Unsupervised Classification of Radar Images Based on Hidden Markov Models and Generalised Mixture Estimation

Roger Fjørtoft[*a], Jean-Marc Boucher[b], Yves Delignon[c], René Garello[b], Jean-Marc Le Caillec[b], Henri Maître[d], Jean-Marie Nicolas[d], Wojciech Pieczynski[e], Marc Sigelle[d], and Florence Tupin[d]

[a]Norwegian Computing Center, P. O. Box 114 Blindern, N-0314 Oslo, Norway
[b]ENST Bretagne, Technopôle de Brest Iroise, BP 832, F-29285 Brest cedex, France
[c]ENIC, Cité scientifique, rue Marconi, F-59658 Villeneuve d'Ascq cedex, France
[d]ENST, 46 rue Barrault, F-75634 Paris cedex 13, France
[e]INT, 9 rue Charles Fourier, F-91011 Evry cedex, France

ABSTRACT

Due to the enormous quantity of radar images acquired by satellites and through shuttle missions, there is an evident need for efficient automatic analysis tools. This article describes unsupervised classification of radar images in the framework of hidden Markov models and generalised mixture estimation. In particular, we show that hidden Markov chains, based on a Hilbert-Peano scan of the radar image, are a fast and efficient alternative to hidden Markov random fields for parameter estimation and unsupervised classification. We also describe how the distribution families and parameters of classes with homogeneous or textured radar reflectivity can be determined through generalised mixture estimation. Sample results obtained on real and simulated radar images are presented.

Keywords: generalised mixture estimation, unsupervised classification, hidden Markov models, radar images

1. INTRODUCTION

Both visual interpretation and automatic analysis of data from imaging radars are complicated by a fading effect called *speckle*, which manifests itself as a strong granularity in detected (amplitude or intensity) images. For example, simple classification methods based on thresholding of grey-levels are generally inefficient when applied to speckled images, due to the high degree of overlap between the distributions of the different classes. Speckle is caused by the constructive and destructive interferences between the backscattered waves from each resolution cell. It is generally modelled as a multiplicative random noise. At full resolution (single-look images), the standard deviation of the intensity is equal to the local mean reflectivity, corresponding to a signal to noise ratio of 0 dB. To overcome this problem, we must exploit spatial dependencies, e.g. the fact that, for most natural scenes, neighbouring pixels are more likely to belong to the same class than pixels that are farther away from each other.

Markov random fields are frequently used to impose a spatial regularity constraint on the classes and to allow a global Bayesian optimisation of the classification result, according to criteria such as the maximum a posteriori (MAP) or the maximum posterior marginal (MPM). However, the computing time is considerable and often prohibitive with this approach. A substantially quicker alternative is to use Markov chains, which can be adapted to two-dimensional analysis through a Hilbert-Peano scan of the image[1, 2, 3, 4].

In the case of unsupervised classification, the statistical properties of the different classes are unknown and must be estimated. For each of the Markov models cited above, we can estimate characteristic parameters with iterative methods such as estimation-maximisation (EM)[5], stochastic estimation-maximisation (SEM)[6], or iterative conditional estimation (ICE)[7, 8]. Classical mixture estimation consists in identifying the parameters of a set of Gaussian distributions corresponding to the different classes of the image. The weighted sum (or mixture) of the distributions of the different classes should approach the overall distribution of the image. In generalised mixture estimation we are not limited to Gaussian distributions, but to a finite set of distribution families[9, 4]. For each class we thus seek both the distribution family and the

[*]Correspondence: Roger Fjørtoft, E-mail: Roger.Fjortoft@nr.no, Phone: (+47) 22 85 26 98, Fax: (+47) 22 69 76 60, or Wojciech Pieczynski, E-mail: Wojciech.Pieczynski@int-evry.fr, Phone: (+33) 1 60 76 44 25, Fax: (+33) 1 60 76 44 33.

parameters that best describe its samples. We have chosen distribution families that are well adapted to single- or multi-look amplitude radar images and to classes with or without texture[10].

In this study, we limit ourselves to the ICE estimation method and the MPM classification criterion. When analysing an image, the only input entered by the user is the number of classes and the list of distribution families that are allowed in the generalised mixture. The estimation and classification schemes are described separately for the two Markov models. We compare their performances and show in particular that the Markov chain method can compete with the Markov random field method in terms of classification accuracy for multi-look images, while being much faster. The impact of the number of classes and the composition of distribution families is examined. Tests have been effectuated on both real and simulated synthetic aperture radar (SAR) data.

The article is organised as follows: In section 2 we introduce the hidden Markov random fields and hidden Markov chains models, as well as the different probabilities that will be needed for parameter estimation and classification. For simplicity, we first describe MPM classification in section 3, assuming known parameters for both Markov models. The ICE parameter estimation methods are presented in section 4, first in the framework of classic mixture estimation (only Gaussian distributions), and then for generalised mixture estimation (several possible distribution families). Specific adaptations to radar images are also mentioned. Estimation and classification results obtained on SAR images are reported in section 5. Our preliminary conclusions are given in section 6.

2. MODELS

Let S be a finite set corresponding to the N pixels of an image. We consider two random processes $\mathbf{Y} = (Y_s)_{s \in S}$ and $\mathbf{X} = (X_s)_{s \in S}$. $\mathbf{Y}$ represents the observed image and $\mathbf{X}$ the unknown class image.[#] Each random variable X_s takes its values from the finite set of classes $\Omega = \{\omega_1, \ldots, \omega_K\}$, whereas each Y_s is a real value. We denote realisations of $\mathbf{X}$ and $\mathbf{Y}$ by $\mathbf{x} = (x_s)_{s \in S}$ and $\mathbf{y} = (y_s)_{s \in S}$, respectively.

We here suppose that the random variables $\mathbf{Y} = (Y_s)_{s \in S}$ are independent conditionally on $\mathbf{X}$, and that the distribution of each Y_s conditional on $\mathbf{X}$ is equal to its distribution conditional on X_s.[§] Hence, all the distributions of $\mathbf{Y}$ conditional on $\mathbf{X}$ are determined by the K distributions of Y_s with respect to $X_s = \omega_1, \ldots, X_s = \omega_K$, which will be denoted $f_1, \ldots, f_K$:

$$P(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \prod_{s \in S} P(Y_s = y_s \mid X_s = x_s) = \prod_{s \in S} f_{x_s}(y_s) \tag{1}$$

$\mathbf{X}$ is in many cases well described by a Markov model. We generally refer to it as a hidden Markov model, as $\mathbf{X}$ is not directly observable. The classification problem consists in estimating $\mathbf{X} = \mathbf{x}$ from the observation $\mathbf{Y} = \mathbf{y}$.

