GENERALISED MIXTURE ESTIMATION AND UNSUPERVISED CLASSIFICATION BASED ON HIDDEN MARKOV CHAINS AND HIDDEN MARKOV RANDOM FIELDS

Roger Fjørtoft, Wojciech Pieczynski and Yves Delignon

Roger.Fjortoft@nr.no, Norwegian Computing Center, P. O. Box 114 Blindern, N-0314 Oslo, Norway
Wojciech.Pieczynski@int-evry.fr, INT, 9 rue Charles Fourier, F-91011 Evry cedex, France
Yves.Delignon@enic.fr, ENIC, rue Marconi, F-59658 Villeneuve d'Ascq cedex, France

ABSTRACT

Hidden Markov chain (HMC) models, applied to a Hilbert-Peano scan of the image, constitute a fast and robust alternative to hidden Markov random field (HMRF) models for spatial regularisation of image analysis problems, even though the latter provide a finer and more intuitive modelling of spatial relationships. In the framework of generalised mixture estimation and unsupervised classification of noisy images, we here show that the two approaches can be combined in a way that conserves their respective advantages. Sample results obtained on simulated radar images are presented.

1. INTRODUCTION

Hidden Markov random field (HMRF) models are often used to impose spatial regularity constraints on the underlying classes of an observed image and to allow Bayesian optimisation of the classification, according to criteria such as the maximum *a posteriori* (MAP) or the maximum posterior marginal (MPM). However, the computing time is often prohibitive with this approach. A substantially quicker alternative is to use a hidden Markov chain (HMC) model, which can be adapted to two-dimensional analysis through a Hilbert-Peano scan of the image [1, 2, 3, 4].

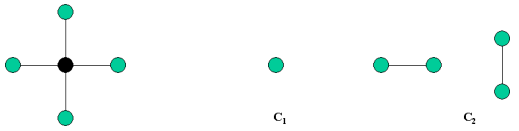
In the case of unsupervised classification, the statistical properties of the different classes are unknown and must be estimated. For each of the Markov models cited above, we can estimate characteristic parameters with iterative methods such as estimation-maximisation (EM) [5], stochastic estimation-maximisation (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. 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 correct distribution family and the parameters that best describe its samples.

In a recent study [10], we compared analysis schemes based on HMRF and HMC models on simulated synthetic aperture radar (SAR) images. We limited ourselves to the ICE estimation method and the MPM classification criterion. It was shown that the HMC method can compete with the HMRF method in terms of classification accuracy, while being much faster. The latter method, however, generally produced more regular region borders and a more visually pleasing result, but the estimation of the regularity parameters often failed. The idea behind the current article is therefore to combine the speed and robustness of the estimation and classification methods based on HMCs with the fine spatial modelling of HMRFs. For the analysis schemes described in this article, the only image-dependent parameter that must be entered by the user is the number of classes. In addition the list of distribution families that are allowed in the generalised mixture must be adapted to the image type. All other parameters are computed automatically.

The article is organised as follows: In section 2 the HMRF and HMC models are briefly introduced. Section 3 gives an overview of the ICE parameter estimation methods and the MPM classification schemes for these Markov models. More detailed descriptions of models and methods, as well as implementation details, are given in [10]. The main issue of this article is a new approach, based on a combination of the two schemes mentioned above. Estimation and classification results obtained on simulated SAR images are reported in section 4, and our conclusions are resumed in section 5.

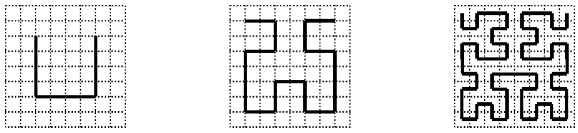
2. MODELS

Let S be the finite set corresponding to the N pixels of an image. We consider two random processes $\mathbf{Y} = (Y_s)_{s \in S}$ and $\mathbf{X} = (X_s)_{s \in S}$, where $\mathbf{Y}$ represents the observed image and $\mathbf{X}$ the unknown class image. Typically, each observed pixel amplitude Y_s depends on several factors, including the radiometric properties of the underlying class X_s, the transfer function of the imaging system, and various noise phenomena. Each random variable X_s takes its values from a finite set of classes (labels), whereas each realisation of Y_s is a real (or complex) value. We here suppose that the random variables $\mathbf{Y}$ 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. This basically means that



(a) (b)

Fig. 1. (a) Pixel neighbourhood structure V and (b) associated click families C_1 and C_2 in the case of 4-connectivity.



(a) (b) (c)

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

we suppose spatially uncorrelated noise.

