Fuzzy Random Fields and Unsupervised Image Segmentation

Hélène Caillol, Alain Hillion, *Associate Member, IEEE,* and Wojciech Pieczynski

***Abstract*—This paper deals with the statistical unsupervised image segmentation using fuzzy random fields. We introduce a new fuzzy model containing two components: a "hard" component, which describes "pure" pixels and a "fuzzy" component, which describes "mixed" pixels. First, we introduce a procedure to simulate this fuzzy field based on a Gibbs sampler step followed by a second step involving white or correlated Gaussian noises. Then we study the different steps of unsupervised image segmentation. Four different blind segmentation methods are performed: the conditional expectation, two variants of the maximum likelihood, and the least squares approach. As our methods are unsupervised, the parameters required are estimated by the stochastic estimation maximization (SEM) algorithm, which is a stochastic variant of the expectation maximization (EM) algorithm, adapted to our model. These "fuzzy segmentation" methods are compared with a classical "hard segmentation" one, without taking the fuzzy class into account. Our study shows that our "fuzzy" SEM algorithm provides reliables estimators, especially regarding the good robustness properties of the segmentation methods. Furthermore, we point out that this "fuzzy segmentation" always improves upon the "hard segmentation" results.**

***Index Terms*— Fuzzy random fields, unsupervised segmentation, fuzzy segmentation, SEM algorithm, Bayesian segmentation.**

I. Introduction

REMOTE sensing image classification may be put into different terms. The usual statistical models assume that the segmented image is a random field $\xi = \{\xi_1, \cdots, \xi_n\}$ taking its values in a finite set of classes Ω numbered from 1 to q. Each class represents the true nature of the ground, for instance the class 1 is "vegetation," the class 2 "water", and so on. The data set $x = \{x_1, \cdots, x_n\}$ is interpreted as a realization of some real random field $X = \{X_1, \cdots, X_n\}$ which is a noisy version of the original image $\xi = \{\xi_1, \cdots, \xi_n\}$. In statistical terms, segmentation is finding the initial realization of ξ from the consideration of the data set $x = \{x_1, \cdots, x_n\}$. More precisely, let us consider the case of the blind (classification without any context) Bayesian segmentation. Noting by S the set of the pixels, $S = \{1, \cdots, n\}$, each ξ_s will be estimated by the element of Ω which minimizes the so-called loss function:

$$L = \sum_{i=1}^{n} d(i,\, k) P_{X_s = s_s}(i)$$

Manuscript received September 8, 1992.; revised March 1, 1993.

H. Caillol and W. Pieczynski are with the Groupe Image Département Systèmes et Réseaux, Institut National des Télécommunications, 91011 Evry Cedex, France.

A. Hillion is with the Groupe Traitement d'Images Département Mathématiques et Systèmes de Communications, Ecole Nationale Superieure des Télécommunications, 29285 Brest Cedex, France.

IEEE Number 9209097.

where d is a class dissimilarity function and $P_{X_s = x_s}$ the distribution of ξ_s given $X_s = x_s$.

Another way to consider this problem is the fuzzy approach. From this point of view, each pixel s is attached with a q-dimensional vector $u(s) = [u_{is}]$. Roughly speaking, u_{is} which is the grade of membership of the sth pixel to the ith class, is the proportion of area s belonging to class i. Two constraints need to be introduced:

$$\forall s \in S \sum_{i=1}^{q} u_{is} = 1 \quad \text{and} \quad \forall i \, 1 \le i \le q \sum_{s=1}^{n} u_{is} > 0.$$

Like statistical segmentation, fuzzy segmentation can be expressed as the minimization of a given objective function. For instance the objective function used to perform the fuzzy c-means algorithm is [1]

$$Q = \sum_{i=1}^{q} \sum_{s=1}^{n} u_{is}^{m}\, d(x_s,\, m_i) \qquad m > 1$$

where m_i is the center value of the class i, d a dissimilarity function and m a weighting exponent. In [14] Pedrycz wrote a survey on the use of fuzzy representation in pattern recognition with an sizeable bibliography.

Classical statistical modeling forces each pixel to be associated with exactly one class. This assumption may be not realistic, particulary in the case of satellite data, whereas the fuzzy approach allows the possibility of mixed pixels. On the other hand, statistical methods present the advantage of being well adapted to strong and particularly highly correlated noises. One can find in [9] an application of the fuzzy c-means algorithm which indicates confusion of classes in the case of correlated noise. Some approaches mixing fuzzy and statistical modeling have also been developed. For instance, Wang proposed in [19] a generalization of the statistical parameters to a fuzzy representation which allows a fuzzy version of classical statistical classification rules such as maximum likelihood. The consistency of the stochastic model is ensured by the knowledge of the probability distribution of $U_s = \{U_{s1}, \cdots, U_{sq}\}$ which is often assumed to be of continuous type and even to obey suitable transformations of Gaussian distribution; see the model given by Kent and Mardia in [10]. This assumption has some drawback, because it implies that every pixel is a mixture of thematic classes, although some pixels are definetely "pure" water or "pure" vegetation. Accordingly we propose a statistical model which includes fuzzy membership and "hard" thematic classes. For the sake of simplicity, we confine the study to the case $q = 2$

and put $\xi_s = \xi_{s2} = 1 - \xi_{s1}$. We relax the continuity assumption and admit that the distribution of ξ_s, which appears as the proportion of area s belonging to class "vegetation," may include Dirac measures at 0 and 1.

The paper is organized as follows. In Section II, we describe a way of simulating realizations of field $\xi = (\xi_s)_{s \in S}$, by two-step adaptation of the classical Gibbs sampler ([6]) which takes account of the spatial interactions between pixels. In Section III, where the distribution of X conditioned on ξ is assumed to be Gaussian, we define four estimations of ξ_s from the observation of the intensity X_s of pixel s. Namely, a modified version of the maximum posterior probability estimation, called relative maximum likelihood, the maximum likelihood, the conditional expectation, and the minimum mean square estimator are defined.

Insofar as our segmentation methods are completely unsupervised, Section IV is devoted to estimating the unknown parameters necessary for the implementation of the previous methods. Naturally, this estimation is based only on the data $x = \{x_1, \cdots, x_n\}$. Taking advantage of some of our previous works in the context of unsupervised segmentation, [11], [13], we introduce a fuzzy version of the so-called stochastic estimation maximization (SEM) mixture estimation algorithm [4], [13]. All the applications, results and comments, for several noise types, are included in Section V: they include estimates by SEM algorithm, errors, and robustness curves and segmentation of an image. The last section contains the conclusions.