2.1 Hidden Markov Random Fields

Let V_s denote a neighbourhood of the pixel s whose geometric shape is independent of $s \in S$. $\mathbf{X}$ is a Markov random field if and only if

$$P\left(X_s = x_s \middle| (X_t = x_t)_{t \in S, t \neq s}\right) = P\left(X_s = x_s \middle| (X_t = x_t)_{t \in V_s}\right), \tag{2}$$

i.e., if the probability that the pixel s belongs to a certain class x_s conditional on the classes attributed to the pixels in the rest of the image is equal to the probability of x_s conditional on the classes of the pixels in the neighbourhood V_s.

On certain conditions, which are generally verified in digital imagery, the Hammersley-Clifford theorem[11] establishes the equivalence between a Markov field, defined with respect to a neighbourhood structure V, and a Gibbs field whose potentials are associated with V. The elementary relationships within the neighbourhood V are given by the system of clicks C. Figure 1 shows the neighbourhood and the associated clicks in the case of 4-connectivity.

[#] As a practical example, consider a radar image $\mathbf{Y}$ covering an area with agricultural fields. We can imagine a corresponding class image $\mathbf{X}$, where the different crops are identified by discrete labels. Each observed pixel amplitude depends on several factors, including the characteristic mean radar reflectivity and texture of the underlying class, the speckle phenomenon, the transfer function of the imaging system, and thermal noise.

[§] In the case of radar images, this means that we suppose uncorrelated speckle.

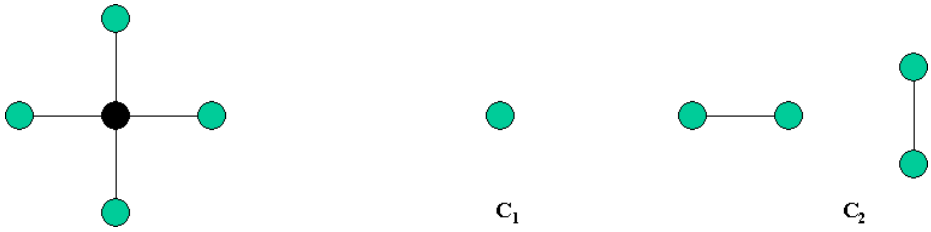


Figure 1 Pixel neighbourhood and associated click families C_1 and C_2 in the case of 4-connectivity.

For a Gibbs field

$$P(\mathbf{X}=\mathbf{x})=\frac{1}{Z}\exp\left(-U(\mathbf{x})\right), \tag{3}$$

where $U(\mathbf{x})=\sum_{c\in C}U_c(\mathbf{x})$ is the energy of $\mathbf{X}$ and $Z=\sum_{\mathbf{x}\in\Omega^N}\exp\left(-U(\mathbf{x})\right)$ is a normalising factor. The latter is in practise impossible to compute due to the very high number of possible configurations (K^N). The Hammersley-Clifford theorem makes it possible to relate local and global probabilities. Indeed, the local conditional probabilities can be written:

$$P\left(X_s=x_s\middle|(X_t=x_t)_{t\in V_s}\right)=\frac{1}{Z_s}\exp\left(-U_s\left(x_s,(x_t)_{t\in V_s}\right)\right), \tag{4}$$

where $U_s\left(x_s,(x_t)_{t\in V_s}\right)$ is the local energy function and $Z_s=\sum_{\omega_i\in\Omega}U_s\left(\omega_i,(x_t)_{t\in V_s}\right)$ is a normalising factor. There are several ways of computing U_s. For simplicity, we shall restrict ourselves to Potts model, 4-connectivity and clicks of type C_2, in which case $U_s\left(x_s,(x_t)_{t\in V_s}\right)$ is the number of pixels $t\in V_s$ for which $x_t\neq x_s$, minus the number of pixels $t\in V_s$ for which $x_t=x_s$, multiplied by a regularity parameter λ. The energy U in (3) is computed in a similar way, except that we sum the potentials over all vertical and horizontal C_2 clicks in the image. As will be explained later, it can be advantageous to have different regularity parameters λ_x and λ_y horizontally and vertically.

Bayes' rule and the conditional independence of the samples (1) allow us to write the *a posteriori* probability as

$$P(\mathbf{X}=\mathbf{x}\mid\mathbf{Y}=\mathbf{y})=\frac{P(\mathbf{Y}=\mathbf{y}\mid\mathbf{X}=\mathbf{x})P(\mathbf{X}=\mathbf{x})}{P(\mathbf{Y}=\mathbf{y})}=\frac{1}{Z'}\exp\left(-U(\mathbf{x})+\sum_{s\in S}\log f_{x_s}(y_s)\right) \tag{5}$$

and the corresponding local conditional distributions as

$$P\left(X_s=x_s\middle|Y_s=y_s,(X_t=x_t)_{t\in V_s}\right)=\frac{1}{Z_s'}\exp\left(-U_s\left(x_s,(x_t)_{t\in V_s}\right)+\log f_{x_s}(y_s)\right), \tag{6}$$

where Z' and Z_s' are normalizing factors obtained by summing over all possible nominators, similar to Z and Z_s. Substituting $U'(x)=U(\mathbf{x})-\sum_{s\in S}\log f_{x_s}(y_s)$ into (5), we see that $\mathbf{X}$ conditional on $\mathbf{Y}$ is a Gibbs field (3).

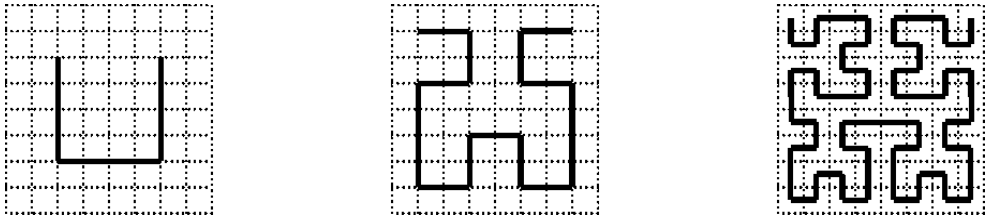
It is not possible to create *a posteriori* realisations of $\mathbf{X}$ according to (5) directly, but they can be approximated iteratively with the Metropolis algorithm[12] or Gibbs sampler[13]. We shall here only consider Gibbs sampler, which includes the following steps:

- We start from a random class image $\mathbf{x}^0$.
- The image is traversed repeatedly[£] until convergence. For each iteration q and for each pixel s,
  - the local *a posteriori* distribution given by (6) is computed, and
  - the pixel is attributed to a class drawn randomly according to this distribution.[$]

In a similar way, *a priori* realisations of $\mathbf{X}$ obeying (3) can be computed iteratively using (4).

