5 Real Multisensor Image Segmentation

We consider in this section unsupervised segmentation of a real multisensor radar image of the area surrounding the town of Sunbury, Pennsylvania. The colors in the image are assigned to different frequencies and polarizations of the SIR-C radar. We only consider two sensors which seem complementary by their different visual aspects, as presented in Fig. 8. We present the results of two-sensor ICE-GAUS and ICE-KOLM based MPM segmentations in 2, 3, and 4 classes in Fig. 9. Rational comparison of the results is difficult in the absence of clear knowledge of the ground truth. However, note that the nature of the components varies with the class, and the ICE-KOLM based MPM segmentation seems richer.

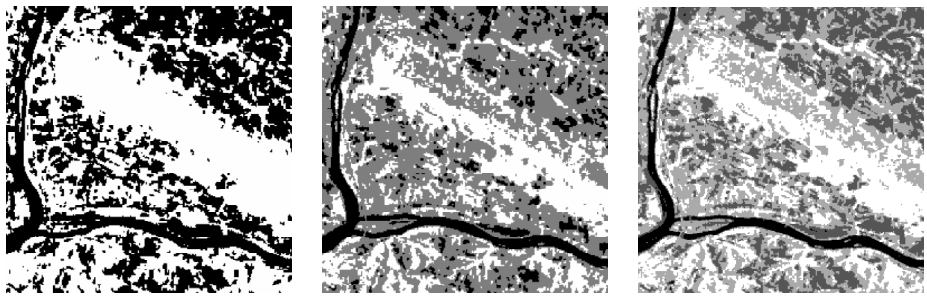


Sensor 1 Sensor 2

Fig. 8. Sensor 1: Blue color horizontally transmitted, vertically received. Sensor 2: Red color horizontally transmitted, vertically received.

Two classes Three classes Four classes

Two-sensor ICE-GAUS-based MPM segmentations.



Two classes Three classes Four classes

Two-sensor ICE-KOLM-based MPM segmentations.

Fig. 9. Two-sensor ICE-GAUS- and ICE-KOLM-based MPM segmentations. Components detected by ICE-KOLM are following. Two classes: Sensor 1 (Gamma, Normal), Sensor 2 (Beta, Beta). Three classes: Sensor 1 (Gamma, Gamma, Normal), Sensor 2 (Gamma, Normal, Beta). Four classes: Sensor 1 (Gamma, Normal, Normal, Normal), Sensor 2 (Gamma, Beta, Beta, Normal).

7 Discussion

We addressed in this work the problem of estimating generalized multisensor mixtures with applications to generalized multisensor hidden Markov chains and, more specifically, to unsupervised statistical image segmentation problems. The contribution of this paper rests, with respect to earlier works [1], [6], [26], on three points:

1) A more general family, based on Iterative Conditional Estimation (ICE), of generalized mixture estimation methods is proposed. In particular, this family is valid in the i.i.d. case, hidden Markov fields, hidden Markov chains, and any family $\Psi = \{F_1, ..., F_M\}$, provided each F_i is parametrized and can be estimated separately.
2) The effectiveness of a particular method based on ICE and Kolmogorov distance in the frame of generalized multisensor hidden Markov chains is shown using simulations.
3) Applications to unsupervised image segmentation generalize results obtained in the Gaussian monosensor case [1].

Concerning the results obtained using generalized hidden Markov chains, we may state, as a general conclusion, that component recognition and parameter estimation are quite efficient. However in the multisensor case further investigations would be desirable in order to choose judiciously which sensors among those available should be used.

Concerning the unsupervised image segmentation, the results presented allow us to put forth two conclusions. First, the form of the noise can indeed change with the class in real images, which renders the proposed techniques more effective than classical methods which assume the same form of noise for all classes. Second, little can be said about the relative value of the proposed method and classical hidden Markov field based approaches in the general case, but when inhomogeneous zones, like urban areas, are

present in the image the Markov chain based methods can be more appropriate than the Markov field ones.

We close with directions for future work. As concerns the proposed general method, it would be desirable to introduce nonparametric components and study how they are detected and estimated with ICE-GEMI. Another direction is the unsupervised generalized segmentation of sequences of multisensor images, or even 3D multisensor images. The flexibility of the proposed model renders it well suited to such tasks: it suffices to define a Hilbert-Peano scan in a three-dimensional set of pixels. Preliminary results of spatio-temporal segmentation of mono-sensor Gaussian images, presented in [1], are encouraging, although the important problem of estimating the number of classes remains. Devising a reliable method for estimating the number of classes remains of definite interest.

ACKNOWLEDGMENTS

The authors thank Associate Editor D.M. Titterington, and the anonymous reviewers for their help in improving the readability of the paper. This work was supported by the Convention CNET-INT 93 PE 74 05. The authors also thank Bruno Choquet and Danielle Pele, CCETT Rennes, for their collaboration. The image "Amazonia" was provided by CIRAD SA under the project GDR ISIS.

REFERENCES

[1] B. Benmiloud and W. Pieczynski, "Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images," *Traitement du Signal*, vol. 12, no. 5, pp. 433-454, 1995.

[2] J. Besag, "On the Statistical Analysis of Dirty Pictures," *J. Royal Statistical Soc.*, Series B, vol. 48, pp. 259-302, 1986.