II. The Fuzzy Model and Its Simulation

Recall that each ξ_s is the proportion of the class 2 on the pixel s and takes the value 0 if s is "pure" class 1, the value 1 if s is "pure" class 2 and any value in $]0,1[$ if the both classes are simultaneously present on s. We have to define a probability distribution which takes these situations into account and find a method to simulate realizations of ξ. The Gibbs sampler is currently performed in image processing to sample Markovian random fields. It is natural to envisage a "fuzzy" version of this algorithm. Let us expose briefly the Gibbs sampler:

We consider a Markovian and stationary random field $Z = \{Z_1, \cdots, Z_n\}$ taking its values in Ω (the set of classes) and suppose that Z is Markovian with respect to a spatial type of neighborhood, for instance the four nearest neighbors. The neighborhood of a pixel s will be noted V_s and its spatial shape is assumed to be independent of s. The assumption of any null probability configurations allows the distribution of Z to be expressed as

$$P_Z[z] = Ke^{-U(z)} \qquad (^*)$$

where U is called the energy function and K is a normalization constant.

A clique C is a subset of S (set of sites) defined by one of these propositions:

1) C is a single pixel,
2) two different elements of C are neighbors.

For instance, if we consider the four nearest neighbors, there are three types of cliques: a single pixel, two vertical neighbors and two horizontal neighbors.

According to the Hammerley–Clifford theorem, the energy function U can be expressed by

$$U(z) = \sum_i \sum_{C \in T_i}^{n} \phi_i(z_C)$$

where i represents the type of the clique, T_i the set of the cliques of type i in S, and ϕ_i the energy potential function defined on this type of clique.

We need only define the potential functions on each type of clique to be able to compute the energy of each configuration z of Z. Since the constant K in (*) is usually uncomputable due to the high cardinal of the set of configurations $(|\Omega|^n)$, knowing U is not sufficient to evaluate the distribution of Z. However, conditional distributions of Z_s given Z_{V_s} are generally easy to compute, which allows an iterative procedure, the so-called Gibbs sampler, based on these conditional distributions. Starting from an initial configuration z^0, at each step k, we compute the conditional distribution of each Z_s^{k-1} based on the realization z_V^{k-1} and draw a new realization of Z_s^k, z_s^k, according to this distribution. We obtain stochastic series of realizations $z^0, z^1, \cdots, z^q$ which converge to the realization of Z defined by (*).

Now we adapt this classical sampler to our fuzzy model by considering at first the "fuzzy" class to be a real class and simulating its realizations as a second step. Letting

$$\pi_0 = P[\xi_s = 0] \qquad \pi_1 = P[\xi_s = 1]$$

$$\text{and} \quad \pi_F = 1 - (\pi_0 + \pi_1)$$

and denoting by F the open interval $]0,1[$ we define an auxiliary random field $Y = (Y_s)_{s \in S}$ by

$$Y_s = 0 \Leftrightarrow \xi_s = 0$$
$$Y_s = 1 \Leftrightarrow \xi_s = 1$$
$$Y_s = F \Leftrightarrow \xi_s \in]0,1[\ .$$

Supposing that Y is a Markov random field, we can perform the Gibbs sampler to sample its realizations. This is the first step of our procedure. Since we have a three-class Markov random field, it will be easy, for each pixel belonging to the class F, to draw a value in $]0,1[$ according to a continuous distribution. Finally compiling step 1) and step 2) we obtain a fuzzy image containing "pure" pixels of class 1 and 2 and fuzzy pixels taking their values in $]0,1[$.

Step 1) The first step of our procedure is simulating Y. We consider the neighborhood of the four nearest neighbors and define the potential function only on the cliques pairs, which entails that the two types of clique pairs have the same potential function and that the potential function of the single clique is a constant. As the class F has not the same behavior as the two "hard" classes, we need to define three different

(a) (b)

Fig. 1. Examples of the conditional density of ξ_s given ξ_{V_0}.

expressions for the potential function:

$$\Phi(s, t) = \begin{cases} -b \text{ if } (Y_s, Y_t) = (0, 0) \text{ or } (1, 1) \\ b \text{ if } (Y_0, Y_t) = (0, 1) \text{ or } (1, 0) \\ d \text{ if } (Y_s, Y_t) = (0, F) \text{ or } (1, F) \\ \quad \text{or } (F, 0) \text{ or } (F, 1) \\ e \text{ if } (Y_s, Y_t) = (F, F) \end{cases}$$

where

b defines the size of the regions of the two pure classes.
d manages the attraction between the fuzzy class and the other classes.
e controls the area of the fuzzy class.

Different numerical choices for these three parameters generate a large diversity of simulated images. The image is the more homogeneous that factor b is more significant. To modify the localization of the fuzzy pixels towards the "hard" pixel, factor d is modified, and if you want to diminish the proportion of fuzzy pixel, e is reduced.

Step 2) The previous step produced a realization of a three-class (0 and 1 and F) random field Y. The second step concerns pixels of Y which are in class F, the objective being to attribute at each "fuzzy" pixel a value in $]0, 1[$. In order to take spatial interactions into account, we have to consider the conditional distribution of ξ_s, given the values ξ_t, for t belonging to V_s, the four nearest neighbors of s.

We shall suppose that the density of this distribution, g, depends upon two parameters a and v, i.e.,

$$g(x) = g(x, a, v)$$

where a is a shape parameter and v is the local sum of ξ_t over the likelihood

$$v = \sum_{t \in V_s} \xi_t.$$

The family of densities is chosen so that they have a strong mode in 0 or 1, according to the proximity of v to 0 or 4.

In our simulation study g is of the affine shape: $g(x) = a(v - 2)x + \beta$ where β is a normalizing constant. Let us specify that a is chosen for all pixels, v is computed for every pixel and then β is taken in such a way that g is a probability density. Thus, roughly speaking, g is of the type (a) in Fig. 1, for $v = 0$ and of type (b) for $v = 4$, respectively.