[£]For each iteration, the pixels should ideally be visited in random order, but it is generally done line by line.

[$]Random sampling of a class according to a distribution can be done in the following way: We consider the interval $[0,1]$ and attribute to each class a subinterval whose width is equal to the probability of that class. A uniformly distributed random number in $[0,1]$ is generated, and the class is finally selected according to the subinterval in which this random number falls.



(a) (b) (c)

Figure 2 Construction of a Hilbert-Peano scan for an 8x8 image: (a) initialisation, (b) intermediate stage and (c) result.

2.2 Hidden Markov Chains

A 2D image can easily be transformed into a 1D chain, e.g. by traversing the image line by line or column by column. Another alternative is to use a Hilbert-Peano scan[1], as illustrated in Figure 2. Generalised Hilbert-Peano scans[2] can be applied to images whose length and width are not powers of 2. In a slightly different sense than above, let now $\mathbf{X} = (X_1, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, \ldots, Y_N)$ be the vectors of random variables ordered according to such a transformation of the class image and the observed image, respectively. Their realisations will consequently be denoted $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_N)$. According to the definition, $\mathbf{X}$ is a Markov chain if

$$P(X_n = x_n \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = P(X_n = x_n \mid X_{n-1} = x_{n-1}) \tag{7}$$

for $1 < n \leq N$. The distribution of $\mathbf{X}$ will consequently be determined by the distribution of X_1, denoted by π, and the set of transition matrices $(\mathbf{A}^n)_{1 \leq n < N}$, whose elements are $a_{ij}^n = P(X_{n+1} = \omega_j \mid X_n = \omega_i)$. We will in the following assume that the probabilities

$$c_{ij} = P(X_n = \omega_i, X_{n+1} = \omega_j) \tag{8}$$

are independent of n. The initial distribution then becomes

$$\pi_i = P(X_1 = \omega_i) = \sum_{1 \leq j \leq K} c_{ij} \tag{9}$$

and the stationary transition matrix $\mathbf{A}$ is given by

$$a_{ij} = P(X_{n+1} = \omega_j \mid X_n = \omega_i) = \frac{c_{ij}}{\sum_{1 \leq l \leq K} c_{il}}. \tag{10}$$

Hence, the *a priori* distribution of $\mathbf{X}$ is entirely determined by the K^2 parameters $(c_{ij})_{1 \leq i \leq K, 1 \leq j \leq K}$, and we can write

$$P(\mathbf{X} = \mathbf{x}) = P(X_1 = \omega_{i_1}, \ldots, X_N = \omega_{i_N}) = \pi_{i_1} a_{i_1 i_2} \cdots a_{i_{N-1} i_N}. \tag{11}$$

The so-called *forward* and *backward* probabilities

$$\alpha_n(i) = P(X_1 = \omega_i, Y_1 = y_1, \ldots, Y_n = y_n) \tag{12}$$

and

$$\beta_n(i) = P(Y_{n+1} = y_{n+1}, \ldots, Y_N = y_N \mid X_n = \omega_i) \tag{13}$$

can be calculated recursively. Unfortunately, the original forward-backward recursions derived from (12) and (13) are subject to serious numerical problems[15, 3]. Devijver et al.[15] have proposed to replace the joint probabilities by *a posteriori* probabilities:

$$\alpha_n(i) \approx P(X_1 = \omega_i \mid Y_1 = y_1, \ldots, Y_n = y_n) \tag{14}$$

$$\beta_n(i) \approx \frac{P(Y_{n+1} = y_{n+1}, \ldots, Y_N = y_N \mid X_n = \omega_i)}{P(Y_{n+1} = y_{n+1}, \ldots, Y_N = y_N \mid Y_1 = y_1, \ldots, Y_n = y_n)} \tag{15}$$

In the following we use the numerically stable forward-backward recursions resulting from this approximation:

- Initialisation: $\alpha_1(i) = \dfrac{\pi_i f_i(y_1)}{\sum_{j=1}^{K} \pi_j f_j(y_1)}$ for $1 \le i \le K$ (16)

- Induction: $\alpha_{n+1}(i) = \dfrac{f_i(y_{n+1}) \sum_{j=1}^{K} \alpha_n(j) a_{ji}}{\sum_{l=1}^{K} f_l(y_{n+1}) \sum_{j=1}^{K} \alpha_n(j) a_{jl}}$ for $1 \le i \le K$, $1 \le n < N$ (17)

- Initialisation: $\beta_N(i) = 1$ for $1 \le i \le K$ (18)

- Induction: $\beta_n(i) = \dfrac{\sum_{j=1}^{K} a_{ij} f_j(y_{n+1}) \beta_{n+1}(j)}{\sum_{l=1}^{K} f_l(y_{n+1}) \sum_{j=1}^{K} \alpha_n(j) a_{jl}}$ for $1 \le i \le K$, $1 \le n < N-1$ (19)

The joint probability of two subsequent classes given all the observations

$$\Psi_n(i,j) = P(X_n = \omega_i, X_{n+1} = \omega_j \mid \mathbf{Y} = \mathbf{y}) \tag{20}$$

can be written as a function of the forward-backward probabilities:

$$\Psi_n(i,j) = \frac{\alpha_n(i) a_{ij} f_j(y_{n+1}) \beta_{n+1}(j)}{\sum_{l=1}^{K} f_l(y_{n+1}) \sum_{j=1}^{K} \alpha_n(j) a_{jl}} = \frac{\alpha_n(i) a_{ij} f_j(y_{n+1}) \beta_{n+1}(j)}{\sum_{l=1}^{K} \sum_{m=1}^{K} \alpha_n(l) a_{lm} f_m(y_{n+1}) \beta_{n+1}(m)} \tag{21}$$

The marginal *a posteriori* probability, i.e., the probability of having class ω_i in element number i given all the observations $\mathbf{Y}$, can also be expressed in terms of the forward-backward probabilities:

$$\xi_n(i) = P(X_n = \omega_i \mid \mathbf{Y} = \mathbf{y}) = \sum_{j=1}^{K} \Psi_n(i,j) = \frac{\alpha_n(i) \beta_n(i)}{\sum_{l=1}^{K} \alpha_n(l) \beta_n(l)} \tag{22}$$