$\mathbf{X}$ is in many cases well described by a Markov model. In this context we 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

Markov random fields are defined with respect to a neighbourhood structure V, such as the one shown in Fig. 1 (a). $\mathbf{X}$ is a Markov random field if and only 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 its neighbourhood V_s.

On certain conditions, which are generally assumed in digital imagery, the Hammersley-Clifford theorem [11] establishes the equivalence between a Markov random field, defined with respect to a certain neighbourhood structure V, and a Gibbs field whose potentials are associated with V. The potentials are computed on the system of clicks $\mathbf{C}$, which describes the elementary relationships within the neighbourhood. We have restricted ourselves to a 4-neighbourhood and click family C_2, as shown in Fig. 1, and Potts model with regularity parameter λ [10].

The Hammersley-Clifford theorem, Bayes' rule and the conditional independence of the samples allow us to derive analytical expressions for the global *a posteriori* distribution $P(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y})$ and the local *a posteriori* probabilities $P(X_s = x_s|Y_s = y_s, (X_t = x_t)_{t \in V_s})$ in the framework of the HMRF model.

It is not possible to create an *a posteriori* realisation of $\mathbf{X}$ according to $P(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y})$ directly, but it can be approximated iteratively, e.g. with Gibbs sampler [12]. This algorithm can also be used to create *a priori* realisations of $\mathbf{X}$. Due to their iterative nature and slow convergence, the computing time is considerable.

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 by Fig. 2. (c). 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 $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ be the vectors of random variables ordered according to such a transformation of the class image and the observed image, respectively.

According to the definition, $\mathbf{X}$ is a first order Markov chain if and only if the probability that $X_n = x_n$, given the classes of all the preceding elements in the chain $X_1 = x_1, \ldots, X_{n-1} = x_{n-1}$, is equal to the probability of $X_n = x_n$ conditional on $X_{n-1} = x_{n-1}$ only.

The regularity parameters of the HMC model are the elements of a stationary transition matrix $\mathbf{A}$, and *a posteriori* realisations can be computed directly, i.e., without iteration, through so-called *forward-backward recursions*. Unfortunately, the original forward-backward recursions are subject to serious numerical problems [13, 3], so we have used the alternative method proposed by Devijver et al. [13, 10]. The fact that no iterative procedures are necessary to simulate *a posteriori* realisations makes this approach much faster than estimation and classification schemes based on HMRF models.

3. METHODS

In practise, the regularity parameters and the parameters of the distributions of the classes are often unknown and must be estimated from the observation. The problem is then double: We neither know the characteristics of the classes nor 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 only consider the latter here.

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 criteria, such as the MAP, which aims at maximising the global *a posteriori* probability $P(\mathbf{X}|\mathbf{Y})$, and the MPM, which consists in maximising the posterior marginal distribution $P(X_s|\mathbf{Y})$ for each pixel. We here only consider MPM classification.

The methods are only described very briefly in this article, and we focus on the differences between the two Markov models. As already mentioned, more thorough descriptions

