SEM Algorithm and Unsupervised Statistical Segmentation of Satellite Images

Pascale Masson and Wojciech Pieczynski

Abstract—**This work addresses Bayesian unsupervised satellite image segmentation. We propose, as an alternative to global methods like MAP or MPM, the use of contextual ones, which is partially justified by previous works. We show, via a simulation study, that spatial or spectral context contribution is sensitive to image parameters such as homogeneity, means, variances, and spatial or spectral correlations of the noise. From this one may choose the best context contribution according to the estimated values of the above parameters. The parameter estimation step is treated by the SEM, a densities mixture estimator which is a stochastic variant of the EM algorithm. Another simulation study shows good robustness of the SEM algorithm with respect to different image parameters. Thus modification of the behavior of the contextual methods, when the SEM-based unsupervised approaches are considered, remains limited and the conclusions of the supervised simulation study stay valid. We propose an "adaptive unsupervised method" using more relevant contextual features. Furthermore, we apply different SEM-based unsupervised contextual segmentation methods to two real SPOT images and observe that the results obtained are consistently better than those obtained by a classical histogram based method.**

Index Terms— **Random fields, image segmentation, mixture estimation, Bayesian classification, unsupervised segmentation.**

I. Introduction

Our study deals with unsupervised image segmentation. We adopt the statistical approach which assumes modeling by random fields. With S the set of pixels, two collections of random variables $X = (X_s)_{s\in S}, Y = (Y_s)_{s\in S}$ called "random fields" are considered. The first one models the field of "classes": each X_s takes its values in a finite set $\Omega = \{\omega_1, \omega_2, \cdots, \omega_m\}$. The second one is the field of "measurements" or "observations": each Y_s takes its values in R^d, with $d \in \{1, 2, \cdots, n, \cdots\}$. Thus, in modeling we adopt $Y_s = y_s$ as a "noisy" observation of the "class" $X_s = x_s$. Let us clarify the two concepts "noise" and "class" used in what follows. Let us suppose that we have to segment an image where "forest" and "water" are present. Thus, in our modeling, we have two classes. The class "forest" does not provide a unique measure because of the "natural variability", and it is the same for the class "water". However, although this "natural variability" is not a noise in the common "anti-information" meaning, we will in order to simplify things call it "noise". It will be seen as a part of the "global" noise, the other part being the "real noise" due to the transmission of the data. Let us note that our model does not allow us to take into account pixels where "forest" and "water" are simultaneously present; to do so we would have to incorporate "fuzzy" modeling [3], [15].

In the statistical context the problem of segmentation is that of estimating the "hidden" or "unobserved" realization of X from the "observed" realization of Y. The Bayesian approach, which is widely used, gives rise to two families of methods: global and local.

The global methods take into account all the information $Y = y$ simultaneously. There are two global Bayesian methods: the MAP method consists of estimating the unobservable realization of X by x, whose *a posteriori* (conditioned upon $Y = y$) probability is maximized. The MPM chooses for the estimated realization $x = (x_s)_{s\in S}$ such that for each $s \in S, x_s$ maximizes the *a posteriori* marginal probability (i.e., the distribution of X_s conditioned upon $Y = y$). In fact, both MAP and MPM are Bayesian solutions corresponding to different loss functions. Neither MAP nor MPM are computable and one must resort to iterative methods. Such methods exist under the following two hypotheses: X is a Markov random field and the random variables (Y_s) are independent conditionally to every realization of X. The solution of MAP can be approached by the simulated annealing algorithm [11] and one can use Marroquin *et al.*'s algorithm [18] in order to approach the solution of the MPM. When the simulation model verifies the above hypotheses and when all useful parameters are known both MAP and MPM give excellent results.

The local, or contextual, methods consist of estimating the realization of each X_s from the information contained in a small neighborhood of s. These methods are much simpler: When all parameters are known one need only compute the classical discrimination functions.

In the unsupervised case, as described in this work, things are more difficult. One has to estimate, in a previous step or simultaneously, all parameters relevant to the chosen segmentation method. This is a difficult problem in the case of global methods. The estimation step can be performed by EM, Stochastic Gradient (SG, [32]) or Iterative Conditional Estimation (ICE, [24], [25]) methods. The use of the EM algorithm is nontrivial and its theoretical justification requires strong hypotheses [6], [8], [26]. In particular, most authors assume that the noise variance is the same for all classes. SG and ICE seem better suited in this context but the important problem of the choice of the energy form (see next section) remains. Furthermore, we showed in our previous work [16],

Manuscript received March 23, 1992; revised November 11, 1992.
P. Masson is with the Groupe Traitement d'Images, Département Mathématiques et Systèmes de Communications, Ecole Nationale Supérieure des Télécommunications de Bretagne, B.P. 832-29285 Brest cedex, France.
W. Pieczynski is with the Groupe Image, Département Systèmes et Réseaux, Institut National des Télécommunications, 91011 Evry cedex, France.
IEEE Log Number 9207811.

0196-2892/93$03.00

[17] that these methods can become very unstable, even when the exact form of the energy is known, in the case of nonhomogeneous images and correlated noise. Thus, we presume that these methods are not well adapted to deal with SPOT images containing urban areas. In fact, these areas are not particularly homogeneous, and estimations described below show that the noise is usually strongly correlated and that its variance depends on the class.

Thus we opt for local methods. By doing so we do not exhaust the choice problem. In fact, when dealing with SPOT data we have several spectral bands and we are rapidly faced with computational problems. For instance, if there are eight classes and if we want to take into account two spectral bands and use a context containing just one neighbor we have already a mixture of 64 distributions defined on R^4. This leads us to the principal goals of this work. First, in what circumstances is the use of a context relevant? If it is, should we choose to exploit the information concentrated in several spectral bands or the information contained in some spatial context? What could be criteria to make a better choice? We try to answer these questions via numerous simulations. In our method the estimation step is treated by the recent Stochastic Estimation Maximization (SEM, [4], [5]) algorithm which is a stochastic improvement of the well-known Estimation Maximization (EM, [7]) method.

The organization of this paper is as follows. In the next section we briefly recall the hierarchical model generally used in problems of statistical image segmentation. The third section is devoted to the SEM algorithm, which is described in some detail. The fourth section contains some simulation results and in the fifth we show the correct behavior of the SEM algorithm in the context considered. The results of unsupervised segmentations of two real images are given in the sixth section and the seventh section contains the concluding remarks.

II. Hierarchical Model

As mentioned in the introduction we consider two random fields: the field of "classes" $X = (X_s)_{s\in S}$, and the field of "measurements" $Y = (Y_s)_{s\in S}$. Each X_s takes its values in a finite set $\Omega = \{\omega_1, \omega_2, \cdots, \omega_m\}$ of classes and Y_s in R^d, with $d \in \{1, 2, \cdots, n, \cdots\}$. The problem of segmentation is the problem of estimating an "unknown" realization of X from an "observed" realization of Y. We will suppose that realizations of Y depend on realizations of X as well as two noises of a different nature: "natural variability" or "texture," and "transmission." The distribution of (X, Y) is defined by P_X, distribution of X, and the family P_Y^ϵ of distributions of Y conditioned on $X = \epsilon$. The field X will be assumed Markovian with all realizations possible; thus its distribution is a Gibbs distribution

$$P[X = \epsilon] = P_X[\epsilon] = ce^{-U_\alpha(\epsilon)} \quad (1)$$

where c is unknown and the form of U ("energy") is simple enough to allow the computation of conditional distributions $P[X_t/X_s, s \neq t]$ of each X_t. The field modeling the "natural variability" will be denoted Z.

The distribution of Z is defined by m distributions $P^1, P^2, \cdots, P^m$ of Z conditional to m uniform realizations ($X_s = \omega_1$ for each $s, X_s = \omega_2$ for each $s, \cdots, X_s = \omega_m$ for each s, respectively) and the hypothesis according to which of the random variables Z_i (with Z_i restriction of Z to $Q_i = \{s \in S/\zeta_s = \omega_i\}$) are independent. In order to allow simulations of $P^1, P^2, \cdots, P^m$ we will suppose that they are Gibbs distributions (each "texture" or "natural variability" is modeled by a Markov random field). The "transmission noise" will be denoted by $N = (N_s)_{s\in S}$. Once more the field N will be assumed to be Markovian; furthermore, the random variables (X, Z) and N will be supposed independent. Finally the observed field Y will depend on (Z, N) in some way: $Y = \Psi(Z, N)$. When considering satellite images one often takes $Y = Z + N$ for optical and $Y = ZN$ for radar data.

We will not use explicitly the above model when dealing with the local methods except for image simulations. Let us denote by W the set containing the considered pixel and the context used. The aim of the previous step is to estimate the distribution of (X_W, Y_W). In the sequel we will suppose the field Y is Gaussian conditionally to each realization of X. Then the distribution of (X_W, Y_W) is defined by the parameter (α, β), where α designates the priors (distribution of X_W) and β defines the distributions of Y_W conditional to X_W. So, if W contains q pixels, α has m^q components (it is a distribution on Ω^q) and β is composed of m^q mean vectors $\mu_1, \cdots, \mu_{m^q}$ in R^{qd} and m^q covariance matrices $\Gamma_1, \cdots, \Gamma_{m^q}$ of size $qd \times qd$ (each Y_s takes its values in R^d). Thus we have to estimate (α, β) from a sequence $Y_{W_1}, Y_{W_2}, \cdots, Y_{W_n}$ of restrictions of Y to a sequence $W_1, W_2, \cdots, W_n$ of sets of shape W in S. The distribution of Y_W is a mixture of distributions, and thus the previous problem is that of mixture estimation. We propose the use of the SEM algorithm, which is a stochastic version the EM method and is well adapted to the problem [4], [5].