It can be shown that the *a posteriori* distribution of $\mathbf{X}$, i.e., $p(\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y})$, is that of a non stationary Markov chain, with transition matrix

$$t_{ij}^{n+1} = P(X_{n+1} = \omega_j \mid X_n = \omega_i, \mathbf{Y} = \mathbf{y}) = \frac{a_{ij} f_j(y_{n+1}) \beta_{n+1}(j)}{\sum_{l=1}^{K} a_{il} f_l(y_{n+1}) \beta_{n+1}(l)}. \tag{23}$$

We can therefore simulate *a posteriori* realisations of $\mathbf{X}$ directly, i.e. without iterative procedures as in the case of hidden Markov fields, using (23) recursively. The class of the first pixel is drawn randomly according to the marginal *a posteriori* distribution $\xi_1(i)$ (22). Subsequently, for each new pixel, the transition probabilities (23) are computed, the class of the precedent pixel ω_i being fixed, and the class ω_j is obtained by random sampling according to this distribution.

3. CLASSIFICATION

Let us first assume that we know the distribution f_i and the associated parameters Θ_i of each class ω_i, as well as the regularity parameters of the underlying Markov model (λ or t_{ij}). In a Bayesian framework, the goal of the classification is to determine the realisation $\mathbf{X} = \mathbf{x}$ that "best" explains the observation $\mathbf{Y} = \mathbf{y}$, in the sense that it minimises a certain cost function. Several cost functions can be envisaged, leading to different estimators, such as the MAP, which aims at maximising the global *a posteriori* probability $P(\mathbf{X} \mid \mathbf{Y})$, and the MPM, which consists in maximising the posterior marginal distribution $P(X_s \mid \mathbf{Y})$ for each pixel. We here only consider MPM classification, as the computing time for the MAP solution is much higher, in particular when simulated annealing[13] is used for hidden Markov random fields.

3.1 Hidden Markov Random Fields

In the case of hidden Markov random fields, the MPM[14] is computed as follows:

- A series of independent *a posteriori* realisations of $\mathbf{X}$ are computed iteratively, using Gibbs sampler (or the Metropolis algorithm) based on the local *a posteriori* probability function (6).
- For each pixel we retain the class that has occurred the most frequently there.

The required number of *a posteriori* realisations and iterations per realisation will be discussed in section 5.

3.2 Hidden Markov Chains

The MPM solution can be calculated directly for hidden Markov chains, based on one forward-backward computation:

- For every element n in the chain, and for every possible class ω_i, we compute
  - the forward probabilities $\alpha_n(i)$ (17),
  - the backward probabilities $\beta_n(i)$ (19) and
  - the marginal *a posteriori* probabilities $\xi_n(i)$ (22).
- Each pixel n is attributed to the class ω_{i_n} that maximises ξ_n.

4. MIXTURE ESTIMATION

In practise, the regularity parameters and the parameters of the distributions of the classes are often unknown and must be estimated from the observation $\mathbf{Y} = \mathbf{y}$. The problem is then double: we do not know the characteristics of the classes, and we do not know which pixels are representative for each class. We first present *classic mixture estimation*, where all classes are supposed to be Gaussian, and then introduce *generalised mixture estimation*, where several distribution families are possible for each class. There are several iterative methods for mixture estimation, including EM[5], SEM[6] and ICE[7, 8]. We will only consider the latter here. Moreover, we will only present what is necessary for the implementation, and not the underlying theory.

4.1 Hidden Markov Random Fields

The ICE algorithm iteratively creates *a posteriori* realisations and recalculates the class and regularity parameters. In the framework of classic mixture estimation, the following computation steps are carried out for each iteration q:

- A certain number of *a posteriori* realisations (with index r) are computed with Gibbs sampler, using (6) and the parameters Θ_i^{q-1} (defining f_i^{q-1}) obtained in the previous iteration.
- The class parameters $\Theta_i^{q,r}$ (which are the mean value and the variance for a scalar Gaussian distribution) are estimated for each realisation and then averaged to obtain Θ_i^q.
- We have chosen to estimate the regularity parameter λ^q in each ICE iteration with a stochastic gradient approach, realised through a series of *a priori* realisations computed with Gibbs sampler using (4).

In order to reduce the computation time, we generally compute only one *a posteriori* realisation for each iteration of the ICE and only one *a priori* realisation for each iteration of the stochastic gradient. This simplification does not imply any significant performance loss.

We use a coarse approximation of the stochastic gradient equation. Let $U^q_{a\ posteriori}$ be the energy of the current *a posteriori* realisation and $U^{q,r}_{a\ priori}$ the energy of the *a priori* realisation in iteration number r of the stochastic gradient. Setting $\lambda^{q,0}$ to the value obtained in the precedent ICE iteration λ^{q-1}, we repeatedly compute until convergence:

$$\lambda^{q,r} = \lambda^{q,r-1} + \frac{1}{r}\frac{U^{q,r}_{a\ priori} - U^q_{a\ posteriori}}{U^q_{a\ posteriori}} \tag{24}$$

The ICE algorithm needs an initial class image from which the initial parameters of the classes are computed. We have used the K-means algorithm, which subdivides the grey-levels in K distinct classes iteratively. The initial class centres are

uniformly distributed over the range of grey-levels. The image is traversed repeatedly until stability, attributing each pixel to the class with the closest centre, and recomputing the class centres from all the attributed samples at the end of each iteration. It should be noted that this method only can be used to initiate classes with different mean values. It is, for example, not suited when classes with the same mean values but different textures must be distinguished. K-means is basically a thresholding method, so if there is much overlap between the true class distributions, as for radar images, the resulting class image will be quite irregular and the initial class statistics will not be very representative. The consequences of this, and a possible remedy, will be commented on in section 5. The initial value of λ is predefined (typically $\lambda = 0.5$).