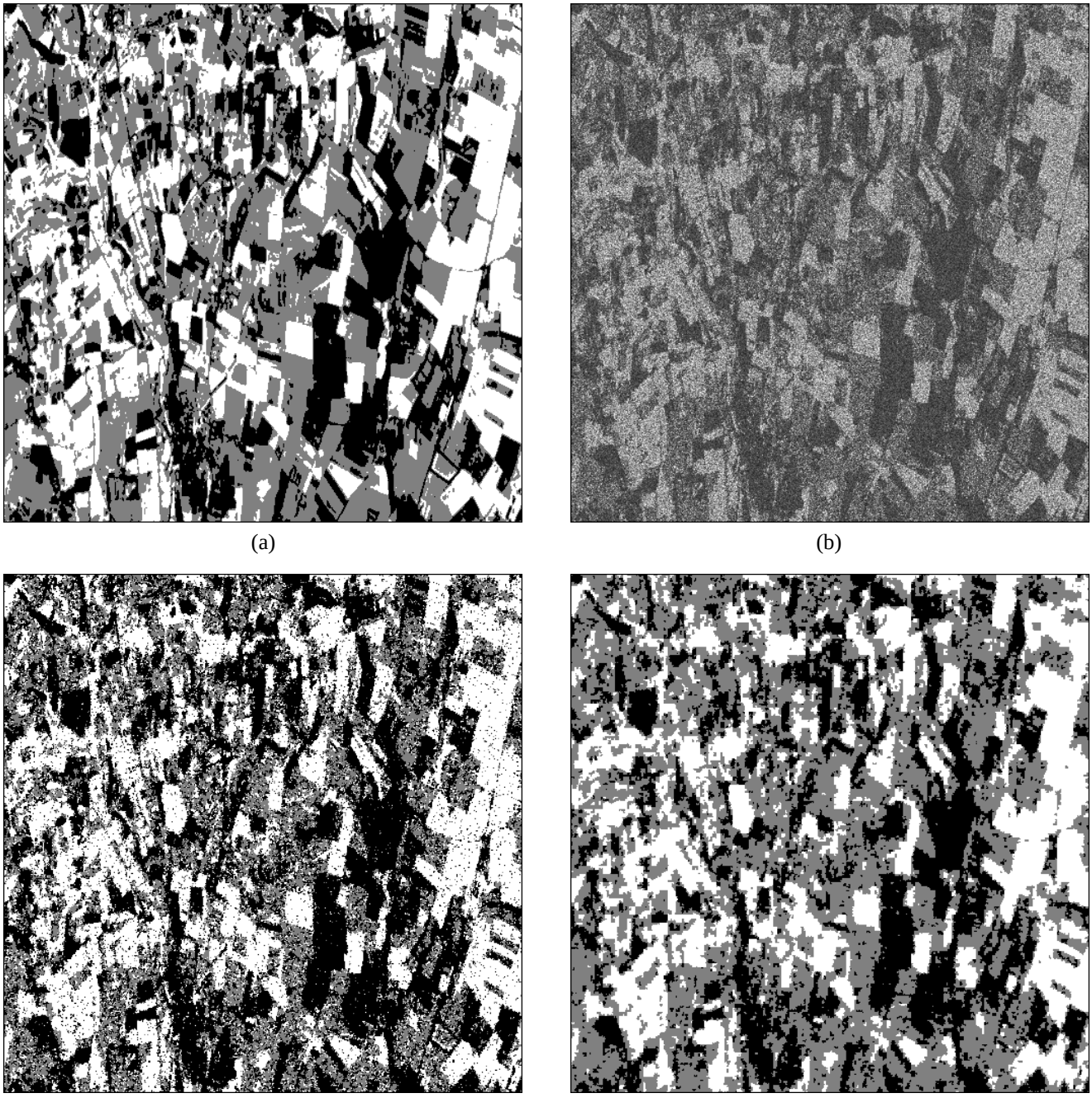


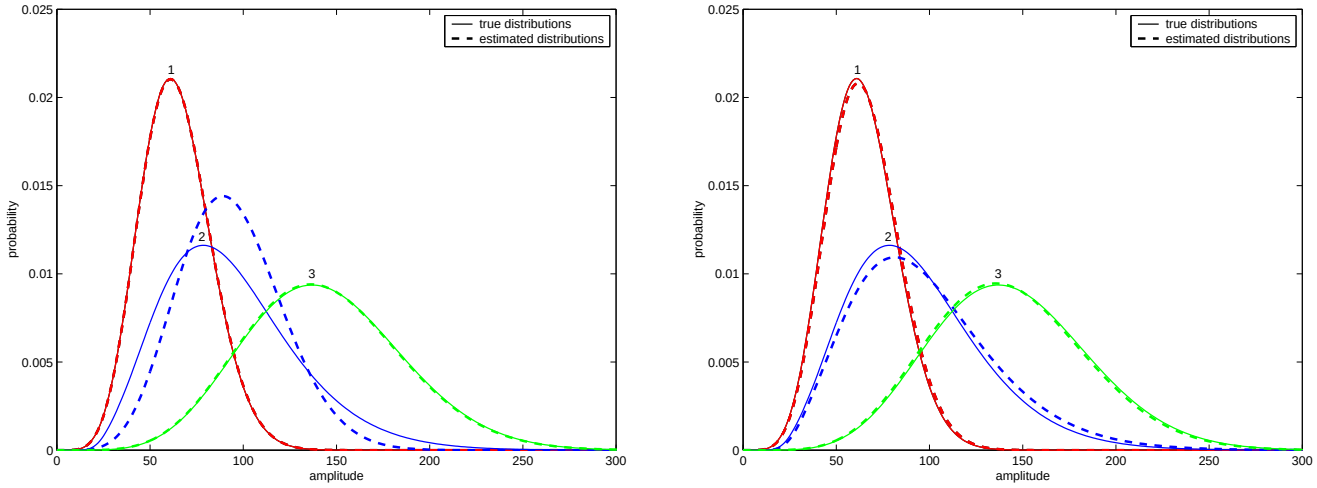
Fig. 3. Classification of a simulated 3-look SAR image into 3 classes: (a) Ideal class image. (b) Speckled amplitude image. Results obtained (c) with the HMRF method and (d) with the HMC method.

and implementation details are given in [10]. However, we here explain how these two approaches can be combined in order to improve the efficiency.