As we said in the introduction, contrary to global methods, we can use directly the discriminating functions, once the parameters (α, β) are known. Let us clarify this point. Each pixel is classified, i.e., each unknown realization of X_s is estimated, by the class ω_i whose probability conditional on Y_W is maximal. The latter condition is equivalent to:

$$f_i(y_W) = \sup_{1\leq j\leq m} f_j(y_W)$$

$f_1, f_2, \cdots, f_m$ being the discriminating functions defined by:

$$f_j(y_w) = \sum_{\epsilon\in\Omega^{q-1}} P[X_s = \omega_j, X_{W-\{s\}} = \epsilon] f_\epsilon(y_w)$$

where f_ϵ is the density of the distribution of Y_W conditioned on $X_W = (\omega_j, \epsilon)$.

For instance, if we consider two classes, $m = 2$, one spectral band, $d = 1$, and the context reduced to one neighbor, $q = 2$, we have two discriminating functions defined on R^2

$$f_i = \pi_{i1} f_{i1} + \pi_{i2} f_{i2}$$

where $\pi_{ij} = P[(X_s, X_t) = (\omega_i, \omega_j)]$ and f_{ij} is the density of the distribution of (Y_s, Y_t) conditional on $(X_s, X_t) = (\omega_i, \omega_j)$. The parameter α is then $\alpha = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$ and, when f_{ij} are assumed Gaussian, $\beta = (m_1, m_2, \sigma_1, \sigma_2, \rho_{11}, \rho_{12}, \rho_{22})$, where m_i and σ_i are respectively mean and standard deviation of Y_s conditional on $X_s = \omega_i$ and ρ_{ij} is the covariance of Y_s, Y_t conditional on $(X_s, X_t) = (\omega_i, \omega_j)$.

III. SEM Algorithm

The mixture estimation problem above is a particular case of the following one: let us consider a pair of random variables $(\zeta, Y), \zeta$ taking its values in a finite space $E = \{e_1, \cdots, e_K\}$ and Y in R^ν. We change notation in order to simplify things: we have $\nu = qd$ and $K = m^q$ according to the previous notation where q is the cardinal of W, d is the number of spectral bands and m is the number of classes in the image to be segmented. Let us denote by $\Pi_m = P[\zeta = e_m]$ the *a priori* distribution and f_m the Gaussian density (mean μ_m and covariance matrix Γ_m) of the distribution of Y conditioned upon $\zeta = e_m$. Let us put:

$$p_m(y) = P[\zeta = e_m / Y = y]$$

which are the *a posteriori* distributions. We observe a sample $y_1, y_2, \cdots, y_N$ of realizations of Y whose distribution has for density:

$$f_Y = \sum_{p=1}^{K} \Pi_p f_p \tag{2}$$

and we have to estimate the parameter $(\Pi_m, \mu_m, \Gamma_m), 1 \leq m \leq K$.

There are a number of iterative methods for dealing with this problem; one can find in [5] numerous references. All methods, except the SEM, have the following two limitations:

- the number of classes is assumed to be known
- the solution depends strongly on the initialization

The first limitation is a serious handicap in satellite image processing: in fact, in real images the number of classes is frequently difficult to determine *a priori.*

The SEM algorithm is an improvement of the EM algorithm obtained by the addition of a stochastic component. This addition results, with respect to the EM algorithm, in the following improvements:

- it is sufficient to know an upper bound on the number of classes
- the solution is essentially independent of the initialization
- the speed of convergence is appreciably improved.

The SEM algorithm works as follows:

Let M be an upper bound on the number of classes and let $\delta \in]0, 1[$ be a chosen threshold.

Initialization

Take, for every observation y_i, a probability $p_m^0(y_i)$ of its belonging to the class $e_m, 1 \leq m \leq K$. In the absence of any information take for $p_m^0(y_i), 1 \leq m \leq K$ the uniform distribution. The superscript will denote the iteration number.

For every $n \geq 0$:

Step S (Stochastic):

For each y_i select from the set of classes $\{e_1, e_2, \cdots, e_K\}$ an element according to the distribution $p_1^n(y_i), \cdots, p_K^n(y_i)$. This selection defines a partition $Q_1^n, Q_2^n, \cdots, Q_K^n$ of the sample $y_1, y_2, \cdots, y_N$.

Step M (Maximization)

The idea of the SEM algorithm is to suppose that every y_i belonging to Q_m^n, for each $1 \leq m \leq K$, is realized according to the distribution defined by f_m, the density corresponding to the class e_m. By denoting $C_m^n = card(Q_m^n), Q_m^n = (y_{1,m}^n, y_{2,m}^n, \cdots, y_{C_m^n,m}^n)$ we can estimate the mean μ_m and the covariance matrix Γ_n^{m+1} by an empirical mean and covariance matrix:

$$\mu_m^{n+1} = \frac{1}{C_m^n} \sum_{i=1}^{C_m^n} y_{i,m}^n \tag{3}$$

$$\Gamma_m^{n+1} = \frac{1}{C_m^n} \sum_{i=1}^{C_m^n} (y_{i,m}^n - \mu_m^n)(y_{i,m}^n - \mu_m^n)^t \tag{4}$$

and the priors by the frequencies:

$$\Pi_m^{n+1} = \frac{C_m^n}{N}. \tag{5}$$

If for m the prior estimated by (5) is inferior to the threshold δ eliminate the corresponding class and go to the initialization.

Step E (Estimation)

For each y_i define the next distribution—$[p_1^{n+1}(y_i), \cdots, p_K^{n+1}(y_i)]$ on the set of classes by the *a posteriori* distribution based on the current parameter $(\Pi_m^{n+1}, \mu_m^{n+1}, \Gamma_m^{n+1}), 1 \leq m \leq K$:

$$p_m^{n+1}(y_i) = \frac{\Pi_m^{n+1} f_m^{n+1}(y_i)}{\sum_{q=1}^{K} \Pi_q^{n+1} f_q^{n+1}(y_i)} \tag{6}$$

where f_m^{n+1} designates the Gaussian distribution corresponding to $\mu_m^{n+1}, \Gamma_m^{n+1}$.

Return to step S.

Numerous simulations [4] show the correct behavior of the SEM. Its theoretical study is performed in the case of the mixture of two Gaussian distributions [5].

The SEM algorithm has been presented by authors as a viable alternative to the EM algorithm which is widely used in mixture estimation problems. EM is a deterministic, i.e., without stochastic sampling, iterative method. The next value $(\Pi_m^{n+1}, \mu_m^{n+1}, \Gamma_m^{n+1})_{1 \leq m \leq K}$ of the parameter is obtained from the current value $(\Pi_m^n, \mu_m^n, \Gamma_m^n)_{1 \leq m \leq K}$ and the sample

$y_1, y_2, \cdots, y_N$ by the formulae:

$$\Pi_m^{n+1} = \frac{1}{N}\sum_{i=1}^{N} P_m^n(y_i)$$

$$\mu_m^{n+1} = \frac{\sum_{i=1}^{N} y_i P_m^n(y_i)}{\sum_{i=1}^{N} P_m^n(y_i)}$$

$$\Gamma_m^{n+1} = \frac{\sum_{i=1}^{N} (y_i - \mu_m^{n+1})(y_i - \mu_m^{n+1})^t P_m^n(y_i)}{\sum_{i=1}^{N} P_m^n(y_i)}.$$

The main drawback of this method is the large dependence of the solution on the initialization.

IV. Contextual Segmentation

In this section we do not consider the parameter estimation problem. Our purpose is to study the relevance of using the spatial or spectral context. We present the results of numerous simulations which express the contribution of spatial or spectral context with respect the number of neighbors used, signal-to-noise ratio (SNR), homogeneity of the image, and correlations of the noise. Let us point out that we do not need a precise definition of the "signal-to-noise ratio" in what follows. We will say that this ratio decreases when the image becomes noisy and, in each case, this evolution will be clear with the evolution of specific parameters. What we are interested in is above all the evolutions of different methods when images become noisy. The aim of this study is to deduce the most relevant choice according to the parameters above. Our study is limited to binary images and Gaussian noise.