4.2 Hidden Markov Chains

We initiate the ICE algorithm in a similar way for hidden Markov chains, using K-means to define the class parameters Θ_i^0 and thus the marginal conditional distributions f_i^0, uniformly distributed *a priori* probabilities π_i^0 and a generic transition matrix $\mathbf{A}^0 = \{a_{ij}^0\}$ where $a_{ij}^0 = 0.5$ when $i = j$ and $a_{ij}^0 = 0.5/(K-1)$ when $i \neq j$. Each ICE iteration q is based on one forward-backward computation and includes the following steps:

- For every element n in the chain, and for every possible class ω_i, we compute
  - the forward probabilities $\alpha_n^q(i)$ (17) based on a_{ij}^{q-1}, π_i^{q-1} and f_i^{q-1} (i.e. Θ_i^{q-1}),
  - the backward probabilities $\beta_n^q(i)$ (19) based on a_{ij}^{q-1}, π_i^{q-1} and f_i^{q-1} (i.e. Θ_i^{q-1}), and
  - the marginal *a posteriori* probabilities $\xi_n^q(i)$ (22).
- This allows us to compute
  - the new joint conditional probabilities $\Psi_n^q(i,j)$ (21),
  - the new elements of the stationary transition matrix

$$a_{ij}^q = \frac{\sum_{1 \le n \le N} \Psi_n^q(i,j)}{\sum_{1 \le n \le N} \xi_n^q(i)} \tag{25}$$

  - and the new initial probabilities

$$\pi_i^q = \frac{1}{N} \sum_{1 \le n \le N} \xi_n^q(i) . \tag{26}$$

- We compute a series of *a posteriori* realisations based on f_i^{q-1} and (23), as explained in section 2.2. For each realisation (with index r) we estimate the class parameters $\Theta_i^{q,r}$, which are averaged to obtain Θ_i^q and thus f_i^q.

As for hidden Markov random fields, we generally limit the number of *a posteriori* realisations per ICE iteration to one.

4.3 Generalised Mixture Estimation

In generalised mixture estimation, the distribution f_i of each class is not necessarily Gaussian, but can belong to any distribution family in a predefined set. This implies the following modifications to the above ICE algorithms:

- The parameters of all possible distribution families are computed from the a posteriori realisations for each class.
- The Kolmogorov-Smirnov test is used to determine the most appropriate distribution family for each class[4]:
  - We compute the cumulative distributions for all the distribution families, based on the estimated parameters of the class.
  - The cumulative normalised histogram of the class is computed.
  - We retain the distribution family having the smallest maximum difference between the cumulative distribution and the cumulative normalised histogram of the class.

4.4 Application to Radar Images

For simplicity, we here consider only two distribution families, corresponding to homogeneous and textured classes, respectively. Assuming the speckle to be spatially uncorrelated, the observed intensity in a zone of constant reflectivity is Gamma distributed. The corresponding amplitude distribution is

$$f_i(y) = \frac{2y}{\Gamma(L)}\left(\frac{L}{R}\right)^L y^{2(L-1)} e^{-y^2 L/R} \tag{27}$$

for $y \geq 0$ and $f_i(y) = 0$ for $y < 0$. L is here the equivalent number of independent looks of the image, which should be provided by the data supplier. Let the estimated moments of order m be denoted by μ_m. The estimated mean radar reflectivity $R = \mu_2$ in (27) is computed over the pixels attributed to class i.

If we assume that the radar reflectivity texture of a class is Gamma distributed, the observed intensity will obey a K distribution. The corresponding amplitude distribution is

$$f_i(y) = \frac{2b}{\Gamma(L)\Gamma(a)}\left(\frac{by}{2}\right)^{a+L-1} \mathrm{K}_{a-L}(by) \tag{28}$$

for $y \geq 0$ and $f_i(y) = 0$ for $y < 0$. K is here the modified Bessel function of the second kind, and the estimated parameters a and b are computed as follows: Let $C_1 = \sqrt{L}\,\Gamma(L)\mu_1 / (\sqrt{\mu_2}\,\Gamma(L+0.5))$ and $C_2 = L\mu_4 / ((L+1)\mu_2^2)$. If $C_1 < 1$, a is obtained by solving the equation $C_1\sqrt{a}\,\Gamma(a) - \Gamma(a+0.5) = 0$. Otherwise $(C_1 \geq 1)$ we set $a = 1/(C_2 - 1)$, provided that $C_2 > 1$. In both cases $b = 2\sqrt{La/\mu_2}$. If $C_1 \geq 1$ and $C_2 \leq 1$, we consider (28) as unsuited and we make sure that it is not selected. Moreover, if the texture is very weak (typically $a > 20$), we approximate (28) by (27).

For simplicity, we will in the following refer to (27) and (28) as Gamma and K distributions, respectively, even though they actually are the amplitude distributions corresponding to Gamma and K distributed intensities.

5. RESULTS

5.1 Real SAR Image

Figure 3 (a) shows an extract of a JERS 3-look amplitude image of a tropical forest with some burnt plots related to agricultural activity. The equivalent number of independent looks $L = 2.7$. It was decided to classify the image into three classes. Figure 3 (b) shows the initial classification obtained with the K-means algorithm. The classes are represented by different grey-levels.[€] We note that the K-means class image is very irregular due to the speckle. A more regular initial classification could be obtained by applying an adaptive speckle filter prior to the K-means algorithm. Our tests indicate that this generally has little influence on the final result, but it can sometimes facilitate a correct choice of distribution family.

Let us first consider the ICE and MPM algorithms based on the hidden Markov random fields model. We use 30 iterations for the ICE algorithm, with only one *a posteriori* realisation per iteration. The initial value of the regularity parameter is $\lambda = 0.5$. Within each ICE iteration, the maximum number of iterations for the stochastic gradient is set to 10, with one *a priori* realisation per iteration. We interrupt the iteration earlier if λ differs less than 0.01 from its previous value. Except for the first ICE iterations, the stochastic gradient estimation generally requires very few iterations. Gibbs sampler with as much as 100 iterations is used to generate the *a priori* and *a posteriori* realisations. The convergence of the global energies is in fact quite slow, especially for realisations according to the *a priori* distribution. The MPM classification based on the hidden Markov random fields model relies on 10 *a posteriori* realisations. Figure 3 (c) presents the obtained classification result. The regularity parameter was in this case estimated to $\lambda = 0.75$ and the Gamma distribution was retained for all three classes. The classification result corresponds quite well to our thematic conception of the image based on visual inspection, except that many fine structures seems to be lost and the region borders are somewhat rounded.

The number of ICE iterations is set to 30 also for the corresponding analysis scheme based on hidden Markov chains, and we compute only one *a posteriori* realisation per iteration. Figure 3 (d) shows the result of the MPM classification. It is generally less regular than the classification in Figure 3 (c), but it represents small structures more precisely. The K-distribution was here selected for the darkest class, whereas the Gamma distribution gave the best fit for the two others.

The overall quality is comparable for the two classification results, but the computing time is quite different: The programme based on hidden Markov random fields spent about 53 minutes on a PC with a Pentium III 733 MHz processor running Linux, whereas the programme based on hidden Markov chains only needed 2 minutes and 27 seconds.