3.1. Mixture Estimation

ICE consists in computing the conditional expectation of the parameters. As the ideal class image $\mathbf{X}$ is unknown, the ICE algorithm iteratively creates *a posteriori* realisations and recalculates the class and regularity parameters. In the framework of classic mixture estimation and HMRFs, the following computation steps are carried out for each iteration:

- A certain number of *a posteriori* realisations are computed with Gibbs sampler, based on the parameters obtained in the previous iteration.
- The class parameters (which are the mean value and the variance for a scalar Gaussian distribution) and the regularity parameters are estimated for each realisation and then averaged.



(a) (b)

Fig. 4. Fit of the generalised mixture estimation performed (a) by the HMRF method and (b) by the HMC method on the simulated 3-look SAR image with 3 classes.

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

In the case of HMRFs, the regularity parameter λ cannot be estimated directly. We have chosen to estimate λ in each ICE iteration with a modified stochastic gradient algorithm, relying on a series of *a priori* realisations computed with Gibbs sampler [10].

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.

The procedure is similar for the analysis scheme based on HMCs [10], except that *a posteriori* realisations can be simulated without iteration, through forward-backward recursions. Moreover, the regularity parameters can be computed directly from the conditional probabilities. This approach is therefore considerably faster.

3.2. Generalised Mixture Estimation

In generalised mixture estimation, the distribution 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 distance is used to determine the most appropriate distribution family for each class [4, 10].

3.3. Classification

When the distribution families and the associated parameters of each class have been determined, we can proceed to the classification. In the case of HMRFs, the MPM [14] is computed as follows:

- A series of independent *a posteriori* realisations of $\mathbf{X}$ are computed iteratively, using Gibbs sampler.
- For each pixel we retain the class that has occurred the most frequently there.

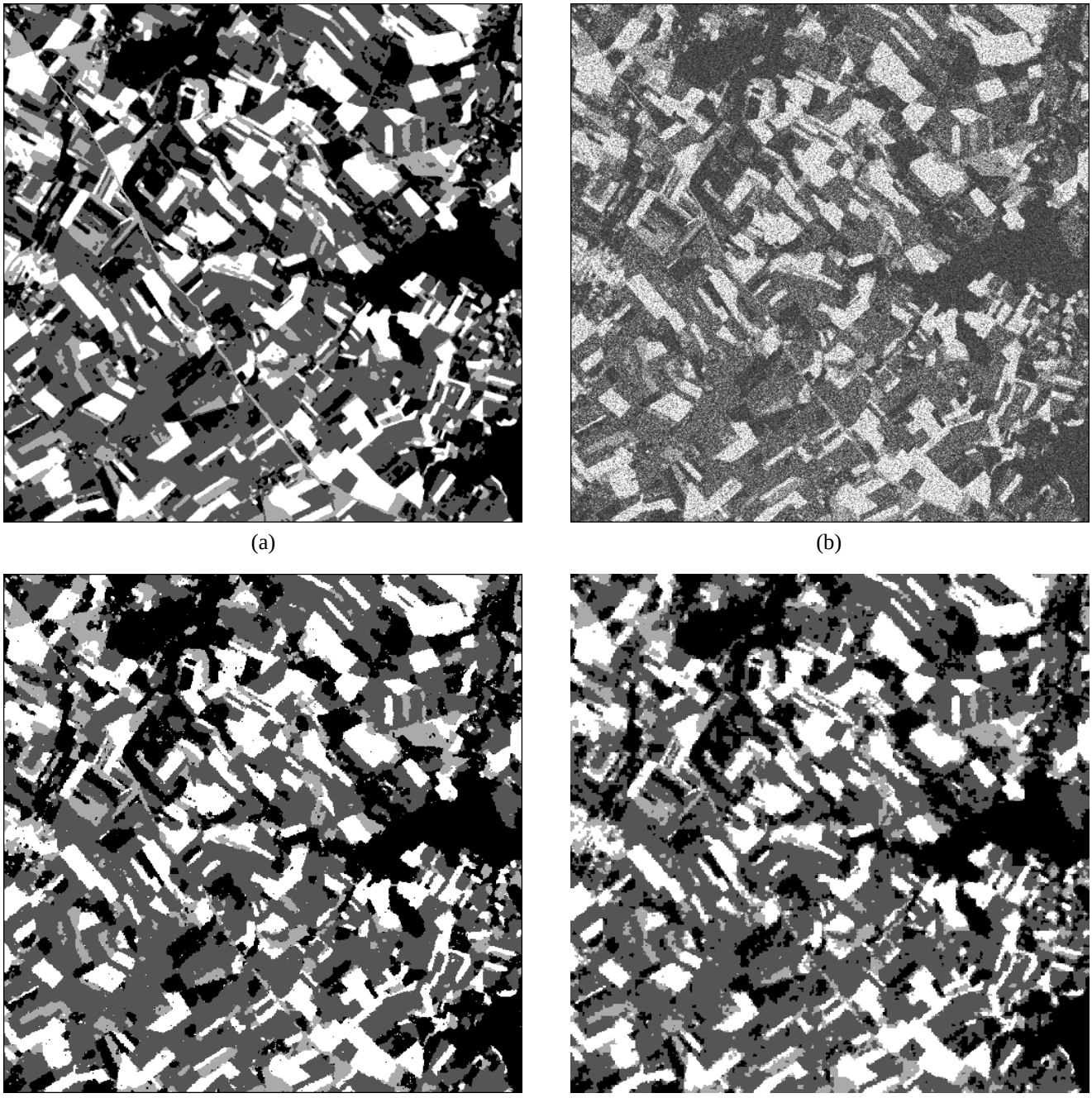
The required number of *a posteriori* realisations and iterations per realisation will be discussed in section 4.