In order to distinguish different cases let us introduce the following notations. The MD will designate the "means discriminating" case: at each pixel the two Gaussian distributions corresponding to two classes only differ by the means (they have the same variance). The "variance discriminating" case, denoted by VD, is the case where these distributions, having the same mean, only differ by variances. These cases are extreme: in real situations we have generally to deal with "mixed", i.e., "means" and "variances" discriminating, case. In each of them we can be interested in the contribution of the spectral or spatial context: we will put spe in the first case and spa in the second one. Thus, for instance, speVD is the case where the model is "variance discriminating" and we are interested in spectrally contextual and spatially blind segmentation. In each of cases speMD, speVD, spaMD, spaVD the noise can be independent or correlated: we will denote speMDI, speVDI, spaMDI, spaVDI, speMDC, speVDC, spaMDC, spaVDC the corresponding possibilities. In "spe" I and C designate the spectral independency and correlation respectively and in "spa" they designate the spatial ones. Let us note that speMDI and spaMDC can refer to a same model: The noise can be spectrally independent and spatially correlated. The difference is that in the first case we look at the spectral context contribution and in the second case at the spatial one. The same remark can be made about the cases speMDC and spaMDI and these two remarks remain valid when "MD" is replaced by "VD". As we said above the "spe" cases are spatially blind, i.e., the corresponding segmentation methods use just one pixel, and thus the homogeneity of the image does not play any role. On the other hand, the homogeneity can be important when considering the "spa" cases [16]. Thus, in each of four "spa" cases we will consider two possibilities: a homogeneous image: spaMDIH, spaVDIH, spaMDCH, spaVDCH and a nonhomogeneous one: spaMDIN, spaVDIN, and spaMDCN, spaVDCN.

Thus we have 12 cases to study. In each of them we look at contexts of different sizes. This point is important given that the computational time strongly depends upon the size of the context used. Furthermore, in each of them we look at different signal to noise ratios: In MD cases we vary the distance between the means and, in VD cases, the distance between the standard deviations.

Finally, we propose below the study of the following cases: Spe MDI, Spe MDC (Table I), Spe VDI, Spe VDC (Table II), Spa MDIH, Spa MDCH (Table IV), Spa MDIN, Spa MDCN (Table V) and Spa VDI, Spa VDC (Table VI). The results are given in percent of incorrectly classified pixels.

A. Contribution of the Spectral Context

Let us denote by $m_1, m_2, \sigma_1, \sigma_2$ the noise means and standard deviations, which will be assumed to be the same in each spectral band, corresponding to the classes ω_1, ω_2 considered. σ will designate the spectral noise correlation. Furthermore, let us put $\Delta = m_2 - m_1$ and let us denote by d the number of spectral bands. The elements of tables below are the percent of wrongly classified points, which can be explicitly computed. We do not consider the spatial context, so the homogeneity does not intervene in this section. We give results concerning three cases: means discriminating $(m_1 \neq m_2, \sigma_1 = \sigma_2)$, standard deviations discriminating $(m_1 = m_2, \sigma_1 \neq \sigma_2)$, and mixed $(m_1 \neq m_2, \sigma_1 \neq \sigma_2)$. We put the case $d = 1$ apart, as it corresponds to blind segmentation and is a "reference case".

1) Means Discriminating Case: $(m_1 \neq m_2, \sigma_1 = \sigma_2)$: This case is the most commonly studied. As we mentioned in the introduction the condition $\sigma_1 = \sigma_2$ is required when using an unsupervised global method based on the EM algorithm [6], [8], [26].

Thus Table I corresponds to the cases Spe MDI, Spe MDC (we look at the contribution of the spectral context, the cases studied are "mean discriminating", the noise is independent or correlated).

The results obtained are not surprising, however we note that in the case of strong correlation, $\sigma = 0.8$, the contribution of the spectral context is negligible.

2) Standard Deviations Discriminating Case: $(m_1 = m_2, \sigma_1 \neq \sigma_2)$: It is possible to show that in this case the Bayesian classification error is independent of the spectral

TABLE I
PERCENT OF THE MISCLASSIFIED PIXELS IN THE CASES speMDI AND speMDC WITH $\Delta = m_2 - m_1$, d NUMBER OF SPECTRAL BANDS, σ SPECTRAL CORRELATION OF THE NOISE

	Δ=1				Δ=2				Δ=3			
d=1	30.9				16				6.5			
	d=2	d=5	d=9	d=100	d=2	d=5	d=9	d=100	d=2	d=5	d=9	d=100
σ=0	23.9	13.1	6.7	0.0	7.5	2.2	1.4	0.0	1.7	0.05	0.0	0.0
σ=0.4	27.4	24.5	23.2	21.7	11.7	8.2	7.2	5.8	3.7	1.9	1.4	0.6
σ=0.8	29.8	29.4	29.1	28.9	14.7	13.8	13.6	13.1	5.5	5.1	4.9	4.7

TABLE II
PERCENT OF THE MISCLASSIFIED PIXELS IN THE CASES speVD WITH σ_1 AND σ_2 STANDARD DEVIATIONS OF THE NOISE CORRESPONDING TO THE CLASSES, $r = \sigma_1/\sigma_2$ AND d NUMBER OF SPECTRAL BANDS

	r=2				r=3				r=4			
d=1	33.8				25.8				20.9			
	d=2	d=5	d=9	d=20	d=2	d=5	d=9	d=20	d=2	d=5	d=9	d=20
σ=0	26.4	14.7	7.7	1.0	16.1	5.1	1.3	0.03	10.8	2.1	0.3	0.00

TABLE III
PERCENT OF THE MISCLASSIFIED PIXELS IN THE MIXED CASE: $m_2 m_1 = 2$ $\sigma_1 = 1, \sigma_2 = 2$ IN EACH OF d SPECTRAL BANDS

d=1	22.4		
	d=2	d=5	d=9
σ=0	13.5	2.7	0.9
σ=0.4	16.4	8.0	4.8
σ=0.8	18.3	11.4	9.7

correlation of the noise. Let us put $r = \sigma_1/\sigma_2$. The results obtained are given in Table II.

We observe that the context contribution is quite significant in this case.

3) Mixed Case: $(m_1 \neq m_2, \sigma_1 \neq \sigma_2)$: The explicit computation of the Bayesian error classification is impossible, so the results in Table III are obtained by simulations using the Monte Carlo method. As it was possible to guess, the results obtained are halfway between the two previous cases. In practice, we must nearly always deal with the third case. Thus we can say in conclusion that the contribution of the spectral context is relevant when the considered case is close to the "standard deviations discriminating" one and useless, except in the case of a little noise correlation, when it is close to the "means discriminating" one.

B. Contribution of the Spatial Context

When considering the contribution of the spatial context the homogeneity of the image considered is important; indeed, it influences the discriminating functions via the priors (see Section II). So we will consider, for each possibility above, two kinds of priors: Π_1 will denote priors corresponding to a homogeneous image (Im.1) and Π_2 those corresponding to a nonhomogeneous one (Im.2). For each context considered Π_1, Π_2 are estimated from Im.1, Im.2, respectively. The Bayesian classification errors below are estimated by the Monte Carlo method. We consider one spectral band and N denotes the number of pixels (the pixel considered and the neighbors in the context used).

Im.1

Im.2

1) Means Discriminating Case: $(m_1 \neq m_2, \sigma_1 = \sigma_2)$: The results corresponding to Π_1 are in Table IV and those corresponding to Π_2 are in Table V.

We can notice that, when the means discriminating case is considered, the homogeneity of images plays a major part in the behavior of Bayesian segmentation. When the image is homogeneous and the noise independent it is interesting to use the spatial context. This tendency is reversed in the case of a nonhomogeneous image: the contribution of the spatial context increases with the spatial correlation. This property of the local methods is particularly interesting. In fact, in real satellite images the noise is often strongly correlated and some of them contain very few homogeneous parts, like urban areas. On the other hand, we showed in our previous work that global methods are completely ineffective in this case [16], [17].

Finally, the contribution of the spatial context, in the "mean discriminating" case, is significant when the image is very homogeneous and the noise nearly uncorrelated or when the image is not homogeneous and the noise strongly correlated.

TABLE IV
CASES spaMDIH AND spaMDCH WITH $\Delta = m_2 m_1, N$: NUMBER OF PIXELS (THE PIXEL CONSIDERED AND THE NEIGHBORS) AND ρ: SPATIAL CORRELATION OF THE NOISE

	Δ=1			Δ=2			Δ=3		
N=1	30.9			16			6.5		
	N=2	N=5	N=9	N=2	N=5	N=9	N=2	N=5	N=9
ρ=0	25.6	14.6	11.9	10.4	3.8	2.0	3.8	1.1	1.0
ρ=0.4	29.3	26.2	25.8	13.7	10.2	9.3	5.1	2.5	2.1
ρ=0.8	30.9	28.9	30.0	15.0	13.8	12.6	6.0	4.3	4.0

TABLE V
CASES spaMDIN AND spaMDCN WITH $\Delta = m_2 m_1, N$: NUMBER OF PIXELS (THE PIXEL CONSIDERED AND THE NEIGHBORS) AND ρ: SPATIAL CORRELATION OF THE NOISE

	Δ=1			Δ=2			Δ=3		
N=1	30.9			16			6.5		
	N=2	N=5	N=9	N=2	N=5	N=9	N=2	N=5	N=9
ρ=0	30.8	29.0	28.7	15.6	14.5	14.5	6.5	5.9	6.2
ρ=0.4	30.4	28.0	28.2	14.7	13.4	12.6	5.5	4.3	3.3
ρ=0.8	28.0	25.5	24.6	12.7	10.3	7.5	4.3	2.7	1.4

TABLE VI
CASES spaVDI AND spaVDI WITH $r = \sigma_2/\sigma_1, N$: NUMBER OF PIXELS AND ρ: SPATIAL CORRELATION OF THE NOISE

	r=2			r=3			r=4		
N=1	33.0			21.5			20.6		
	N=2	N=5	N=9	N=2	N=5	N=9	N=2	N=5	N=9
ρ=0	27.8	18.6	13.8	17.9	9.6	8.3	13.2	6.9	7.0
ρ=0.4	27.9	18.8	14.8	17.8	10.0	8.8	13.2	7.5	8.0
ρ=0.8	27.8	20.1	19.2	17.9	11.8	12.0	13.2	9.0	10.0

2) *Standard Deviations Discriminating Case:* ($m_1 = m_2, \sigma_1 \neq \sigma_2$): As above we put $r = \sigma_2/\sigma_1$. The homogeneity does not play any part in this case, and the results obtained are shown in Table VI.