As in the Gibbs sampler, we proceed by visiting the fuzzy pixels. During the first visit, pixels belonging to V_s can be of three different kinds: hard (0 or 1), of class F or fuzzy, i.e., in $]0, 1[$ (if the pixel considered has already been sampled according to the distribution given by g). During the first visit, we compute v by allocating to pixels F the value 0.5. From the second visit, pixels F disappear and, for each s of class F in step 1, all pixels in V_s are numbers in $[0, 1]$, and thus v can be evaluated directly. The sequence of fuzzy images obtained becomes visually stabilized after about 200 visits. The variation of parameters b, d, e, a allows many possibilities for the fuzzy images one of which is noised and segmented below.

Now we have at our disposal a fuzzy image. Since we perform an unsupervised segmentation, we shall suppose not having any information, like training data, about this "true" image. We shall perform our segmentation knowing the only realization of a noisy version of ξ, say a realization of the field X.

The real images are obtained from the fuzzy realization above by adding different Gaussian noises. As the conditional distribution of X_s given ξ_s is assumed to be Gaussian and the parameters of the noise on a fuzzy pixel are assumed to depend linearly on the parameters of the both "hard" classes, we just need the mean and variance m_0 and σ_0^2 of the class 1 and the mean and variance m_1 and σ_1^2 of the class 2 to define this distribution:

$$P_{X_s}^{\epsilon} = N[m(\epsilon), \sigma^2(\epsilon)]$$

where

$$m(\epsilon) = (1 - \epsilon)m_0 + \epsilon m_1 \text{ and } \sigma^2(\epsilon) = (1 - \epsilon)\sigma_0^2 + \epsilon\sigma_1^2.$$

In the following, we will denote by $f(\epsilon, x)$ the corresponding Gaussian density, which is

$$f(\epsilon, x) = \frac{1}{\sqrt{2\pi\sigma^2(\epsilon)}} e^{-\frac{(x - m(\epsilon))^2}{2\sigma^2(\epsilon)}}.$$

The noise parameters are fixed in advance and determine the type and the power of the noise. Indeed there are three possible combinations and the four parameters m_0, σ_0^2, m_1 and σ_1^2 determine three types of noise:

- type 1 means discriminating noise $m_0 \neq m_1$ and $\sigma_0^2 = \sigma_1^2$.
- type 2 variances discriminating noise $m_0 = m_1$ and $\sigma_0^2 \leq \sigma_1^2$.
- type 3 means and variances discriminating noise $m_0 \neq m_1$ and $\sigma_0^2 \neq \sigma_1^2$.

Denoting by $\Delta = m_1 - m_0$ and $\rho = \frac{\sigma_0^2}{\sigma_1^2}$, the power of a noise depends on these two parameters.

The correlation is also a criterion of noise comparison. We consider white and correlated noises. White Gaussian noises are easy to obtain by simulating independent realizations of a Gaussian random variable. Correlated noises are simulated by taking weighted mobile averages computed in each pixel from the values in the neighborhood of a white noise. We propose two different types of neighborhood: 8 or 24 neighbors. Note that there is no relation between these neighborhoods and the neighborhoods of the Markovian assumption. In the following, we will add an "A" to the name of a noise if it uses 8 neighbors and a "B" if it uses 24 neighbors. For instance, noise 1A will denote a noise of type 1 (means discriminating) correlated using 8 neighbors.

III. Fuzzy Unsupervised Blind Statistical Segmentation

As stated in the introduction, we have to estimate the realization of $\xi = \{\xi_1, \cdots, \xi_n\}$ from the observation $x = \{x_1, \cdots, x_n\}$. We adopt the unsupervised blind segmentation. The blind segmentation consists of estimating the unobservable realization of each ξ_s from the observation $X_s = x_s$. According to our model, we can compute the conditional distribution of ξ_s given $X_s = x_s$, the so-called posterior distribution, which can be expressed from the distribution P_{ξ_s} of ξ_s, the so-called prior distribution, and the conditional distribution $P^{\epsilon}_{X_s}$ of X_s given $\xi_s = \epsilon$. Unsupervised segmentation means that the parameters required to perform the segmentation methods are unknown but instead estimated from the observation. We shall apply in the next section a fuzzy adaptation of the SEM algorithm previously mentioned in the introduction.

The aim of this section is to describe how the four fuzzy segmentation methods considered behave. Let us recall that a Bayesian estimator is obtained by minimization, with respect to the class k, of a given loss function:

$$L = \sum_{i=1}^{q} d(i, k) P_{X_s = x_s}(i).$$

Each choice of d defines an specific estimator. The classical Bayesian "hard" blind segmentation applies the dissimilarity function defined by

$$d(i, k) = 0 \text{ if } i = k$$
$$d(i, k) = 1 \text{ if } i \neq k.$$

In this case, the minimization of the loss function L consists of choosing, for each pixel $s \in S$, the class which maximizes the posterior distribution of the distribution of ξ_s. In the fuzzy model considered, this distribution contains a "hard" component which is a measure on $\{0, 1\}$ and a "fuzzy" one which is a measure on $]0, 1[$ given by a density with respect to the Lebesgue measure. This continuous component of ξ_s allows a large class of segmentation methods. Four different segmentation methods are proposed in this section.

To be more precise, let us suppose that the distribution of ξ_s is defined by h, a function on $[0, 1]$. We have

$$\pi_0 = h(0) = P[\xi_s = 0] \qquad \pi_1 = h(1) = P[\xi_s = 1]$$

and the restriction of h to $]0, 1[$ describes the fuzzy component of the distribution of ξ_s. Let us note that h can be seen as a density of the distribution of ξ_s with respect to the measure $v = \delta_0 + \delta_1 + \mu$, where δ_0, δ_1 are Dirac's measures and μ the Lebesgue measure. We will suppose that the restriction of h to $]0, 1[$ is constant:

$$h(\epsilon) = 1 - \pi_0 - \pi_1 = \pi_F \quad \forall \epsilon \in]0, 1[$$

Thus h and f defined in the previous section give the distribution of (ξ_s, X_s), whose density with respect to $v \otimes \mu$ is defined on $[0, 1] \times R$ by:

$$h(\epsilon) f(\epsilon, x).$$

So the distribution of X_s is given by the density defined on R by

$$w(x) = h(0) f(0, x) + h(1) f(1, x) + \int_0^1 h(\epsilon) f(\epsilon, x)\, d\epsilon.$$

Finally, the distribution of ξ_s conditional to $X_s = x_s$ is given by the following density (with respect to $v = \delta_0 + \delta_1 + \mu$) on $[0, 1]$:

$$h(\epsilon, x) = \frac{h(\epsilon) f(\epsilon, x)}{w(x)}.$$

This *a posteriori* distribution contains then a hard component given by

$$\pi_{0,x} = h(0, x)$$
$$\pi_{1,x} = h(1, x)$$

and a fuzzy one given by the density $h(\epsilon, x)$ on $]0, 1[$ (x is fixed).