The MPM solution can be calculated directly for HMCs, based on one forward-backward recursion. We simply attribute each pixel to the class that maximises the marginal *a posteriori* probability.

3.4. Hybrid Method

Preliminary tests [10] revealed that the scheme based on HMRFs in many cases has difficulties in estimating the regularity parameters, whereas the estimation method based on HMCs is very robust. The latter method is also much faster. Nevertheless, in cases where the HMRF method estimates the regularity parameters correctly, it provides a more satisfactory classification result. In particular, it produces more regular region borders.

In order to take advantage of the positive sides of both approaches, while avoiding the negative ones, we propose a hybrid method, where we first run the ICE and MPM algorithms in the framework of the HMC model, and then, initialising with the estimated parameters, effectuate only a very restricted number of ICE iterations and the final MPM classification with the HMRF versions of the algorithms.



(a) (b) (c) (d)

Fig. 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 HMRF method and (d) with the HMC method.

4. RESULTS

In order to compare the three approaches quantitatively, we have created two simulated synthetic aperture radar (SAR) images. Both visual interpretation and automatic analysis of data from imaging radars are particularly difficult, due to 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.

The possible distribution families in the generalised mixture has been chosen so that they are well adapted to single- and multi-look amplitude radar images [15]. 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

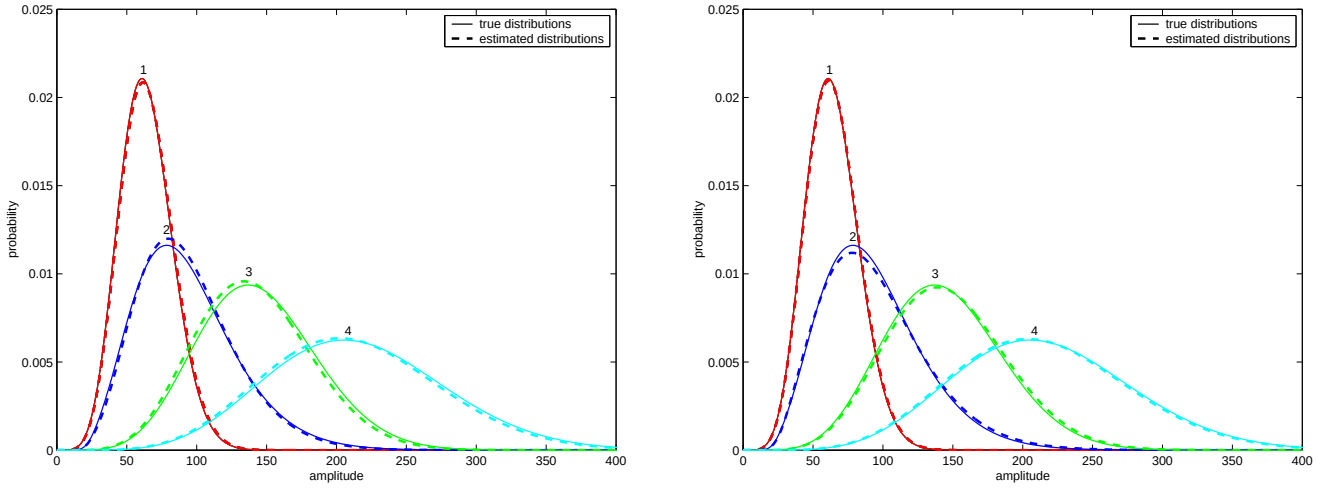


Fig. 6. Fit of the generalised mixture estimation performed (a) by the HMRF method and (b) by the HMC method on the simulated 3-look SAR image with 4 classes.