We note that the spatial context contribution is much less relevant than the spectral context one in this case; however, in some situations it can turn out to be nonnegligible.

C. Conclusions and a "Context Choice" Algorithm

We propose the following evaluation of the context contribution in each of the cases considered. Let ER_b, ER_c designate the blind and contextual errors. The context contribution CC can be defined by:

$$CC = \frac{ER_b - ER_c}{ER_b}.$$

This gives us an idea about the importance of the additional contextual information. To be more specific, CC can be seen as the proportion of pixels well classified by the corresponding contextual method of the pixels wrongly classified by the blind method. CC so defined depends on the number of neighbors in the context used and the SNR. We take four neighbors in the case of the spatial context, thus $N = 5$, and 5 spectral bands, $d = 5$, in the case of the spectral context. In the MD case we will take for CC the mean of the CC corresponding to the cases $\Delta = 1$ *and* $\Delta = 2$, and, in the VD case it will be the mean of the CC corresponding to $r = 2, r = 3$. We will take for the noise correlation $\rho = \sigma = 0.8$ in both spatial and spectral correlated noise cases.

As an example let us compute the CC in the case speMDI. We have to consider the cases $\Delta = 1$ and $\Delta = 2$. According to Table I, ER_b and ER_c are in the first case respectively 30.9 and 13.1 (we look at the errors corresponding to $d = 1$ and $d = 5, \sigma = 0$). Thus the "context contribution" corresponding to $\Delta = 1$ is:

$$CC_{\Delta=1} = \frac{ER_b - ER_c}{ER_b} = \frac{30.9 - 13.1}{30.9} = 0.576.$$

By looking at the column $\Delta = 2$ in Table I we obtain in the same way:

$$CC_{\Delta=2} = \frac{ER_b - ER_c}{ER_b} = \frac{16.0 - 2.2}{16.0} = 0.863.$$

Finally, CC defined as the mean of $CC_{\Delta=1}$ and $CC_{\Delta=2}$ is:

$$CC = \tfrac{1}{2}(0.576 + 0.863) = 0.72.$$

The results obtained are shown in Table VIIa.

When ordering these results according to the context contribution we obtain Table VIIb.

Thus we can put forth the following two conclusions:

1) The VD case is always more relevant that the MD case. Moreover, it is relevant in all situations tested and is,

TABLE VIIa

	CC		CC
		Spa VDC	0.53
Spa MDIH	0.65	Spe MDI	0.72
Spa MDCH	0.10	Spe MDC	0.09
Spa MDIN	0.08	Spe VDI	0.85
Spa MDCN	0.27	Spe VDC	0.85
Spa VDI	0.64		

TABLE VIIb

Spe VDI	0.85	Spa VDC	0.53
Spe VDC	0.85	Spa MDCN	0.27
Spe MDI	0.72	Spa MDCH	0.10
Spa MDIH	0.65	Spe MDC	0.09
Spa VDI	0.64	Spa MDIN	0.08

in both spectral and spatial case, widely independent of the noise correlation.

2) The MD case is more complicated. The spectral context contribution strongly depends on the noise correlation: it can be dominant or insignificant. When using the spatial context the relevant cases are "independent noise and homogeneous image" or "correlated noise and nonhomogeneous image". As we shall see below the latter case is especially relevant when dealing with SPOT images containing urban areas.

The results above lead us to propose the following "context choice" algorithm:

If Spe VD use spectral context, if not see Spe MDI
If Spe MDI use spectral context, if not see Spa MDIH
If Spa MDIH use spatial context, if not see Spa VD
If Spa VD use spatial context, if not see Spa MDCN
If Spa MDCN use spatial context, if not use blind segmentation

This is to be read as follows: first decide if the considered image is of type Spe VD, which is, according to Table VII, the most relevant case. If so, use the spectral context; if not, decide if it is of type Spe MDI which remains relevant when the spectral context contribution is considered. If so, use the spectral context; if not, decide if it is of Spa MDIH type, and so on.

V. Parameter Estimation by the SEM and Unsupervised Segmentation

This section treats the crucial aspect of the problem: what is the behavior of the unsupervised method with respect to the parameters like SNR or noise correlation. The theoretical study of this problem is, without doubt, extremely tedious. In fact, its efficiency depends on three independent factors:

1) the behavior of the contextual method based on the real parameters, an aspect studied in the previous section;
2) the behavior of the parameter estimators used;
3) the robustness of the segmentation method with respect to the parameters.

Here we are interested in the points 2), 3). Note that 2) has its own importance outside the segmentation problem.

A. Blind Unsupervised Segmentation

We begin with the study of the blind segmentation. Points 2), 3) are examined with respect to two discriminating cases (means or standard deviations), SNR and the noise correlation.

1) Means Discriminating Case: We take $\sigma_1 = \sigma_2 = 1$ and consider two cases: $m_2 - m_1 = 1, m_2 - m_1 = 2$. For each of them case A designates the independent noise and case B the strongly correlated one (for t, s neighbors $\mathrm{cov}(Y_t, Y_s) = 0.72$ if $X_t = X_s$ and 0.04 if $X_t \neq X_s$). We designate by E.V. (empirical values) the estimations obtained from the image and its noisy observation by the classical estimators, i.e., frequency and empirical means and variances. The results obtained are shown in Table VIII.

We note that the efficiency of the SEM is excellent in this case and approaches the efficiency of the E.V. This is particularly striking in the case $m_1 - m_2 = 1$ where the noise is significant. Furthermore, the efficiency of the SEM based unsupervised method is comparable with the efficiency of the EV based method which can be seen as a supervised one. This was not automatically implied by the good behavior of the SEM, in fact, if the segmentation method used were weakly robust the efficiency of the unsupervised SEM based one could have been significantly degraded.

2) Standard Deviations Discriminating Case: We take here $m_1 = m_2 = 1$ and study two cases: $(\sigma_1 = 1, \sigma_2 = 2), (\sigma_1 = 1, \sigma_2 = 3)$. The results obtained are shown in Table IX.

We note that, contrary to the former case, the efficiency of the unsupervised method decreases when the noise correlation increases. To be more precise we can say, concerning the points 2), 3) we are interested in, that SEM is as efficient as EV, except for the variance estimation and that the Bayesian method used is sensitive to this parameter.

B. Spatially Contextual Unsupervised Segmentation

We consider the spatial context formed by four nearest neighbors ($N = 5$) and take one spectral band.

1) Means Discriminating Case: The results obtained are found in Table X.

2) Standard Deviations Discriminating Case: The results obtained are found in Table XI.

C. Spectrally Contextual Unsupervised Segmentation

We present the results corresponding to both "means discriminating" and "variances discriminating" cases in Table XII below.

D. Remarks and Unsupervised "Context Choice" Algorithm

According to the numerical results above the efficiency of the SEM seems acceptable. In the case of blind segmentation, where the mixture to estimate contains just two components, the SEM is nearly as efficient as the EV, even in the case of a very weak SNR, which is quite surprising. In the MD case we note that $\sigma_1, \sigma_2, \Pi_1, \Pi_2$ are quite well estimated and the

TABLE VIII

MEAN DISCRIMINATING (MD) CASE. REAL VALUES: $(m_1 = 1, m_2 = 2) or (m_1 = 1, m_2 = 3), \sigma_1 = \sigma_2 = 1, \Pi_1 = \Pi_2 = 0.5$ E. V.: ESTIMATES FROM THE IMAGE AND ITS NOISY VERSION BY FREQUENCIES AND EMPIRICAL MOMENTS, S.E.M.: ESTIMATES FROM THE NOISY VERSION BY S.E.M., A: WHITE NOISE, B: CORRELATED NOISE, τ: RATIO OF WRONGLY CLASSIFIED PIXELS BY THE BLIND BAYESIAN ESTIMATES BASED METHOD.

	$m_2 - m_1 = 1$				$m_2 - m_1 = 2$			
	A		B		A		B	
	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.
m_1	0.96	1.00	0.99	0.91	0.98	0.97	0.99	0.95
m_2	1.96	1.87	1.99	1.99	2.96	2.95	2.99	2.99
σ_1	1.00	1.04	1.03	0.97	1.00	1.00	1.06	1.01
σ_2	0.98	1.05	0.94	0.94	0.99	0.99	0.94	0.88
Π_1	0.52	0.50	0.52	0.49	0.52	0.52	0.52	0.55
Π_2	0.48	0.50	0.48	0.51	0.48	0.48	0.48	0.45
τ (%)	30.7	33.2	31.0	32.5	15.7	14.1	16.3	16.3

TABLE IX

VARIANCE DISCRIMINATING (VD) CASE. REAL VALUES: $m_1 = m_2 = 1, (\sigma_1 = 1, \sigma_2 = 2) or (\sigma_1 = 1, \sigma_2 = 3, \Pi_1 = \Pi_2 = 0.5$. E.V.: ESTIMATES FROM THE IMAGE AND ITS NOISY VERSION BY FREQUENCIES AND EMPIRICAL MOMENTS, S.E.M.: ESTIMATES FROM THE NOISY VERSION BY SEM, A: WHITE NOISE, B: CORRELATED NOISE, τ RATIO OF WRONGLY CLASSIFIED PIXELS BY THE BLIND BAYESIAN ESTIMATES BASED METHOD.