[3] B. Braathen, W. Pieczynski, and P. Masson, "Global and Local Methods of Unsupervised Bayesian Segmentation of Images," *Machine Graphics and Vision*, vol. 2, no. 1, pp. 39-52, 1993.

[4] B. Chalmond, "An Iterative Gibbsian Technique for Reconstruction of m-ary Images," *Pattern Recognition*, vol. 22, no. 6, pp. 747-761, 1989.

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

[6] Y. Delignon, A. Marzouki, and W. Pieczynski, "Estimation of Generalized Mixtures and Its Applications in Miage Segmentation," submitted to *IEEE Trans. Image Proc.*

[7] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood From Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc.*, Series B, vol. 39, pp. 1-38, 1977.

[8] P.A. Devijver, "Hidden Markov Mesh Random Field Models in Image Analysis," *Advances in Applied Statistics, Statistics and Images*, vol. 1, pp. 187-227. Carfax Publishing Co., 1993.

[9] R.C. Dubes and A.K. Jain, "Random Field Models in Image Analysis," *J. Applied Statistics*, vol. 16, no. 2, pp. 131-164, 1989.

[10] G.D. Fornay, "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268-277, 1973.

[11] S. Geman and G. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.

[12] N. Giordana, "Segmentation non supervisée d'images multispectrales par chaînes de Markov cachées," PhD Thesis, Université de Technologie de Compiègne, 1996.

[13] R. Haralick and J. Hyonam, "A Context Classifier," *IEEE Trans. GRS*, vol. 24, pp. 997-1,007, 1986.

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

[15] G.K. Kaleh and R. Vallet, "Joint Parameter Estimation and Symbol Detection for Linear or Nonlinear Uknown Channels," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 2,406-2,413, 1994.

[16] S. Lakshmanan and H. Derin, "Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 799-813, 1989.

[17] A.M. Law and W.D. Kelton, *Simulation Modeling and Analysis*, 2nd ed. McGraw Hill, 1991.

[18] A. Maffet and C. Wackerman, "The Modified Beta Density Function as a Model for Synthetic Aperture Radar Clutter Statistics," *IEEE Trans. GRS*, vol. 29, no. 2, 1991.

[19] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic Solution of Ill-Posed Problems in Computational Vision," *J. Am. Statistical Assoc.*, vol. 82, pp. 76-89, 1987.

[20] P. Masson and W. Pieczynski, "SEM Algorithm and Unsupervised Segmentation of Satellite Images," *IEEE Trans.* GRS, vol. 31, no. 3, pp. 618-633, 1993.

[21] E. Mohn, N. Hjort, and G. Storvik, "A Simulation Study of Some Contextual Classification Methods for Remotely Sensed Data," *IEEE Trans. GRS*, vol. 25, pp. 796-804, 1987.

[22] A. Peng and W. Pieczynski, "Adaptive Mixture Estimation and Unsupervised Local Bayesian Image Segmentation, Graphical Models, and Image Processing," vol. 57, no. 5, pp. 389-399, 1995.

[23] W. Pieczynski, "Mixture of Distributions, Markov Random Fields, and Unsupervised Bayesian Segmentation of Images," Technical Report 122, L.S.T.A., Université Paris 6, 1990.

[24] W. Pieczynski, "Statistical Image Segmentation," *Machine Graphics and Vision*, vol. 1, nos. 1 and 2, pp. 261-268, 1992.

[25] 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. Applied Statistics*, vol. 6, no. 2, pp. 267-282, 1989.

[26] H.C. Quelle, Y. Delignon, and A. Marzouki, "Unsupervised Bayesian Segmentation of SAR Images Using the Pearson System," *Proc. IGARSS '93*, pp. 1,538-1,540, 1993.

[27] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[28] R.A. Redner, and H.F. Walker, "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Rev.*, vol. 26, pp. 195-239, 1984.

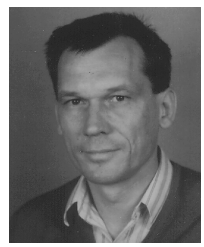
[29] W. Skarbek, *Generalized Hilbert Scan in Image Printing, Theoretical Foundation of Computer Vision*, R. Klette and W.G. Kropetsh, eds., pp. 45-57. *Akademik Verlag*, 1992.

[30] J. Tilton, S. Vardeman, and P. Swain, "Estimation of Context for Statistical Classification of Multispectral Image Data," *IEEE Trans. GRS*, vol. 20, pp. 445-452, 1982.

[31] L. Younes, "Parametric Inference for Imperfectly Observed Gibbsian Fields," *Probability Theory and Related Fields*, vol. 82, pp. 625-645, 1989.

[32] J. Zhang, "The Mean Field Theory in EM Procedures for Blind Markov Random Field Image Restoration, *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 27-40, 1993.

Nathalie Giordana received the Doctorat en Contrôle des Systèmes from the Université Technologique de Compiègne in 1996. Her thesis research was carried out at the Département Signal et Image at the Institut National des Télécommunications d'Evry. Her research interests include statistical image processing, pattern recognition, and theory of evidence.



Wojciech Pieczynski received the Doctorat d'Etat en mathématics from the Université de Paris 6 in 1986. He has been with France Telecom since 1987 and is presently a professor at the Institut National des télécommunications d'Evry. His main research interests include stochastic modeling, statistical techniques of image processing, and theory of evidence.