Gamma distributed. 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 distributions are given in [10]. The equivalent number of independent looks L of the image is assumed to be known. The mean reflectivity R and the texture parameter a are estimated automatically [10].

4.1. Simulated SAR Image with Three Classes

Figs. 3 (a) and (b) represent an ideal class image with 3 classes and its 3-look ($L = 3.0$) speckled counterpart. The darkest and brightest classes are both Gamma distributed, whereas the one in the middle is K distributed with texture parameter $a = 4.0$. The contrast between two consecutive classes is 3.5 dB. The image size is 512×512 pixels.

Let us first consider the ICE and MPM algorithms based on the HMRF model. We use 30 iterations for the ICE algorithm, with only one *a posteriori* realisation per iteration. As the resolution is not the same vertically and horizontally for this image, we allow different regularity parameters in the two directions, with initial values $\lambda_x = \lambda_y = 0.5$. Within each ICE iteration, the maximum number of iterations for the stochastic gradient is set to 10, with one *a priori* realisation per iteration, but we interrupt the iteration earlier if λ_x and λ_y differs less than 0.01 from the previous values. 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 HMRF model relies on 10 *a posteriori* realisations. The regularity parameters were estimated to $\lambda_x = 0.37$ and $\lambda_y = 0.68$, respectively. However, the classification result in Fig. 3 (c) seems far too irregular, and only 72.7 % of the pixels are correctly classified.

The number of ICE iterations is set to 30 also for the corresponding analysis scheme based on the HMC model, and we compute only one *a posteriori* realisation per iteration. Fig. 3 (d) shows the result of the MPM classification. The method based on HMCs here gives a more satisfactory result, with 83.9 % of correctly classified pixels. In particular, the overall regularity of the image seems to be better estimated.

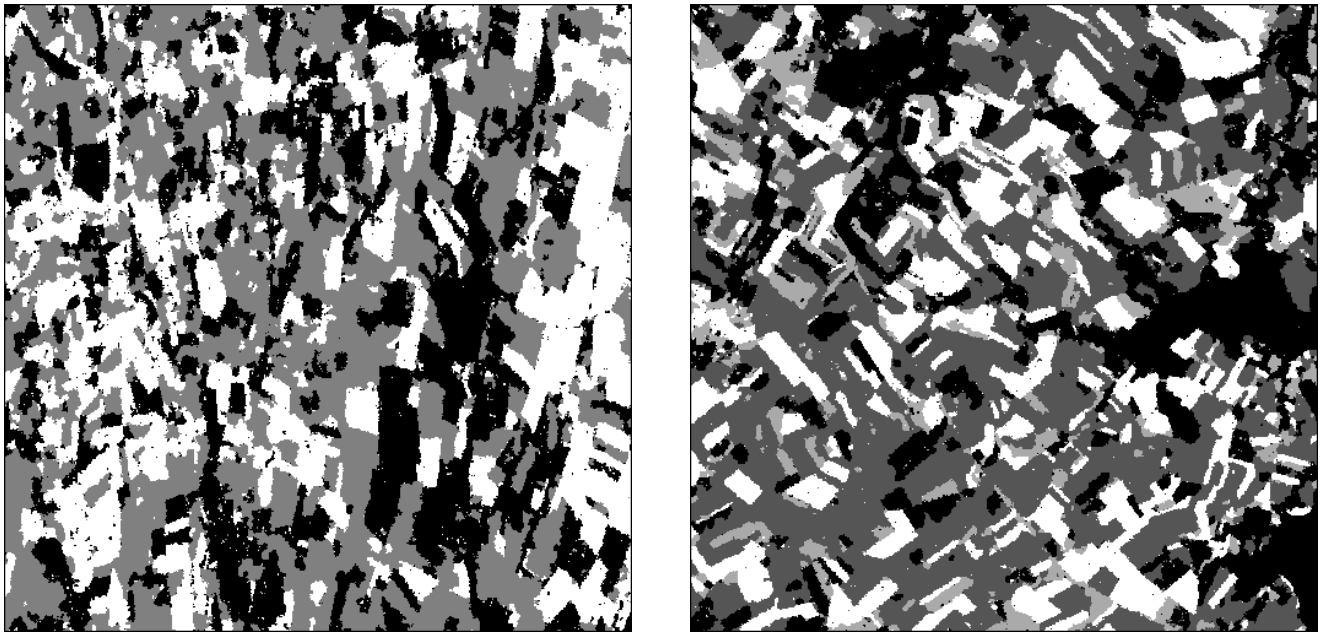
Both methods correctly identify Gamma distributions for the darkest and brightest classes, but only the HMC method found that the class in the middle was K distributed. The fit between the estimated and the true distributions in Fig. 4 is generally very good, except for the case where the HMRF method chooses the wrong distribution family.