[€] The grey-levels do not correspond to the mean amplitudes of the different classes, but they are ordered according to these mean values.

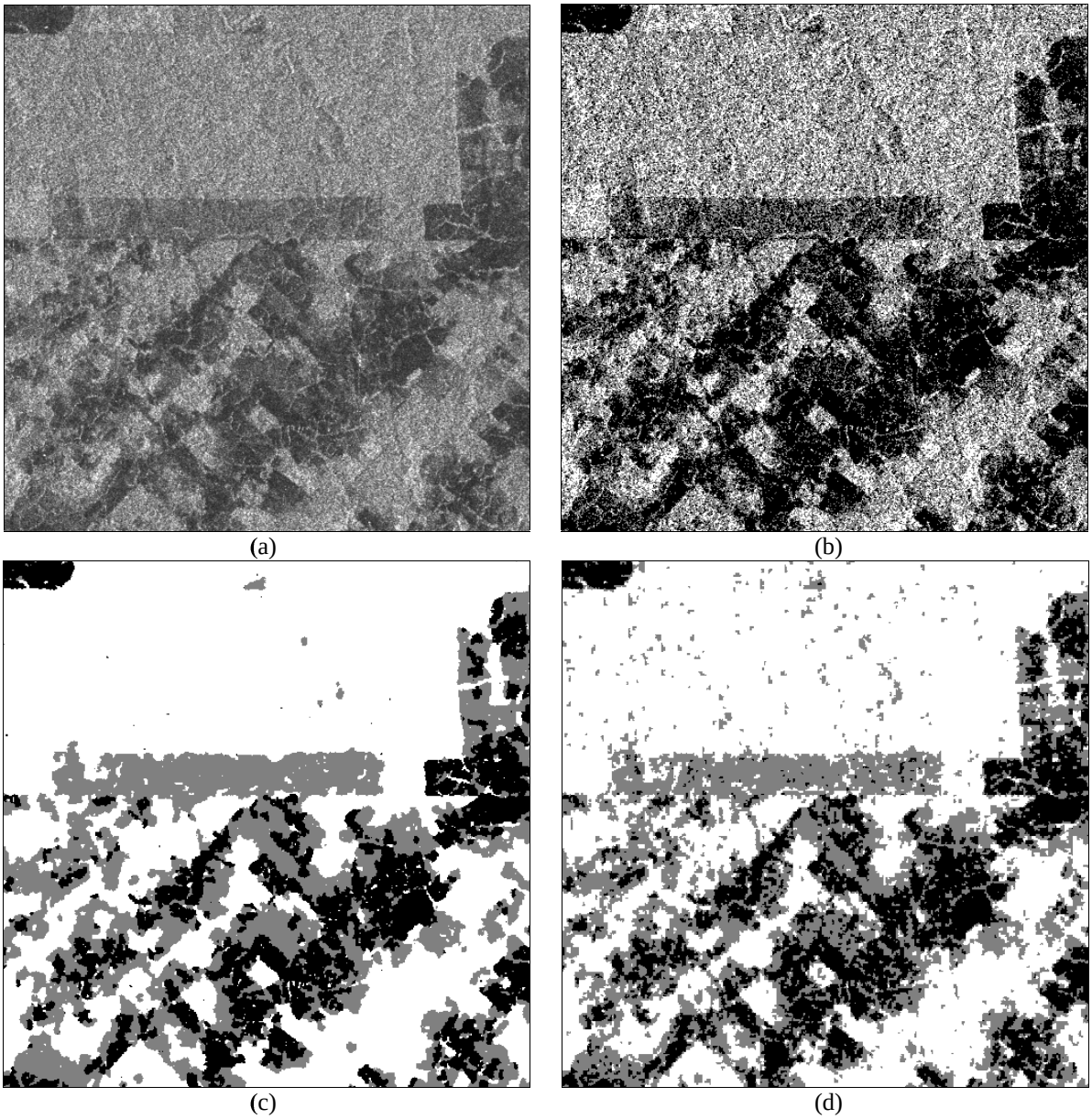


Figure 3 Classification of a JERS 3-look image of a tropical forest into 3 classes: (a) Original amplitude image. (b) Initial K-means classification. Result obtained (c) with the Markov random fields method and (d) with the Markov chains method.

5.2 Simulated SAR Images

In order to examine the performances more carefully, we have created two simulated SAR images. Figure 4 (a) represents an ideal class image with 3 classes, and Figure 4 (b) shows its 3-look ($L = 3.0$) speckled counterpart. The darkest and brightest classes are Gamma distributed, whereas the one in the middle is K distributed with texture parameter $a = 4.0$ (28). The contrast between two consecutive classes is 3.5 dB. The parameter settings described in section 5.1 were applied, except that we allowed different regularity parameters vertically and horizontally for the method based on hidden Markov random fields, as the resolution is not the same in the two directions. The regularity parameters were estimated to $\lambda_x = 0.37$ and $\lambda_y = 0.68$, respectively. However, the classification result in Figure 4 (c) seems far too irregular, and only 72.7 % of the pixels are correctly classified. The method based on hidden Markov chains here gives a more satisfactory result, shown in Figure 4 (d), with 83.9 % of correctly classified pixels. The confusion matrices are given in Tables 1 and 2.