	$(\sigma_1 = 1, \sigma_2 = 2)$				$(\sigma_1 = 1, \sigma_2 = 3)$.			
	A		B		A		B	
	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.
m_1	0.95	0.92	0.99	0.99	0.87	0.87	0.96	0.97
m_2	0.96	0.98	0.97	0.98	0.98	0.97	0.99	0.98
σ_1	0.99	0.89	1.06	1.90	8.93	9.07	1.06	1.85
σ_2	4.07	3.90	3.75	2.93	1.00	0.95	8.43	7.60
Π_1	0.52	0.51	0.52	0.50	0.52	0.47	0.52	0.52
Π_2	0.48	0.49	0.48	0.50	0.48	0.53	0.48	0.48
τ (%)	34.5	34.0	33.5	43.7	25.6	25.8	25.0	30.5

TABLE X

MEAN DISCRIMINATING (MD) CASE. REAL VALUES: $m_1 = m_2 = 1, (\sigma_1 = 1, \sigma_2 = 2) or (\sigma_1 = 1, \sigma_2 = 3, \Pi_1 = \Pi_2 = 0.5$. ESTIMATES FROM THE IMAGE AND ITS NOISY VERSION BY FREQUENCIES AND EMPIRICAL MOMENTS, S.E.M.: ESTIMATES FROM THE NOISY VERSION BY SEM, A: WHITE NOISE $(\rho_{11} = \rho_{12} = \rho_{21} = \rho_{22} = 0)$ B: CORRELATED NOISE, τ:RATIO OF WRONGLY CLASSIFIED PIXELS BY THE CONTEXTUAL (FOUR NEIGHBORS) BAYESIAN ESTIMATES BASED METHOD.

	$m_2 - m_1 = 1$				$m_2 - m_1 = 2$			
	A		B		A		B	
	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.
m_1	0.98	0.92	0.99	1.20	0.98	1.01	0.99	0.91
m_2	1.96	0.98	1.99	1.99	2.96	2.96	2.99	2.95
σ_1	1.00	0.89	1.06	1.17	1.00	1.05	1.06	0.98
σ_2	1.05	3.90	0.94	0.95	0.99	1.00	0.94	0.97
ρ_{11}	0.05	-0.09	0.71	0.72	0.05	0.06	0.71	0.68
ρ_{12}	0.02	-0.06	0.55	0.58	-0.04	-0.06	0.55	0.53
ρ_{13}	-0.04	-0.08	0.50	0.52	0.02	-0.04	0.50	0.46
ρ_{21}	-0.01	-0.01	0.63	0.64	0.01	-0.02	0.61	0.71
ρ_{22}	-0.04	-0.04	0.57	0.57	-0.03	0.06	0.57	0.60
ρ_{23}	-0.05	-0.06	0.46	0.42	-0.05	-0.01	0.46	0.50
Π_1	0.52	0.51	0.52	0.69	0.52	0.54	0.52	0.50
Π_2	0.48	0.49	0.48	0.31	0.48	0.46	0.48	0.50
τ (%)	17.0	21.7	28.0	38.6	4.6	4.9	12.7	13.1

TABLE XI

VARIANCE DISCRIMINATING (VD) CASE REAL VALUES: $m_1 = m_2 = 1, (\sigma_1 = 1, \sigma_2 = 2) or (\sigma_1 = 1, \sigma_2 = 3, \Pi_1 = \Pi_2 = 0.5$. ESTIMATES FROM THE IMAGE AND ITS NOISY VERSION BY FREQUENCIES AND EMPIRICAL MOMENTS, S.E.M.: ESTIMATES FROM THE NOISY VERSION BY SEM, A: WHITE NOISE, B: CORRELATED NOISE, τ: RATIO OF WRONGLY CLASSIFIED PIXELS BY THE CONTEXTUAL (FOUR NEIGHBORS) BAYESIAN ESTIMATES BASED METHOD

	$(\sigma_1 = 1, \sigma_2 = 2)$,				$(\sigma_1 = 1, \sigma_2 = 3)$.			
	A		B		A		B	
	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.
m_1	0.95	0.91	0.99	1.02	0.87	0.87	0.99	1.01
m_2	0.96	0.97	0.97	0.94	0.98	0.96	0.96	0.95
σ_1	4.02	3.88	3.75	3.50	8.93	9.20	8.43	8.09
σ_2	0.99	0.95	0.94	1.03	1.00	1.00	1.06	1.03
ρ_{11}	0.05	0.08	0.71	0.59	0.05	0.02	0.71	0.62
ρ_{12}	-0.04	0.10	0.55	0.49	-0.04	0.13	0.55	0.53
ρ_{13}	0.02	0.09	0.50	0.38	0.02	0.03	0.50	0.42
ρ_{21}	-0.01	0.18	0.63	0.71	-0.01	-0.07	0.63	0.70
ρ_{22}	-0.05	-0.13	0.55	0.58	-0.05	0.11	0.55	0.61
ρ_{23}	0.06	0.11	0.45	0.52	0.04	0.00	0.45	0.53
Π_1	0.52	0.51	0.52	0.52	0.52	0.52	0.52	0.50
Π_2	0.48	0.49	0.48	0.48	0.48	0.48	0.48	0.50
τ (%)	18.0	25.7	20.0	20.4	9.8	13.6	10.2	10.1

TABLE XII

MEANS DISCRIMINATING (MD) AND VARIANCES DISCRIMINATING (VD) CASES. REAL VALUES: $m_{11} = m_{12} = 1, m_{21} = m_{22} = 3, \sigma_{11} = \sigma_{12} = \sigma_{21} = \sigma_{22} = 1, \Pi_1 = \Pi_2 = 0.5$ IN THE M.D. CASE AND $m_{11} = m_{12} = m_{21} = m_{22} = 1, \sigma_{11}, \sigma_{12} = 1, \sigma_{21} = \sigma_{22} = 4, \Pi_1 = \Pi_2 = 0.5$ IN THE CASE VD. E.V.: ESTIMATES FROM THE IMAGE AND ITS NOISY VERSION BY FREQUENCIES AND EMPIRICALMOMENTS, S.E.M.: ESTIMATES FROM THE NOISY VERSION BY SEM, A: WHITE NOISE, B: CORRELATED NOISE, τ: RATIO OF WRONGLY CLASSIFIED PIXELS BY THE CONTEXTUAL (TWO SPECTRAL BANDS) BAYESIAN ESTIMATES BASED METHOD.

	means discrimination $(\sigma_1 = \sigma_2 = 1, m_1 \neq m_2)$				variances discrimination $(\sigma_1 \neq \sigma_2, m_1 = m_2)$			
	C		D		C		D	
	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.	E.V.	S.E.M.
m_{11}	0.98	0.97	1.03	1.25	0.98	1.02	1.03	1.16
m_{12}	1.04	1.04	1.10	1.26	1.00	0.98	1.04	1.17
σ_{11}	1.00	1.00	1.04	1.30	1.04	1.10	1.10	1.03
σ_{12}	0.99	0.99	1.04	1.22	0.99	0.93	1.04	0.97
m_{21}	3.02	3.02	3.04	3.29	1.03	0.98	1.08	0.94
m_{22}	2.99	2.99	3.06	3.33	0.99	1.01	1.13	0.99
σ_{21}	1.00	1.00	0.97	0.74	4.01	3.93	3.89	3.83
σ_{22}	0.94	0.96	0.99	0.76	3.77	3.81	3.97	3.92
Π_1	0.52	0.51	0.52	0.64	0.52	0.52	0.52	0.50
Π_2	0.48	0.49	0.48	0.36	0.48	0.48	0.48	0.50
τ (%)	7.8	7.9	15.0	17.9	26.3	26.8	26.8	27.6

estimation of m_1, m_2 seems to be more problematic. In the VD case, Table IX, we note that the variances are badly estimated in the case of small SNR and correlated noise and this fact strongly degrades the well-classified pixels ratio. The same phenomenon appears in the case $\sigma_1 = 1, \sigma_2 = 3$. Thus we can say in conclusion, when the blind segmentation parameters estimation is concerned, that all estimations are reliable, except possibly the variance estimations in the case of correlated noise.