Now the different distributions are defined, and it remains to introduce the different methods.

1) *Maximization of the Posterior Likelihood* $h(\epsilon, x)$. This method was presented at the beginning of this section in the case of the "hard" blind segmentation. In fact, for x fixed $h(\epsilon, x)$ is the likelihood of the posterior distribution of ξ_s with respect to the measure $v = \delta_0 + \delta_1 + \mu$. We shall maximize the continuous posterior likelihood $h(\epsilon, x)$, like in the hard case the posterior distribution. We examine two different ways of maximizing this likelihood:

a) The first method proceeds in two steps. As in the simulation of the fuzzy image, we consider first that the fuzzy class is a "real" class F, then pixel classified F, will be treated in a second step.

Step 1)

We apply a three-class segmentation by taking the maximum argument in $\{0, 1, F\}$ of $\pi_{0,x}$, $\pi_{1,x}$, $\pi_{F,x}$ where $\pi_{F,x} = 1 - (\pi_{0,x} + \pi_{1,x})$.

Step 2)

The pixels classified as fuzzy, say the pixels where F was the maximum argument in the previous step, are reclassified by taking the value ϵ in $]0, 1[$ which maximizes the posterior density defined by the restriction of $h(\epsilon, x)$ on $]0, 1[$. We call this method the "relative" maximum likelihood.

b) The second way maximizes directly the posterior likelihood: we take for the estimation of ξ_s the maximum argument in $[0, 1]$ of $h(\epsilon, x)$. This method is the real "maximum likelihood".

2) *The Conditional Expectation*. The conditional expectation is the best approximation in the mean square error sense: we choose for d the mean square deviation. The realization of ξ_s is estimated by $E(\xi_s / X_x = x_s)$. This expectation can be computed as follows (we put $\Omega = [0, 1]$):

$$\begin{aligned} E(\xi_s / X_s = x_s) &= \int_\Omega \epsilon\, dP_{X_s = x_s} \\ &= 0\pi_{0,x} + 1\pi_{1,x} + \int_0^1 \epsilon h(\epsilon, x)\, d\epsilon \\ &= \pi_{1,x} + \int_0^1 \epsilon h(\epsilon, x)\, d\epsilon. \end{aligned}$$

3) *The Least Squares Method.* The least squares method is the best linear approximation in the mean square sense. The difference with the conditional expectation lies in the linearity of the approximation, which makes its calculation much easier. We have to find α and β in such a way that

$$\langle \xi - (\alpha - \beta X), 1\rangle_{L_2} = 0 \quad \text{and} \quad \langle \xi - (\alpha - \beta X), \xi\rangle_{L_2} = 0.$$

Hence

$$\alpha = \frac{E(\xi X) - E(\xi)E(X)}{E(X^2) - E(X)^2} \quad \text{and} \quad \beta = E(\xi) - \alpha E(X)$$

as an example, let us calculate $E(\xi X)$. Recalling that $\beta_F = 1 - (\pi_0 + \pi_1)$ we have

$$\begin{aligned} E(\xi X) &= \int_{[0,1]\times R} \epsilon x h(\epsilon) f(\epsilon, x)\, dv \otimes \mu \\ &= 0\pi_0 \int_R \epsilon x f(0, x)\, d\epsilon dx + 1\pi_1 \int_R \epsilon x f(1, x)\, d\epsilon dx \\ &\quad + \pi_F \int_R \epsilon x f(\epsilon, x)\, d\epsilon dx \\ &= \pi_1 m_1 + \pi_F \int_0^1 m(\epsilon)\, d\epsilon \\ &= \pi_1 m_1 + \pi_F \int_0^1 [(1-\epsilon)m_0 + m_1]\, d\epsilon \\ &= \pi_1 m_1 + \frac{1}{2}\pi_F(m_0 + m_1). \end{aligned}$$

In the same way:

$$\begin{aligned} E(\xi) &= \pi_1 + \frac{1}{2}\pi_F \\ E(X) &= \pi_0 m_0 + \pi_1 m_1 + \frac{1}{2}\pi_F(m_0 + m_1) \\ E(X^2) &= \pi_0\sigma_0^2 + \pi_1\sigma_1^2 + \pi_F\left(\frac{1}{3}(\sigma_0 - \sigma_1)^2 + \sigma_0\sigma_1\right)\bigg). \end{aligned}$$

Observing $X_s = x_s$, we take $\xi_s^* = \alpha + \beta x_s$ as an estimation of ξ_s. In this case the estimation takes its values in R and we must "bring" them back in $[0, 1]$. We can truncate the values obtained for estimation or transform them linearly. The simulations show that the linear transformation provides the best results.

We conclude this section with some remarks about the evaluation of the efficiency of our methods. The most current way to define an error between the original fuzzy image, say ξ and the segmented fuzzy image, denoted by ξ^*, is to evaluate the absolute average error by

$$\tau_1 = \frac{\sum_{s\in S}|\xi_s - \xi_s^*|}{Card(S)}.$$

This error can be seen as a "proportion" of the pixels not well classified. We also consider the mean square error which defines another error rate τ_2

$$\tau_2 = \frac{\sqrt{\sum_{s\in S}(\xi_s - \xi_s^*)^2}}{Card(S)}.$$

IV. The Fuzzy SEM Algorithm

As we have seen in Section II, the distribution of (ξ_s, X_s) is defined by the distribution of ξ_s, given by a function h defined on $[0, 1]$, and the family of Gaussian distributions of X_s conditioned on $\xi_s = \epsilon$, given by the family of densities $f(\epsilon, x)$. With h assumed constant on $]0, 1[$, the distribution of ξ_s is given by $\pi_0 = h(0)$ and $\pi_1 = h(1)$. According to the form of the conditional distributions we adopted in the former section, the function f (defined on $[0, 1] \times R$) is completely determined by m_0, m_1, σ_0^2, σ_1^2.