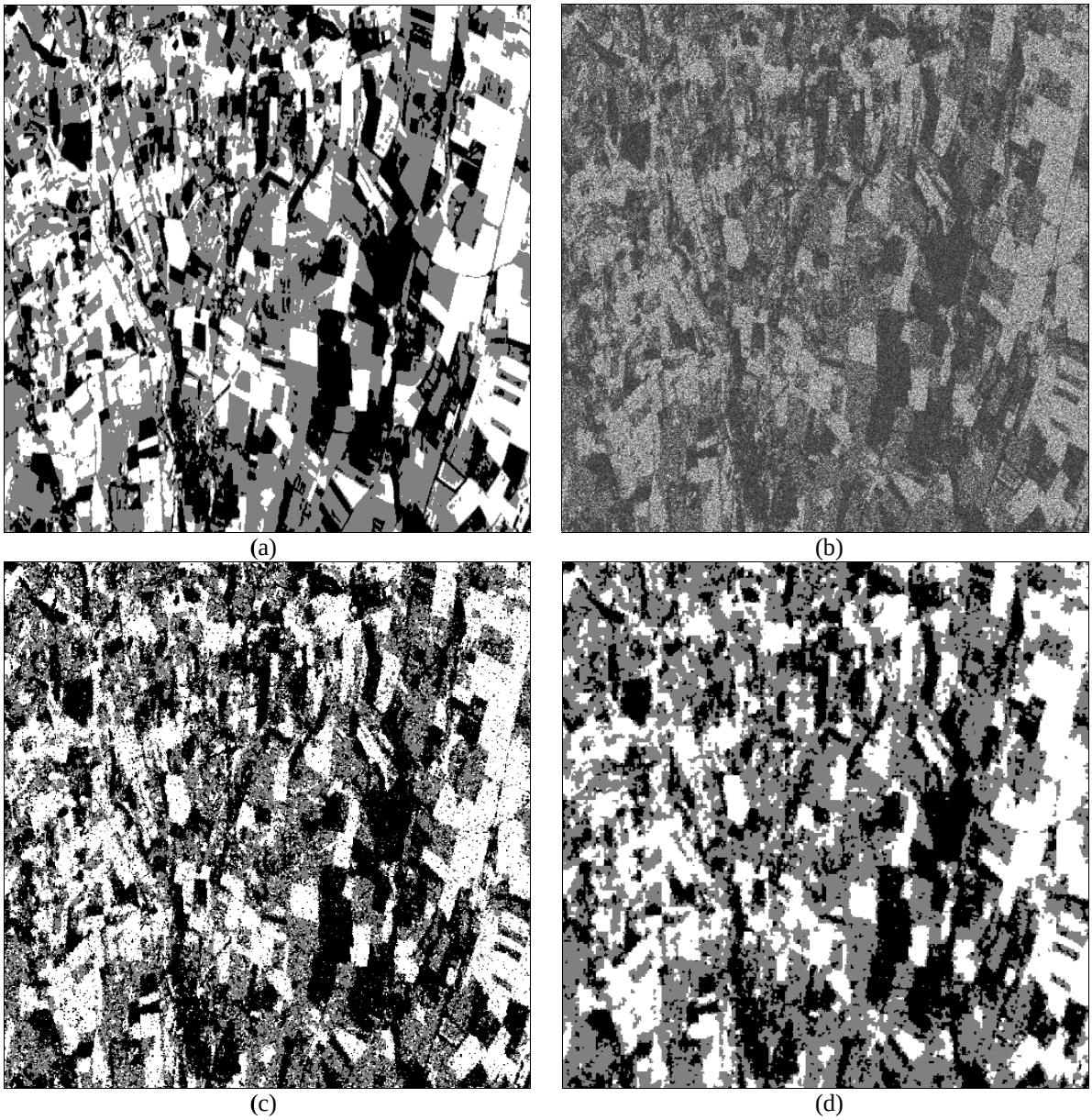


Figure 4 Classification of a simulated 3-look SAR image into 3 classes: (a) Ideal class image. (b) Speckled amplitude image. Results obtained (c) with the Markov random fields method and (d) with the Markov chains method.

0.954646	0.042079	0.003274
0.360625	0.462592	0.176782
0.030055	0.062420	0.907525

Table 1 Confusion matrix for the classification result in Figure 4 (c) obtained with the Markov random fields method.

0.850886	0.147929	0.001185
0.107937	0.831746	0.060317
0.000606	0.158517	0.840878

Table 2 Confusion matrix for the classification result in Figure 4 (d) obtained with the Markov chains method.

Both methods correctly identify Gamma distributions for the darkest and brightest classes, but only the hidden Markov chains method found that the class in the middle was K distributed. This method was here also about 27 times quicker.

Figure 5 (a) represents an ideal and approximately isotropic class image with 4 classes, and Figure 5 (b) shows the corresponding speckled image. The radiometric characteristics of the classes are the same as for the simulated image in Figure 4 (b), except that an additional Gamma distributed class with higher reflectivity has been added. Figure 5 (c) and (d) represent the results obtained with the two approaches, and Tables 3 and 4 show the corresponding confusion matrices.