It would be possible to make the same kind of comments about the results concerning the estimation of the mixture corresponding to spatial or spectral context, MD or VD cases: Tables X–XII. A general tendency does not appear clearly, thus we will just say that in 12 cases studied two are bad: Table X column 4 and Table XI column 2. The first one is due to a bad estimation of priors and the second one to a relatively bad estimation of the correlations. Two other cases: Table IX column 6 and Table XII column 4 are bad, but can be

TABLE XIII
SPATIAL AND SPECTRAL INTRACLASS CORRELATIONS IN IMAGES "GUINEE" AND "LA ROCHELLE"

	Guinée						La Rochelle					
class	1	2	3	4	5	6	1	2	3	4	5	6
σ_{spa}	0.80	0.70	0.51	0.80	0.92	0.72	0.83	0.66	0.48	0.92	0.68	0.53
σ_{spe}	0.71	0.74	0.81	0.49	0.45	0.96	0.86	0.86	0.71	0.84	0.72	0.90

considered as acceptable. All other results are excellent. We have done numerous other simulations, 32 in total, the results of which can be seen in [19]. Six of them can be considered as bad, three of which concern the MD case with $m_1 - m_2 = 1$. Twenty-four of them are excellent.

Thus we put for in the following two conclusions:

1) The SEM is reliable in about 80% of the situations. The MD case with small signal to noise ratio is the most unfavorable.
2) Its efficiency is excellent in about 75% of the situations studied.

Finally, according to the conclusions above, we can propose the following "unsupervised" context choice algorithm:

1) Estimate, using the SEM, means, variances, spectral and spatial correlations, priors.
2) Decide in which case of Table VII they fall.
3) Apply the "context choice" algorithm Section IV.

This "choice" algorithm would be entirely unsupervised if the step 2) above were automatic. That means we would have to define a rule deciding, from the parameters estimated in 1), in what case of Table VII we are. Thus we would have to decide, from the estimated means, variances, spectral and spatial correlations and priors in what case, between Spe MDI, Spe MDC, Spe VDI, Spe VDC, Spa MDIH, Spa MDCH, Spa MDIN, Spa MDCN and Spa VDI, Spa VDC we are. We can decide about the degree of the spatial and spectral correlation from ρ_{spa} and ρ_{spe} which are spatial and spectral correlation coefficients, respectively. We can have an idea about the homogeneity from the distribution, estimated in 1), of (X_t, X_s), for t, s neighbors. In fact, if the image is homogeneous, this distribution charges $\{(\omega_1, \omega_1), (\omega_2, \omega_2)\}$ and charges $\{(\omega_1, \omega_2), (\omega_2 \omega_1)\}$ otherwise. We propose to assume that the following "homogeneity" parameter $\rho_h = P[\{(X_t, X_s) = (\omega_1, \omega_1)\}] + P[\{(X_t, X_s) = (\omega_2, \omega_2)\}]$ measures this homogeneity: ρ_h is close to 1 if the image is homogeneous and it is close to 0 if not. The last parameter we have to define should allow the decision between the "means" and "variances" discrimination. Thus we have to define a function of $m_1, m_2, \sigma_1, \sigma_2$, respectively means and variances corresponding to two classes and estimated in 1), which would be close to, for instance, 1 in the VD case and close to 0 in the MD case. One possible choice is:

$$\rho_d = e^{-|m_1 - m_2/\sigma_1 - \sigma_2|}$$

if $\sigma_1 \neq \sigma_2$ and equal to 0 if $\sigma_1 = \sigma_2$. Thus we have to decide, from $\rho_{spa}, \rho_{spe}, \rho_h, \rho_d$ evaluated from the estimates obtained in 1), in which case among Spe MDI, Spe MDC, Spe VDI, Spe VDC, Spa MDIH, Spa MDCH, Spa MDIN, Spa MDCN and Spa VDI, Spa VDC we are. The choice we can propose at this stage, which is the simplest, consists of deciding, for each $\rho_{spa}, \rho_{spe}, \rho_h, \rho_d$, the possibility corresponding to 0 if it is inferior to 0.5 and the possibility corresponding to 1 if it is superior to 0.5. For instance, if we have $\rho_{spa} = 0.2, \rho_{spe} = 0.8, \rho_h = 0.9, \rho_d = 0.3$ we are in the case SpaMDIH and SpeMDC. Thus, according to Table VIIb is, SpaMDIH is more appropriate and we should use the spatially contextual segmentation. If we have $\rho_{spa} = 0.7, \rho_{spe} = 0.8, \rho_h = 0.9, \rho_d = 0.3$ we are in the case SpaMDCH and SpeMDC. Thus, according to Table VIIb use of any context is not useful and we should call on the simple blind segmentation.

Finally, we can present the following automatical algorithm:

1) Using the SEM estimate the means, variances, spectral and spatial correlations and priors.
2) Compute $\rho_{spa}, \rho_{spe}, \rho_h, \rho_d$.
3) Decide in which case among Spe MDI, Spe MDC, Spe VDI, Spe VDC, Spa MDIH, Spa MDCH, Spa MDIN, Spa MDCN and Spa VDI, Spa VDC the data lie.
4) Choose the better suited method according to the "choice algorithm" Section IV.C.
5) Perform the segmentation.

Let us note that the rule we propose in order to solve 3) is rather simple and further studies should allow us to refine it.

VI. Real Images Segmentation

We consider in this section two real SPOT images, each of them being taken with two spectral bands Y^1, Y^2. The visualization of the data of Y^1 of each of them gives Im.3, Im.4, respectively. The first one represents a region of Guinee Bissau and the second one the town La Rochelle, France.

We propose four unsupervised segmentation methods based on the SEM for each of the two real images available. The blind method will be denoted by B, the spectrally contextual (two spectral bands) and spatially blind method by Cspe, the spectrally blind and spatially contextual (one neighbor) method Cspa and the spectrally and spatially contextual (two spectral bands, one neighbor) one by Cspespa. Koon designates the Koontz *et al.* method [13] obtained from the bidimensional, taken from the two spectral bands, histogram.

A. Parameter Estimation Problem

In both images the blind SEM finds five classes and the contextual one six. The noise is strongly correlated. For instance the estimated intraclass correlations, spatial and spectral are shown in Table XIII.

The quality of the estimations is difficult to appreciate in absence of the real values. We can roughly perform such an evaluation by comparison of the histograms with the probabil-

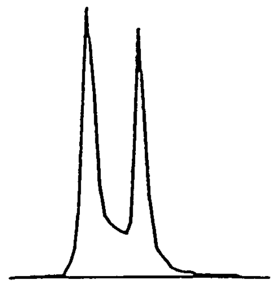


Fig. 1. Histogram of "Guinee" taken from Y^1.

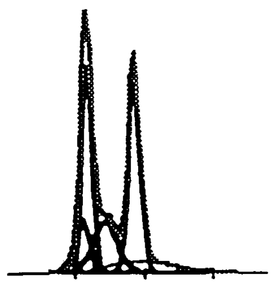


Fig. 2. Density mixture based on the parameters estimated from "Guinee" by SEM.

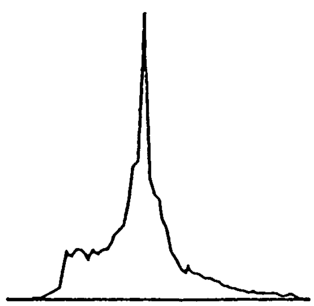


Fig. 3. Histogram of "La Rochelle" taken from Y^1.

ity densities mixture based on the estimated parameters. Fig. 1 represents the histogram of Im3, taken from Y^1, and Fig. 2 the density mixture based on the estimated parameters. It is the same, with respect of Im4, for Figs. 3 and 4.

We note that in both cases "Guinee" and "La Rochelle" histograms seem quite close to the mixture densities based on parameters estimated by SEM. The fact that SEM finds six classes is rather surprising and indicates a good appropriateness of this algorithm in this context.

The bidimensional histograms of Im3, Im4 taken from the two spectral bands are represented by Figs. 5 and 7, and the corresponding mixture densities based on the estimated parameters are given by Figs. 6 and 8, respectively.

The SEM estimation seems to give good results in the two real image cases considered. In particular, the classical methods would have some difficulties in finding six classes from the monodimensional histogram of Im 3 (Fig. 1).

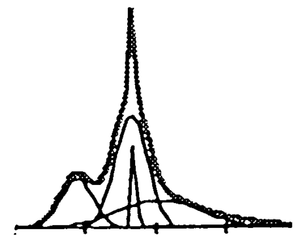


Fig. 4. Density mixture based on the parameters estimated from "La Rochelle" by SEM.

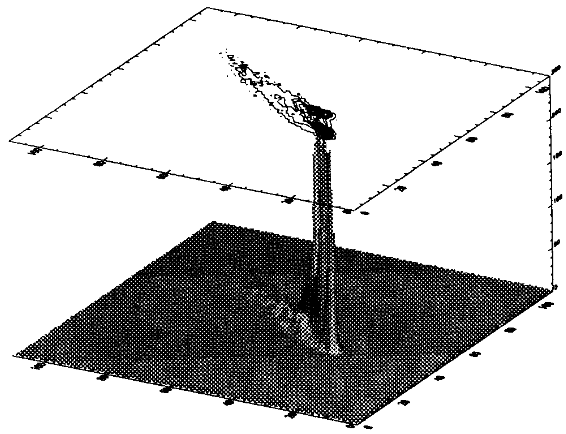


Fig. 5. Bidimensional histogram of "Guinee" taken from the two spectral bands Y^1, Y^2.