Finally, the problem we deal with in this section is to estimate the parameter $\theta = (\pi_0, \pi_1, m_0, m_1, \sigma_0^2, \sigma_1^2)$ from a sample of X_s. Let $x_1, x_2, \cdots, x_n$ be the data. The proposed fuzzy SEM can be seen as an adaptation of the SEM algorithm, whose good behavior in the image segmentation context has been shown in previous work [11], [13], [17], [18]. It is an iterative method which runs as follows:

(i) Initialization.

For each $i = 1, \cdots, n$ sample a value in $\{0, 1\}$ according to the uniform distribution 0.5, 0.5. This gives a partition Q_0^0, Q_1^0 of the set $\{1, 2, \cdots, n\}$, with $i \in Q_0^0$ if 0 has been sampled and $i \in Q_1^0$ if 1 has been sampled. Consider that the sub-sample $(x_i)_{i\in Q_1^0}$, is issued from the class 0 and estimate m_0, σ_0^2 by empirical mean and variance (by the using only x_i such that i is in Q_0^0). Do the same for estimating m_1, σ_1^2 by the use of Q_1^0. This gives the first value

$$\theta^0 = (\pi_0^0, \pi_1^0, m_0^0, m_1^0, (\sigma_0^0)^2, (\sigma_1^0)^2)$$

of the parameter (with $\pi_0^0 = 0.5$, $\pi_1^0 = 0.5$).

(ii) Calculation of θ^{k+1} from θ^k and $x_1, x_2, \cdots, x_n$.

For each $i = 1, \cdots, n$ calculate $h^k(0, x_i)$, $h^k(1, x_i)$ which are probabilities of 0 and 1 conditioned on $X_i = x_i$ and based on θ^k (see the definition of h in Section III). Sample, for each $i = 1, \cdots, n$, a value in $\{0, 1, F\}$ according to the distribution:

$$h^k(0, x_i),\ h^k(1, x_i),\ 1 - h^k(0, x_i) - h^k(1, x_i).$$

This gives a partition Q_0^k, Q_1^k, Q_F^k of $\{1, 2, \cdots, n\}$. As in the initialization use Q_0^k, Q_1^k in order to estimate empirical means and variances. This gives the next values m_0^{k+1}, m_1^{k+1}, $(\sigma_0^{k+1})^2$, $(\sigma_1^{k+1})^2$ of these parameters. The next values of π_0^{k+1}, π_1^{k+1} are given by the frequencies of 0 and 1 in the obtained sample:

$$\pi_0^{k+1} = \frac{1}{n} card(Q_0^k) \quad \pi_1^{k+1} = \frac{1}{n} card(Q_1^k).$$

This gives the next value

$$\theta^{k+1} = (\pi_0^{k+1}, \pi_1^{k+1}, m_0^{k+1}, m_1^{k+1}, (\sigma_0^{k+1})^2, (\sigma_1^{k+1})^2)$$

of the parameter. Thus we obtain a stochastic sequence of parameters whose principal advantage, in the hard case, with respect to the deterministic one obtained by the classical EM is to avoid the local maxima of the likelihood. Moreover, these estimates converge in a large number of situations [3].

Remark: We have not used Q_F^k in the second step of the fuzzy SEM above because of the simplicity of the function

$h(\xi)$ considered, but more complex situations can be treated. If $h(\xi)$ is of any kind, for each i in Q_F^k we should have to sample a value in $[0, 1]$ according to the *a posteriori* density $h^k(t, x_i)$ (x_i is fixed and t varies in $]0, 1[$, see Section III) based on the current value θ^k of the parameter. This would give a "fuzzy" subsample allowing the reestimation of the parameters defining the fuzzy component, i.e., the restriction to $]0, 1[$ of $h(\xi)$, of the distribution of ξ_s.

V. Applications and Results

A. Simulated Image and Noises

We first apply the simulation algorithm presented in the second section. Recall that we have to fix the three parameters of the potential function, the number of iterations for the Gibbs sampler, the shape parameter and finally the number of iteration for sampling of the fuzzy class realizations. Finally, we need six parameters. We propose the following values:

$b = 1 \quad d = 0,8 \quad e = 0,95 \quad a = 1$
20 iterations of the Gibbs sampler
200 iterations for the simulation of the values on the fuzzy class

The fuzzy image obtained is presented at the end of this section, see "simulated image". It will be our original reference fuzzy image. Concerning the additive Gaussian noises, we saw in the second section that we shall consider three types of noises and two correlation degrees. Here we fix the values of the parameters for the three types of noise:

- Noise 1: $m_0 = 1 \quad m_1 = 2 \quad \sigma_0^2 = 1 \quad \sigma_1^2 = 1$, $\Delta = 1 \quad \rho = 1$
- Noise 2: $m_0 = 1 \quad m_1 = 1 \quad \sigma_0^2 = 1 \quad \sigma_1^2 = 3$, $\Delta = 0 \quad \rho = 1,73$
- Noise 3: $m_0 = 1 \quad m_1 = 1,5 \quad \sigma_0^2 = 1 \quad \sigma_1^2 = 0,185$, $\Delta = 0,5 \quad \rho = 0,43$

As we note above, we also consider three degrees of correlation—uncorrelated, type A, and type B— for each type of noise. Thus we obtain nine different noise types. Insofar as our segmentation methods are blind, it would be needless to apply them in the nine cases. We need only to apply these segmentations in the case of uncorrelated noise to test the effectiveness of the segmentations. On the other hand, as we shall see in the applications, the fuzzy SEM algorithm is sensitive to the correlation of the noise, which is why we present the estimation results also in two cases of correlated noise in addition to the uncorrelated case: noises 2A and 3B. These three cases provide a complete idea of the behavior of the estimation results. We present also the estimates provided by the binary SEM algorithm, that is, without taking account of the fuzzy class and considering the only two "hard" classes. The properties of the SEM algorithm in the "hard" context are studied in [13].

B. Estimation by the SEM Algorithm

We provide in Tables I–III, three results of the application of "fuzzy" and classical "binary" SEM to our fuzzy image corrupted by the noise. The fuzzy SEM is less sensitive to the "variances discriminating" nature of the noise and its correlation, the two factors which seem to degrade the effectiveness of the binary one. This could be due to the fact that the binary SEM works on fuzzy data which are not suited to its principle. In fact, the very good behavior of the binary SEM is shown in [13] and this fuzzy version does not degrade its properties. Furthermore, in the case of fuzzy data, the fuzzy SEM is necessary to estimate the *a priori* probabilities. Let us note that the binary SEM can encounter some difficulties in estimating the variances.