The computing time is quite different for the two approaches: The programme based on HMRFs spent about 120 minutes on a PC with a Pentium III 733 MHz processor running Linux, whereas the programme based on HMCs only needed 3 minutes and 12 seconds. Hence the latter method was 37 times quicker for this image.

4.2. Simulated SAR Image with Four Classes

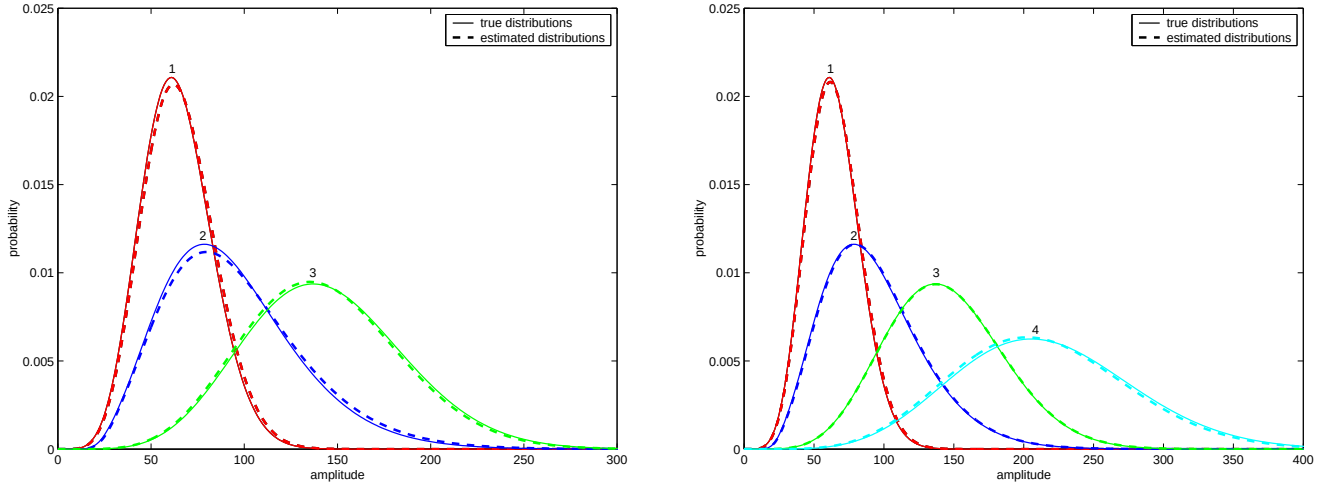
Fig. 5 (a) represents an ideal and approximately isotropic 512×512 class image with 4 classes, and Fig. 5 (b) shows the corresponding speckled image. The parameter settings and the radiometric characteristics of the classes are the same as for the image in Fig. 3, except that an additional Gamma distributed class with higher reflectivity has been added. Fig. 5 (c) and (d) represent the results obtained with the HMRF and HMC methods, respectively.

The regularity parameter for the HMRF method is here $\lambda = 0.65$, which visually gives a very satisfactory result, and the proportion of correctly classified pixels is 87.0 %, whereas it is 85.2 % for the method based on HMCs. The borders are slightly more irregular for the latter method, but



(a) (b)

Fig. 7. Classification results for the simulated 3-look SAR images with (a) 3 classes and (b) 4 classes, obtained with the new method combining HMC and HMRF models.



(a) (b)

Fig. 8. Fit of the generalised mixture estimation for the simulated 3-look SAR images with (a) 3 classes and (b) 4 classes, obtained with the new method combining HMC and HMRF models.

some of the narrow structures are better preserved. Both methods correctly identified the distribution families of all four classes, and the fit of the estimated distributions is comparable, as can be seen from Fig. 6. The HMC algorithm was, however, 25 times faster. The computing time was here 3 minutes and 19 seconds, against 84 minutes and 21 seconds for the HMRF method.