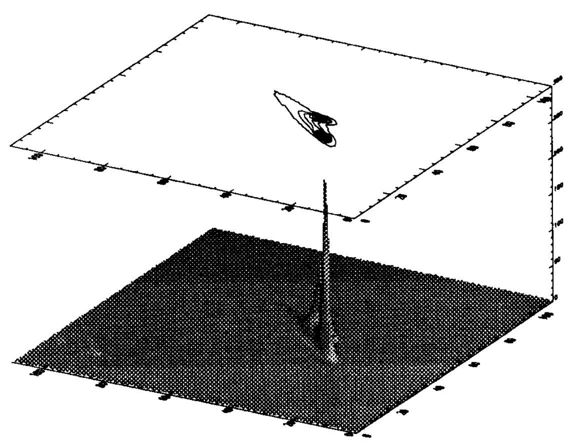


Fig. 6. The density of the distribution of (Y^1, Y^2) based on the parameters estimated from "Guinee" by SEM.

B. Unsupervised Segmentation

This is relatively difficult to connect with our simulation study; the large number of classes in the two images makes the conclusions of our simulation study, limited to binary images, uncertain. However, we can note that the spatial and spectral correlations are always significant. This would favor the use, in this context, of local methods.

1) Image Guinee Bissau: Im 3 below represents the data of Y^1 of the considered image and Koon(Im3), B(IM3), Cspa(IM3), Cspe(Im3), Cspaspe(Im3) the results of its segmentation by Koontz *et al.*'s algorithm [13] and the four SEM-based methods mentioned above.

We notice a real improvement of Cspa over Koon at the class detection level. First, the class 3, situated between the two peaks of the histogram, is ignored by Koon. Furthermore, Cspa allows the identification of two other classes, 4 and 5, which are very "standard deviation discriminating". In a general manner the methods based on histograms are ineffective in finding the "standard deviation discriminating" classes.

The comparison between B and Cspa is difficult. In fact, Cspa is more efficient at the class detection level (6 instead of 5) but B seems to better restore the fine structures of the ground.

Concerning the Cspe we notice that the spectral noise correlations are strong (Table XIV). Our simulation study showed that the spectral context contribution can be insignificant (case IV.A.1), or good (case IV.A.2), when all parameters are known. Here we are at the presence of the two cases and the first one seems to be predominant. In fact, the results obtained by Cspe are worse than those obtained by B or Cspa. This can also be due to the degradation of the parameter estimation efficiency. To be more precise, the SEM could turn out to be much more efficient in case B, where the number of parameters to estimate remains limited, than in case Cspe, where this number is important and the strong correlation of data limits the additional information. In conclusion, we have to opt for B or Cspa, according to which things we are interested in.

2) Image La Rochelle: Im 4 represents the data of Y^1 of the considered image and Koon(IM4), B(IM4), Cspa(Im4), Cspe(Im4), Cspaspe(Im4) the results of its segmentation as above.

We note that Koon does not distinguish the "sea" class from the "town" one, which are, according to the estimated parameter values, very "standard deviations discriminating". The blind SEM finds these two classes but the corresponding B method does not allow their detection in an unerring manner. This weakness is almost entirely deleted by the use of Cspa. This is in accordance with our simulation study. In fact, we have seen that the spatial context contribution can be significant in this case (IV.B.2). The quality of the segmentation performed by Cspaspe seems to be quite superior to the quality obtained by the three methods below: the distinction between the classes "sea" and "town" is quite clear. This can be explained by the following property, which we have not included in the simulation study. Let us denote by

Image "Guinee", spectral band Y^1.

Im3

Unsupervised segmentation of "Guinee" by Koontz et al. method

Koon(Im3)

Image "La Rochelle", spectral band Y^1.

Im4

Unsupervised segmentation of "La Rochelle" by Koontz et al. method

Koon(Im4)

Unsupervised segmentation of "Guinee" by the blind method. Parameters estimated by SEM.

B(Im3)

Unsupervised segmentation of "Guinee" by spatially contextual (one neighbor) method. Parameters estimated by SEM.

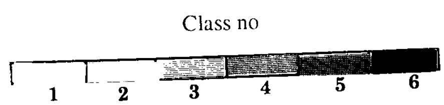
Cspa(Im3)

Unsupervised segmentation of "Guinee" by the spectrally contextual (two spectral bands) method. Parameters estimated by SEM.

Cspe(Im3)

Unsupervised segmentation of "Guinee" by spatially and spectrally contextual (one neighbour, two spectral bands) method. Parameters estimated by SEM.

Cspaspe(Im3)



a_1 and b_1 the means in two spectral bands corresponding to a class A and by a_2, b_2 those corresponding to a class B. When $b_2 - b_1 = a_2 - a_1$ the separability of the Guassian densities on R^2 decreases when the spectral correlation of the noise increases (Fig. 9), but when $b_2 - b_1 \neq a_2 - a_1$ this separability can strongly increase (Fig. 10).

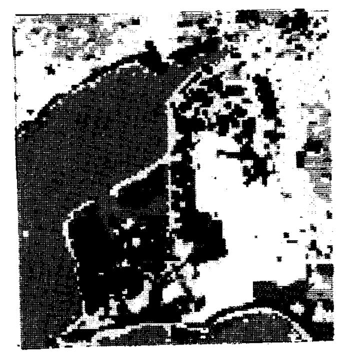
To be more specific, let us denote by $f_{\rho,1}, f_{\rho,2}$ the Gaussian densities corresponding to two classes with ρ the correlation coefficient. Let us put $M_1 = (a_1, b_1)$ and $M_2 = (a_2, b_2)$ as the means points. When $\rho \to 1$ the density $f_{\rho,1}$ becomes concentrated on a 45° line passing through M_1 and likewise for $f_{\rho,2}$ and M_2. This leads to two quite different cases. If the line segment M_1M_2 has unit slope, the two lines above are the same and, when $\rho \to 1$, the theoretical classification error tends to a finite value different from 0. When this segment does not have unit slope the two lines considered are different, and thus the densities $f_{\rho,1}, f_{\rho,2}$ become concentrated, when $\rho \to 1$, on two different subsets of R^2. This implies that the classification error tends to 0. In our simulation study we consider the first case, in fact $a_1 = b_1, a_2 = b_2$. This means that in real situations the spectral context contribution could turn out to be still more relevant.

We notice, through this example, that the conclusions of our simulation study are not immediately generalizable to an arbitrary number of classes, but can be quite useful when we are interested in detecting two given classes. In conclusion, we have to opt to Cspaspe in the case considered.

Unsupervised segmentation of "La Rochelle" by the blind method. Parameters estimated by SEM.
B(IM4)

Unsupervised segmentation of "La Rochelle" by spatially contextual (one neighbor) method. Parameters estimated by SEM.
Cspa(Im4)

Unsupervised segmentation of "La Rochelle" by the spectrally contextual (two spectral bands) method. Parameters estimated by SEM.
Cspe(Im4)



Unsupervised segmentation of "La Rochelle" by spatially and spectrally contextual (one neighbor, two spectral bands) method. Parameters estimated by SEM.
Cspaspe(Im4)

VII. Conclusion

The aim of this work was to study, in the framework of local Bayesian unsupervised satellite image segmentation, the relevance of the use of the spectral or spatial context contribution. We showed via a simulation study that the information carried by the context strongly varies with specific image parameters such as homogeneity, standard deviations, as well as spatial and spectral correlations of the noise. In particular, in numerous situations the efficiency of fast blind segmentation cannot be improved and it is useless to refer to any context. The same study shows that the VD case is more relevant, in any kind of situation, than the MD case. Furthermore, we proposed the use of a recent mixture estimation algorithm, SEM, in order to estimate the parameters above. Another simulation study, Section V, shows the correct behavior of the SEM in the context used. The parameter estimation step is of prime importance: it was shown that the efficiency of global methods like ICM or MPM can be seriously degraded when the estimated values of the parameters move away from the real ones [9]. This phenomenon does not appear when considering the unsupervised method based on the SEM and the contextual segmentation. Thus we can say, roughly speaking, that the properties brought to fore in Section IV are "saved" when switching to the unsupervised approaches based on the SEM. The interest of the results contained in these two sections goes beyond satellite image segmentation problems and can be applied in any problem of unsupervised Bayesian classification. In many detection problems occurring in medical imagery, industrial radiography, or infrared imagery, we know *a priori* that there are two classes. In such situations the automatic "adaptive" algorithm we proposed in Section V could turn out to be quite useful.

Our real image study, Section VI, shows that the SEM-based unsupervised methods are always more efficient than Koontz *et al.* 's histogram method. Furthermore, it appears, according to the visual evaluation, that the context contribution strongly

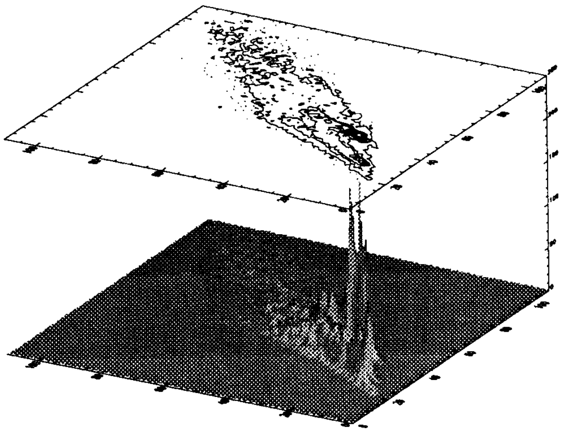


Fig. 7. Bidimensional histogram of "La Rochelle" taken from the two spectral bands Y^1, Y^2.