TABLE I
True Parameter Values and Estimates, Noise 1: Uncorrelated, Means Discriminating

	"Fuzzy SEM"	"Binary SEM"
$m_0=1$	0,99	0,84
$m_1=2$	1,8	1,97
$\sigma_0^2=1$	1,06	0,93
$\sigma_1^2=1$	1,08	1,05
$\pi_0=0,33$	0,3	
$\pi_1=0,33$	0,36	

TABLE II
True Parameter Values and Estimates, Noise 2A: Variances Discriminating, Correlated Using 8 Neighbors

	"Fuzzy SEM"	"Binary SEM"
$m_0=1$	1,02	0,93
$m_1=1$	0,98	0,99
$\sigma_0^2=1$	0,94	1,36
$\sigma_1^2=3$	3,13	3,26
$\pi_0=0,33$	0,36	
$\pi_1=0,33$	0,27	

TABLE III
True Parameter Values and Estimates, Noise 3B: Variances Discriminating, Correlated Using 24 Neighbors

	"Fuzzy SEM"	"Binary SEM"
$m_0=1$	0,89	0,92
$m_1=1,5$	1,49	1,47
$\sigma_0^2=1$	1,03	0,93
$\sigma_1^2=0,185$	0,183	0,95
$\pi_0=0,33$	0,4	
$\pi_1=0,33$	0,3	

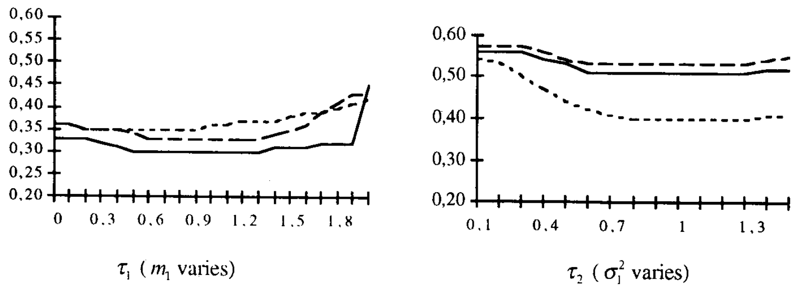
C. Segmentation Results

The effectiveness of an unsupervised segmentation method depends on three independent factors:

1) the efficiency of the parameter estimation method used,
2) the theoretical error of the segmentation method used,
3) the robustness of the segmentation method used with respect to the parameters needed.

	Rel. Max. Likeli.	Max. Likelihood	Cond. Expectation	Least Squares
τ_1	0,30	0,33	0,35	0,49
τ_2	0,51	0,53	0,40	0,51

(a)



(b)

	Rel. Max. Likeli. fuzzy	Rel. Max. Likeli. binary	Max. Likelihood	Cond. Expectation	Least Squares
τ_1	0,33	0,38	0,33	0,36	0,49
τ_2	0,50	0,58	0,49	0,41	0,52

(c)

Fig. 2. Noise 1: Uncorrelated, mean discriminating. (a) Estimated classification errors. (b) Robustness curves. (c) Image segmentation.

The theoretical error of a given method is the expectation of the loss function L defined in the introduction. It can be interpreted as the optimal segmentation error that we could expect by performing this method. Unfortunately, the theoretical errors are not directly computable in the fuzzy case, due to the expression of $f(\epsilon, x)$. It is, however, possible to obtain an approximation of these errors: we sample a Gaussian mixture and we classify this mixture by using the true parameters. These estimated errors are computed using the two types of errors τ_1 and τ_2 presented at the end of the second section.

If a given parameter is poorly estimated but the segmentation method is robust with respect to this parameter, the performance degradation of the corresponding unsupervised method can be negligible. Since we have at our disposal four different methods, we propose, at first, a study of their robustness with respect to the parameters. The estimated errors must also be taken into account. A method which presents a high theoretical error, even if very robust, could not be competitive with a less robust method which presents a lower theoretical error. In this present case, the theoretical errors show that the least squares are not competitive and thus we are led to reject it in this context.

For each different noise we present the theoretical efficiency of classification of the four methods, given by the estimated errors, two significant robustness curves, and the results of the segmentation of the noisy image, both with respect to τ_1 and τ_2. The robustness curves represent the variations of the estimated error when one of those parameters deviates. The representations used in all the graphs are as follows:

____ Relative maximum likelihood
— — Maximum likelihood
– – – Conditional expectation

The transformation applied in the least squares segmentation leads to a falsification of the theoretical results, which is why we omit the robustness curves in this case.

Noise 1 results are shown in Fig. 2; noise 2 in Fig. 3; and noise 3 in Fig. 4.

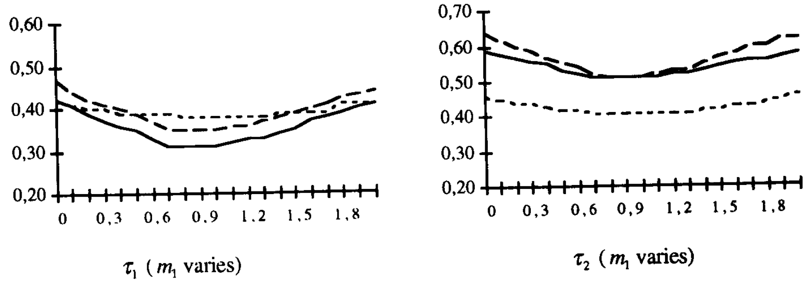
At first, we note that these methods are robust, particularly the conditional expectation method. Recalling the estimation values provide by the fuzzy SEM, we can observe that these values correspond to segmentation errors which are close to the simulated error. This means that this estimation algorithm is well adapted. A more precise analysis of these results can allow us to favor one of these methods, depending on the nature and the correlation of the noise. For instance, in the case of noise 1 (means discriminating noise), the relative maximum likelihood provides the best estimated error τ_1, and the best robustness. Whereas, the maximum likelihood is, on the other hand, less sensitive to the variations of the deviation, especially near zero. We may prefer the relative maximum likelihood in favorable cases for the estimation, low correlation of the noise, or, on the contrary, the maximum likelihood if the noise is highly correlated. The conditional expectation seems to be less sensitive to a "variances discriminating" noise and can in some situations be more reliable than the two maximum likelihoods, even if the error τ_1 is concerned.