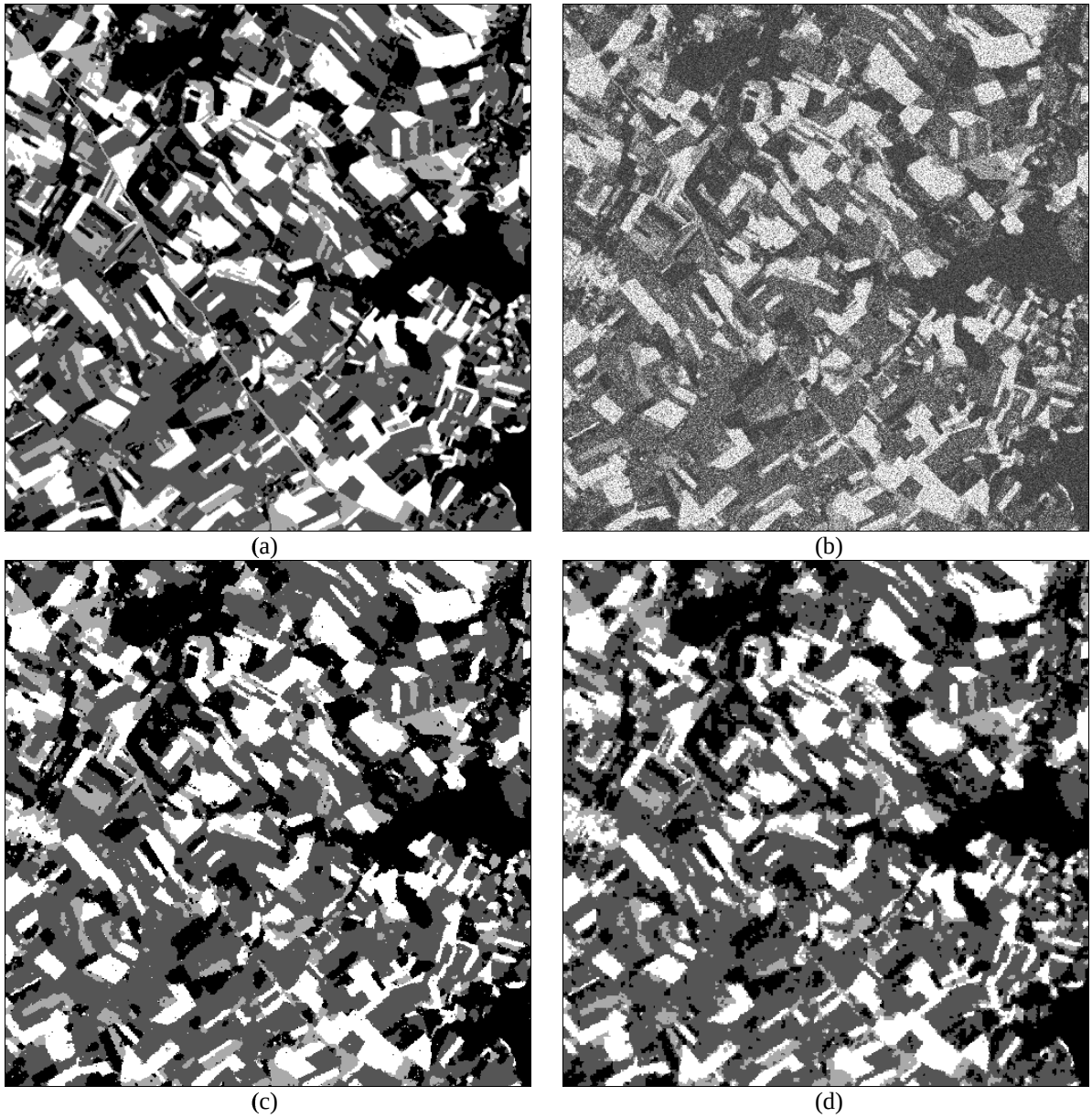


Figure 5 Classification of a simulated 3-look SAR image into 4 classes: (a) Ideal class image. (b) Speckled amplitude image. Results obtained (c) with the Markov random fields method and (d) with the Markov chains method.

0.923290	0.075753	0.000925	0.000031
0.089819	0.862768	0.039894	0.007519
0.017030	0.171414	0.642999	0.168557
0.001991	0.005545	0.021842	0.970622

Table 3 Confusion matrix for the classification result in Figure 5 (c) obtained with the Markov random fields method.

0.819639	0.178788	0.001265	0.000308
0.048103	0.909960	0.036221	0.005716
0.001741	0.244348	0.627520	0.126391
0.000018	0.011179	0.060639	0.928164

Table 4 Confusion matrix for the classification result in Figure 5 (d) obtained with the Markov chains method.

The regularity parameter for the method based on hidden Markov random fields is here $\lambda = 0.65$, which visually gives a very satisfactory result. The proportion of correctly classified pixels is 87.0 %, whereas it is 85.2 % for the method based on hidden Markov chains. The borders are slightly more irregular for the latter method, but some of the narrow structures are better preserved. The distribution families of the four classes were correctly identified by both methods, and the fit of the estimated parameters is comparable. The hidden Markov chains algorithm was, however, 25 times faster.

6. CONCLUSION

This article describes unsupervised classification of radar images in the framework of hidden Markov models and generalised mixture estimation. Hidden Markov random fields are frequently used to impose spatial regularity constraints in the parameter estimation and classification stages. This approach produces excellent results in many cases, but our experiments indicate that the estimation of the regularity parameter is a delicate problem, especially when the signal to noise ratio is low. The considerable computing time is another drawback. Methods based on hidden Markov chains, applied to a Hilbert-Peano scan of the image, constitute an interesting alternative. The estimation of the regularity parameters, which here are the elements of a stationary transition matrix, seems to be much more robust. The region borders often get slightly irregular with this approach, but fine structures are generally better preserved than for the corresponding scheme based on hidden Markov random fields. We describe how the distribution families and parameters of classes with homogeneous or textured radar reflectivity can be determined through generalised mixture estimation. For simplicity, we have restricted ourselves to Gamma and K distributed intensities. In our tests on simulated SAR images, the estimations of distribution families and associated parameters are of similar quality for the two approaches. Besides the robust estimation of regularity parameters, the computing speed is the main advantage of the hidden Markov chains approach. It was about 25 times faster than the programme based on the hidden Markov random fields in our experiments.

ACKNOWLEDGMENTS

This work was effectuated in the framework of the project “Segmentation d’Images Radar” financed by the Groupe des Ecoles des Télécommunications (GET). We thank the GET for the scientific and financial support.

REFERENCES

1. K. Abend, T. J. Harley, and L. N. Kanal, “Classification of binary random patterns”, *IEEE Transactions on Information Theory,* Vol. IT-11, No. 4, 1965.
2. W. Skarbek, “Generalized Hilbert scan in image printing”, *Theoretical Foundations of Computer Vision*, R. Klette and W. G. Kropetsh, editors, pp.45-57, Akademik Verlag, 1992.
3. B. Benmiloud and W. Pieczynski, “Estimation de parametrès dans les chaînes de Markov cachées et segmentation d’images”, *Traitement du Signal*, Vol. 12, No. 5, pp. 433-454, 1995.
4. N. Giordana and W. Pieczynski, “Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 465-475, 1997.
5. L. E. Baum, T. Petrie, G. Soules, and N. Weiss, “A maximization technique occuring in the statstical analysis of probabilistic functions of Markov chains”, *Ann, Math. Statistic.*, Vol. 41, pp. 164-171, 1970.
6. G. Celeux and J. Diebolt “L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités”, *Revue de Statistique Appliquée*, Vol. 34, No. 2, 1986.
7. W. Pieczynski, “Statistical image segmentation”, *Machine Graphics and Vision*, Vol. 1, No. 1/2, pp. 261-268, 1992.
8. W. Pieczynski, “Champs de Markov cachés et estimation conditionnelle iterative”, *Traitement du Signal*, Vol. 11, No. 2, pp. 141-153, 1994.
9. Y. Delignon, A. Marzouki and W. Pieczynski, “Estimation of generalised mixtures and its application in image segmentation”, *IEEE Transactions on Image Processing*, Vol. 6, No. 10, pp. 1364-1375, 1997.
10. Y. Delignon, R. Garello, and A. Hillion, “Statistical modelling of ocean SAR images”, *IEE Radar Sonar and Navigation*, Vol. 144, No. 6, pp. 348-354, 1997.
11. J. Besag, “Spatial interaction and the statistical analysis of lattice systems”, *Journal of the Royal Statistical Society (Series B),* Vol. 36, pp. 192-326, 1974.
12. N. Metropolis, A. W. Rosenbluth, A. H. Teller, M. R. Rosenbluth, and E. Teller, “Equations of state calculations by fast computing machines”, *Journal of Chemical Physics*, Vol. 21, pp. 1087-1091, 1953.
13. S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, pp. 721-741, 1984.
14. 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.
15. P. A. Devijver and M. Dekesel, “Champs aléatoires de Pickard et modélisation d'images digitales”, *Traitement du Signal*, Vol. 5, No. 5, pp. 131-150, 1988.