4.3. Hybrid Method

The rationale of the new hybrid method described in section 3.4 is to combine the computing speed and the robust mixture estimation of the HMC method with the fine spatial definition of HMRF models. We have used 30 ICE iterations for the HMC part, followed by one single ICE iteration with one *a posteriori* realisation and the MPM with 10 *a posteriori* realisations for the HMRF algorithms.

Fig. 7 (a) represents the classification result obtained with the new method on the simulated SAR image with 3 classes. Visually, the classification result is far better than those given in Fig. 3, and the portion of correctly classified pixels has increased to 85.8 %. The distribution families of the different classes are correctly identified, and the estimated regularity parameters are $\lambda_x = 0.61$ and $\lambda_y = 0.67$.

Comparing Fig. 8 (a) with Figs. 4 (a) and (b), we see that the overall fit of the estimated distributions is improved. The computing time is 11 m 38 s, which is nearly 4 times slower than the algorithm based on HMCs, but 10 times faster than the method based on HMRFs.

Likewise, Fig. 7 (b) represents the classification result obtained with the hybrid method on the simulated SAR image with 4 classes. It is very close to the classification result in Fig. 5 (c), obtained with the HMRF method. The portion of correctly classified pixels is the same (87.0 %) and the estimated regularity parameter is very close ($\lambda = 0.67$), but the new method is approximately 7 times faster. The fit of the estimated distributions, shown in Fig. 8 (b), is also slightly better.

5. CONCLUSION

This article describes unsupervised classification of noisy images in the framework of hidden Markov models and generalised mixture estimation. Methods based on HMCs, applied to a Hilbert-Peano scan of the image, constitute an interesting alternative to methods based on HMRFs, mainly because they are considerably faster and because the estimation of the regularity parameters is much more robust. However, HMRFs ultimately offer finer spatial modelling. We therefore propose a new hybrid method, where we use HMCs first and HMRFs only for the final estimation and classification. Tests on simulated SAR images have confirmed that the hybrid method is fast and robust and yields excellent mixture estimation and classification results.

6. ACKNOWLEDGMENTS

Part of this work was effectuated in the framework of the project "Segmentation d'Images Radar", financed by the Groupe des Ecoles des Télécommunications (GET). The authors thank the GET for the scientific and financial support.

7. REFERENCES

[1] K. Abend, T. J. Harley, and L. N. Kanal, "Classification of binary random patterns," *IEEE Trans. Information Theory*, vol. IT-11, no. 4, 1965.

[2] W. Skarbek, "Generalized Hilbert scan in image printing," in *Theoretical Foundations of Computer Vision*, R. Klette and W. G. Kropetsh, Eds. Akademie Verlag, Berlin, 1992.

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

[4] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Trans. 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," *Annals of Mathematical Statistics*, vol. 41, pp. 164–171, 1970.

[6] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 34, no. 3, pp. 618–633, 1993.

[7] W. Pieczynski, "Statistical image segmentation," *Machine Graphics and Vision*, vol. 1, no. 1/2, pp. 261–268, 1992.

[8] B. Braathen, W. Pieczynski, and P. Masson, "Global and local methods of unsupervised Bayesian segmentation of images," *Machine Graphics and Vision*, vol. 2, no. 1, pp. 39–52, 1993.

[9] Y. Delignon, A. Marzouki, and W. Pieczynski, "Estimation of generalised mixtures and its application in image segmentation," *IEEE Trans. Image Processing*, vol. 6, no. 10, pp. 1364–1375, 1997.

[10] R. Fjørtoft, J.-M. Boucher, Y. Delignon, R. Garello, J.-M. Le Caillec, H. Maître, J.-M. Nicolas, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised classification of radar images based on hidden Markov models and generalised mixture estimation," in *Proc. SAR Image Analysis, Modelling, and Techniques V*, Barcelona, Spain, 25–29 September 2000, vol. SPIE 4173.

[11] J. Besag, "Spatial interactions and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society*, vol. B-36, pp. 192–236, 1974.

[12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, November 1984.

[13] 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.

[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, no. 397, pp. 76–89, 1987.

[15] Y. Delignon, R. Garello, and A. Hillion, "Statistical modelling of ocean SAR images," *IEE Proc. Radar, Sonar and Navigation*, vol. 144, no. 6, pp. 348–354, 1997.