D. Some Images

We discuss in this last section some images that illustrate this study. They are shown in Figs. 5 and 6. We have

	Rel. Max. Likeli.	Max. Likelihood	Cond. Expectation	Least Squares
τ_1	0,31	0,35	0,38	0,51
τ_2	0,51	0,52	0,41	0,54

(a)



(b)

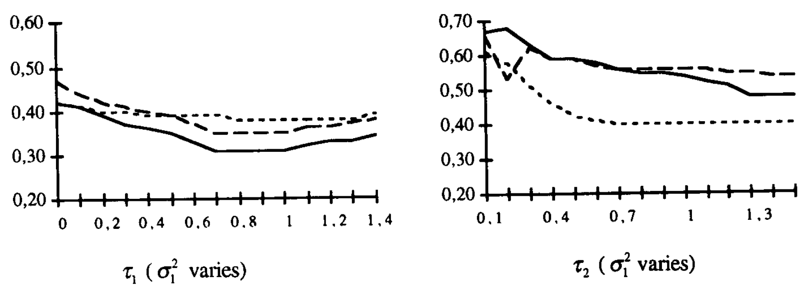
	Rel. Max. Likeli. fuzzy	Rel. Max. Likeli. binary	Max. Likelihood	Cond. Expectation	Least Squares
τ_1	0,36	0,51	0,36	0,39	0,43
τ_2	0,54	0,69	0,53	0,43	0,45

(c)

Fig. 3. Noise 2: Uncorrelated, variances discriminating. (a) Estimated classification errors. (b) Robustness curves. (c) Image segmentation.

	Rel. Max. Likeli.	Max. Likelihood	Cond. Expectation	Least Squares
τ_1	0,33	0,36	0,35	0,51
τ_2	0,53	0,56	0,40	0,53

(a)



(b)

	Rel. Max. Likeli. fuzzy	Rel. Max. Likeli. binary	Max. Likelihood	Cond. Expectation	Least Squares
τ_1	0,33	0,38	0,34	0,36	0,45
τ_2	0,51	0,53	0,50	0,41	0,48

(c)

Fig. 4. Noise 3: Uncorrelated, means and variances discriminating. (a) Estimated classification errors. (b) Robustness curves. (c) Image segmentation.

deliberately chosen strong noises in order to point out the strengths as well as the weaknesses of the estimators and the different segmentation methods. In return, the images corresponding to these studied situations are not visually satisfactory. Accordingly, we present here only the images corresponding to noise 3. In order to get visual feel of the effectiveness of our fuzzy unsupervised segmentation, we also present images which correspond to a lower noise, called noise 4, defined by the following parameters: $m_0 = 1$, $m_1 = 3$, $\sigma_0^2 = 1$, $\sigma_1^2 = 1$.

Fig. 5. Simulated image.

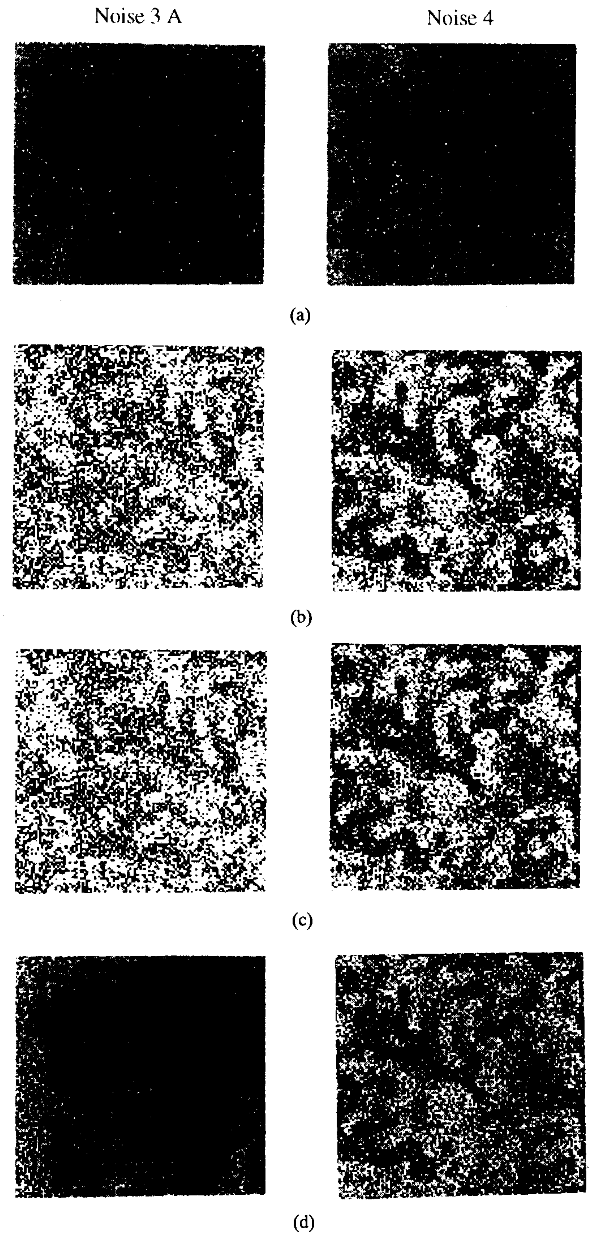


Fig. 6. (a) Corrupted images. (b) Relative maximum likelihood. (c) Maximum likelihood. (d) Conditional expectation.

VI. Conclusions

We have presented in this work a new model for the fuzzy statistical segmentation in the introduction, in priors, of two components of different nature: Dirac's measure, i.e., the "hard" component, and a density, i.e., the "fuzzy" component. We have proposed a procedure of simulation of such a fuzzy random field which takes the spatial interactions into account. Numerous simulations show that this model includes many possibilities which can appear in reality. It remains valid for more than two pure classes and seems to be an alternative to other models using fuzzy membership, such as Kent and Mardia's model [10].