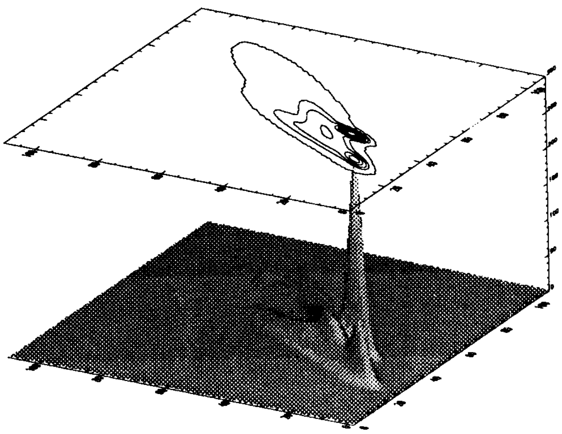


Fig. 8. The density of the distribution of (Y^1, Y^2) based on the parameters estimated from "La Rochelle" by SEM.

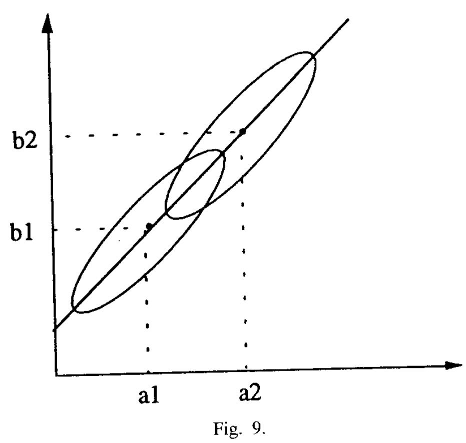


Fig. 9.

varies with the image parameters. The design of an "adaptive" algorithm as above is, without doubt, much more difficult. However, the simulation study of binary images can prove useful at two levels:

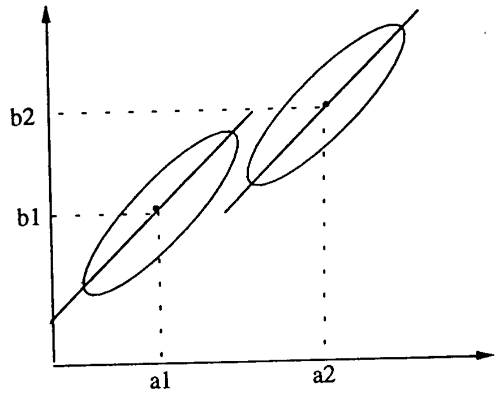


Fig. 10.

1) One can be particularly interested by some part of the image considered, say a 60×60 pixels array. The number of classes in such a window can be small, two or three for instance. The sample being rich enough, one can use the SEM locally and use the "adaptive" algorithm above in order to resegment the window considered.
2) As pointed out in Section VI, the results of the simulation study can be useful in understanding the behavior of the methods considered.

In a general manner, the SEM seems to be well adapted to satellite remote sensing problems. It was also used to estimate the depth and nature of the sea bottom and allowed the introduction of a reliability measure [19], [20]. The acceptable SEM based unsupervised SAR images segmentation [27] was recently improved by its "local" version [28]. Finally, we propose in [3] the use of a "fuzzy" SEM allowing a statistical fuzzy unsupervised segmentation.

As we mentioned above, the use of context is, in some situations, useless. This may be due to the fact that the useful complementary information remains beyond contexts reduced to four or eight neighbors, in which case it would be relevant to look at a global method.

REFERENCES

[1] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, pp. 259–302, 1986.

[2] B. Braathen, N. Marhic, P. Masson, and W. Pieczynski, "Sur une nouvelle approche de segmentation Bayesienne non supervisée d'images," *Proc. Treisième colloque GRETSI* (Juans-les-Pins, France), Sept. 1991.

[3] H. Caillol and W. Pieczynski, "Fuzzy statistical unsupervised Bayesian image segmentation," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS '92)*, (Houston, TX), May 1992.

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

[5] G. Celeux and J. Diebolt, "Comportement asymptotique d'un algorithme d'apprentissage probabiliste pour les mélanges de lois de probabilités," Tech. Rep. INRIA no. 563, 1986.

[6] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognition*, vol. 22, no. 6, pp. 747–761, 1989.

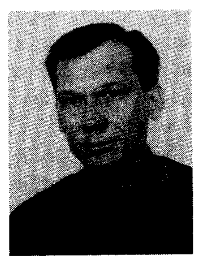
[7] M. M. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.

[8] P. A. Devijver and M. Dekesel, "Champs aléatoires de Pickard et modélisation d'images digitales, *Traitement du Signal*, vol. 5, no. 5, 1988.

[9] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *Journal of Applied Statistics,* vol. 16, no. 2, 1989.
[10] M. Emsalem, H. Caillol, P. Olivié, G. Carnat, and W. Pieczynski, "Fast unsupervised statistical image segmentation," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 92)* (Houston, TX), May 1992.
[11] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intel.,* vol. 6, pp. 721–741, 1984.
[12] R. Haralick and J. Hyonam, "A context classifier," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-24, pp. 997–1007, 1986.
[13] W. L. Koontz, P. M. Naranda, and K. Fukunaga, "A graph-theoretic approach to non parametric cluster analysis," *IEEE Trans.on Computers,*vol. C-25, pp. 936–944, 1976.
[14] P. A. Kelly, H. Derin, and K. D. Hart, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Trans. ASSP,* vol. 36, pp. 1628–1641, 1988.
[15] J. T. Kent and K. V. Mardia, "Spatial classification using fuzzy membership," *IEEE Trans. Pattern Anal. Machine Intel.,* vol. 10, 1991.
[16] N. Marhic, P. Masson, and W. Pieczynski, "Mélange de lois et segmentation non supervisée des donnees SPOT," *Statistique et Analyse des Donnees,* vol. 16, no. 2, 1991.
[17] N. Marhic and W. Pieczynski, "Estimation of mixture and unsupervised segmentation of images," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 91),* (Helsinki, Finland), May 1991.
[18] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of the American Statistical Association,* vol. 82, pp. 76–89, 1987.
[19] P. Masson, "Etude d'algorithmes de classification contextuelle et application à la segmentation d'images satellite," Thèse, Université de Bretagne Occidentale, Sept. 1991.
[20] P. Masson, G. Dubois, G. Le Lann, A. Hillion, and C. Roux, "Classification d'images du satellite SPOT pour la cartographie bathymétrique du littoral," *Proc. PIXILES 90 Journées internationales* (Nouméa, Tahiti), Nov. 1990.
[21] P. Masson and W. Pieczynski, "Segmentation of SPOT images by contextual SEM," *Proc. of EUSIPCO* (Barcelona, Spain), Nov. 1990.
[22] E. Mohn, N. Hjort, and G. Storvic, "A simulation study of some contextual classification methods for remotely sensed data," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-25, pp. 796–804, 1987.
[23] W. Pieczynski, "Estimation of context in random fields," *Journal of Applied Statistics,* vol. 16, no. 2, 1989.
[24] W. Pieczynski, "Mixture of distributions, Markov random fields and unsupervised Bayesian segmentation of images," Tech. Rep. no. 122, L.S.T.A. Université Paris VI, 1990.
[25] W. Pieczynski, "Statistical image segmentation," *Machine Graphics and Vision,* vol. 1, no. 1/2 (Proceedings of GKPO'92, Naleczow, Poland, May 1992).
[26] 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," *Journal of Applied Statistics,* vol. 16, no. 2, 1989.
[27] H. C. Quelle, J. M. Boucher, and W. Pieczynski, "Unsupervised Bayesian classification of SAR images," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 91)* (Helsinki, Finland), May 1991.
[28] H. C. Quelle, J. M. Boucher, and W. Pieczynski, "Local parameter estimation and unsupervised segmentation of SAR images," *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 92),* (Houston, TX), May 1992.
[29] P. Swain, S. Vardeman, and J. Tilton, "Contextual classification of multispectral image data," *Pattern Recognition,* vol. 13, pp. 429–441, 1981.
[30] J. Tilton, S. Vardeman, and P. Swain, "Estimation of context for statistical classification of multispectral image data," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-20, pp. 445–452, 1982.
[31] A. Veijanen, "A simulation-based estimator for hidden Markov random fields," *IEEE Trans. PAMI,* vol. 13, no. 8, 1991.
[32] L. Younes, "Parametric inference for imperfectly observed Gibbsian fields," *Probability Theory and Related Fields,* vol. 82, pp. 625–645, 1989.

Pascale Masson received the "Doctorat en Eléctronique" from the Université de Bretagne Occidentale, on September 1991.

Her research has been performed in the Groupe Traitement d'Images, Ecole Nationale Supérieure des Télécommunications de Brest, France Telecom University. Her research areas include statistical image processing, pattern recognition and computer vision.



Wojciech Pieczynski received the "Doctorat d'Etat en Mathématiques" from Université Paris 6, in February 1986.

He has been with France Telecom since 1987. He is presently head of the Image Group at the Institut National des Télécommunications d'Evry. His main research areas include stochastic modelling and statistical techniques of image processing.