We have proposed four blind statistical methods and tested, in different situations, their theoretical effectiveness and robustness, i.e., their behavior when the used parameter value deviates from the real value. The aim of this study was to understand the behavior, and mainly the stability, of the four unsupervised statistical segmentation methods in which parameters are estimated in previous step by our fuzzy SEM. In order to obtain four unsupervised segmentation methods, we have defined an estimator, based on the SEM algorithm, of all parameters of the marginal "fuzzy" mixture, i.e., the distribution of the noisy observation on each pixel. We have tested its behavior in different situations where the same image was corrupted by different noises. The simulation studies lead us to put forward the following conclusions:

- The fuzzy SEM estimation always gives better results than the SEM when the real data are fuzzy. These results seem to us quite reasonable in the context studied, considering that in fact the data are not independent and rather strongly noised.
- The four fuzzy segmentation methods are robust and three of them present a good theoretical error rate. Furthermore, a more precise analysis of the behavior of these three methods can help select which is best suited. This choice depends on the nature and the correlation of the noise.
- This robustness, combined with the good behavior of the fuzzy SEM, allows us to define three new methods of unsupervised fuzzy statistical segmentation which are always better than the classical SEM-based Bayesian segmentation method, without the fuzzy component in priors.

A topic for prospect of future investigations is the design of contextual fuzzy unsupervised segmentation methods. In fact, it is well known that in the "hard" case the use of the spatial context greatly improves the efficiency of the supervised, i.e. based on the true parameters, segmentation. In exchange, the estimation of parameters becomes more difficult. As in studied situations the behavior of the fuzzy SEM is good and, what is more, the proposed segmentation methods are robust, it would be interesting to investigate the possibilities of the contextual methods. Such methods would undoubtedly be very time consuming, especially when the number of classes exceeds two. This leads to the following problem: How to determine situations in which the use of fuzzy methods is relevant? We intend to devote our further research to these different questions.

Finally, we can conclude that when the real data are fuzzy, the use of the fuzzy segmentation is always more effective than the use of the hard one. The proposed segmentation methods are robust and the behavior of the fuzzy SEM is good, which suggests that the use of the contextual methods is relevant. An important problem to be solved is to find an automatic way i.e., from the data $X = x$, for deciding if the fuzzy component of real data is rich enough to justify the use of fuzzy methods.

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition and Fuzzy Objective Function Algorithm.* New York: Plenum Press, 1981.

[2] H. Caillol and W. Pieczynski, "Fuzzy statistical unsupervised image segmentation," in *Proc. IGARSS'92,* Houston, TX, 1992.

[3] J. R. Casas, A. Hillion, C. Roux, L. Torres, and A. Gasull, "Fuzzy classification of remote sensing images: A pseudocolor representation of fuzzy partitions," presented at the Conf. Neural and Stochastic Methods in Image and Signal Processing, 1992.

[4] G. Celeux and J. Diebold, "L'algorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités," *Revue de Statistique Appliquée,* vol. 34, no. 2, 1986.

[5] B. B. Devi and V. V. S. Sarma, "Estimation of fuzzy memberships from histograms," *Information Sciences,* vol. 35, pp. 43–59, 1985.

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

[7] T. Gu and B. Dubuisson, *Similarity of Classes and Fuzzy Clustering, Fuzzy Sets and Systems,* vol. 34. Amsterdam: North Holland, 1990, pp. 213–221.

[8] A. Hillion, "Les approches statistiques pour la reconnaissance des images de télédétection," *Atti della XXXVI Riunione Scientifica,* SIS, vol. 1, pp. 287–297, 1992.

[9] T. L. Hunstberger, C. L. Jacobs, and R. Cannon, "Iterative fuzzy image segmentation," *Patt. Recog.,* vol. 18, no. 2, pp. 131–138, 1985.

[10] J. T. Kent and K. V. Mardia, "Spatial classification using fuzzy membership," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. 10, no. 5, 1991.

[11] N. Marhic, P. Masson, and W. Pieczynski, "Mélange de lois et segmentation non supervisée des données SPOT," *Statistique et Analyse des Données,* vol. 16, no. 2, 1991.

[12] N. Marhic and W. Pieczynski, "Estimation of mixture and unsupervised segmentation of images," in *Proc. IGARRS'91,* Helsinki, Finland, 1991.

[13] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite data," *IEEE Trans. Geosci. Remote Sensing,* vol. 31, no. 3. pp. 618–633, May 1993.

[14] W. Pedrycz, "Fuzzy sets in pattern recognition: Methodology and methods," *Patt. Recog.,* vol. 23, no. 1/2, pp. 121–146, 1990.

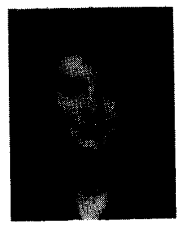
[15] W. Pieczynski, "Estimation of context in random field," *J. Appl. Stat.,* vol. 16, no. 2, 1989.

[16] W. Pieczynski, "Statistical image segmentation," *Machine Graphics and Vision,* vol. 1, no. 1/2.

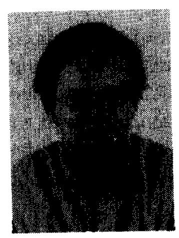
[17] H. C. Quelle, J. M. Boucher, and W. Pieczynski, "Unsupervised Bayesian classification of SAR images," in *Proc. IGARSS'91,* Helsinki, Finland, 1991.

[18] H. C. Quelle, J. M. Boucher, and W. Pieczynski, "Local parameter estimation and unsupervised segmentation of SAR images," in *Proc. IGARSS'92,* Houston, TX, 1992.

[19] F. Wang, "Fuzzy supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sensing,* vol. 28, 1990.

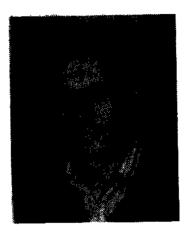


Hélène Caillol was born in France in 1968. She received the Diplome d'Etudes Approfondies degree from the Université de Provence, Marseille, France, in 1991. She is currently completing the Ph.D. degree at the Institut National des Télécommunications, Evry, France. Her research interest include statistical image segmentation, fuzzy image segmentation, and mathematical statistics.



Alain Hillion (A'87) was born in France in 1947. He received the Agregation de Mathématique degree from the Ecole Normale Supérieure in 1970 and the Doctorat d'Etat degree from Pierre et Marie Curie University, Paris, in 1980.

He is currently Professor and Head of the Department of Mathematics and Communication Systems at the Ecole Nationale Supérieure des Télécommunications de Bretagne, Brest, France. His areas of research interest include mathematical statistics, pattern recognition, and signal and image processing.



Wojciech Pieczynski was born in Poland in 1955. He received the Doctorat d'Etat degree from Pierre et Marie Curie University, Paris, France, in 1986.

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