GRAPHICAL MODELS AND IMAGE PROCESSING
Vol. 57, No. 5, September pp. 389–399, 1995

Adaptive Mixture Estimation and Unsupervised Local Bayesian Image Segmentation

ANRONG PENG AND WOJCIECH PIECZYNSKI

Département Signal et Image, Institut National des Télécommunications, 9, rue Charles Fourier, 91011 EVRY Cedex, France

Received May 26, 1993; revised December 22, 1994; accepted May 23, 1995

This paper addresses mixture estimation applied to unsupervised local Bayesian segmentation. The great efficiency of global Markovian-based model methods is well known, but the efficiency of local methods can be competitive in some particular cases. The purpose of this paper is to specify the behavior of different local methods in different situations. Algorithms which estimate distribution mixtures prior to segmentation, such as expectation maximization (EM), iterative conditional estimation (ICE), and stochastic expectation maximization (SEM), are studied. Adaptive versions of EM and ICE, valid for nonstationary class fields, are then proposed. After applying various combinations of estimators and segmentations to noisy images, we compare the estimators' performances according to different image and noise characteristics. Results obtained attest to the suitability of adaptive versions of EM, ICE, and SEM. Furthermore, the local methods turn out to be robust in the sense that the parameter estimation step does not degrade the final segmentation results significantly, and the choice of EM, ICE, or SEM has little importance.

1. INTRODUCTION

Image segmentation underlies many important problems in image processing. Bayesian classification has been widely applied to image segmentation, resulting in Bayesian segmentation. From a statistical viewpoint, an image is composed of different random fields. Let S denote the finite set of image pixels, $X = (X_s)_{s \in S}$ the random field to model the real image which cannot be observed, and $Y = (Y_s)_{s \in S}$ the random field to model the observed image. Each X_s takes its values in a finite set of classes $\Omega = \{\omega_1, \ldots, \omega_m\}$ and each Y_s takes its values in R. The problem of image segmentation is then the problem of estimating the invisible realization of field X from the visible realization of field Y. There are two families of Bayesian approaches: a local family and a global one. When considering a local family, each X_s is estimated from Y_{V_s}, the restriction of Y to a context V_s of s. Assuming V_s of small size, the Bayesian decision rule is computable. Local methods have been used [6, 12, 13, 19, 28] before application of Markov models to image processing problems. The family of global methods allows one to take the whole information available into account: each X_s is estimated from Y. These approaches, like MAP of Geman and Geman [11], MPM of Marroquin *et al.* [17], and ICM of Besag [2], require models by hidden Markov random fields. The great efficiency of global methods has been widely described [2, 3, 5, 7, 10, 11, 14, 15, 17, 23, 29–31], although some of our previous work shows that in some particular situations the local methods are quite competitive [3, 16]. To be more precise, we consider in [16] two binary images, one homogeneous (HI) and another nonhomogeneous (NHI), each corrupted by a white noise (WN) and a correlated noise (CN). Furthermore, one considers two contextual methods M1, M2, and the global Gibbsian EM method of Chalmond [5]. The ratios of wrongly classified pixels are given in Table 1.

Local methods clearly appear as competitive in the NHI + CN case.

Thus one can imagine that in future complex and automated image processing systems, local Bayesian segmentation methods could take some useful place.

The aim of this paper is to present some investigations contributing to an understanding of the behavior of different unsupervised local methods in different situations. "Unsupervised" means that all parameters needed for segmentation are estimated in a previous step from the noisy data. In local methods this previous statistical problem is the mixture estimation one.

We are mainly interested in two points:

(1) Among the three mixture estimation algorithms, namely expectation maximization (EM, [9, 26]), stochastic expectation maximization (SEM, [4, 18]) and iterative conditional estimation (ICE, [22]), which should be used in the context considered?

(2) In adaptive segmentation one considers that priors change with s and the needed parameters are then estimated by "adaptive" versions of EM, SEM [24], or ICE. Is such an adaptive walk to be preferred to classical walk in the context of unsupervised local segmentation?

1077-3169/95 $12.00

TABLE 1

Rate of Wrongly Classified Pixels by Two Local Methods (M1, M2) and a Global Method (Gibbsian EM) Applied to Homogeneous (HI) and Nonhomogeneous (NHI) Images Corrupted with a White Noise (WN) and a Correlated (CN) Noise

	M1 (%)	M2(%)	Gibbsian EM (%)
HI + WN	10.7	11	3.2
HI + CN	14.9	15.6	10.5
NHI + WN	16.1	17.1	13.6
NHI + CN	13.3	13.6	31.7

Note. Theoretical rate of blind segmentation is the same for the four images.

We select from the numerous simulation results given in [20] and put forth some general conclusions.

The paper is organized as follows. In Section 2, we give a brief presentation of local segmentation and mixture estimation problems. Three mixture estimation algorithms EM, ICE, and SEM, are described along with their adaptive versions. The third section is devoted to data considered in the study. The fourth section presents some simulation results and comments. Conclusions are given in Section 5.

2. UNSUPERVISED LOCAL BAYESIAN IMAGE SEGMENTATION

In this section we recall the principle of local segmentation and summarize the different parameter estimation algorithms.

2.1. *Local Bayesian Segmentation*

Let (X_V, Y_V) be the restriction of the fields X, Y to $V \subset S$. The Bayesian rule of decision $\hat{s}_B$ is defined from the distribution of (X_V, Y_V), assumed here independent from the position of V in S, by

$$[\hat{X}_s = \hat{s}_B(y_V) = x_i] \Leftrightarrow [DF(x_i, y_V) = \operatorname*{Max}_{x_j \in \Omega} DF(x_j, y_V)], \quad (1)$$

where the discriminating functions DF are

$$DF(x_i, y_V) = \sum_{x \in \Omega^{|V|-1}} P[X_s = x_i, X_{V-\{s\}} = x, Y_V = y_V] \quad (2)$$

for every x_i in Ω and $|V|$ cardinal of V. Thus the segmented image is obtained by applying the rule above to each pixel s.

2.2. *Mixture Estimation*

The calculation of these discriminating functions requires knowledge of the distribution of (X_V, Y_V). In the following, Y_V will be assumed Gaussian conditionally on X_V. Let us denote by $V_1, \ldots, V_n$ the sequence of contexts chosen in S and, in order to simplify things, $X_1 = X_{V_1}, \ldots X_n = X_{V_n}$ and $Y_1 = Y_{V_1}, \ldots, Y_n = Y_{V_n}$ the restrictions of fields X, Y to $V_1, \ldots, V_n$. For $m = \text{Card}(\Omega)$ and $r = \text{Card}(V)$, the random variables $X_1, \ldots, X_n$ take their values in a finite set of cardinal $M = m^r$, and the random variables $Y_1, \ldots, Y_n$ in R^r. Thus the common distribution of $Y_1, \ldots, Y_n$ is a mixture of $M = m^r$ Gaussian distributions on R^r. For $x_1, \ldots, x_M$ possible values of each X_i (or possible configurations of X_V), let us denote by $\alpha_1 = P[X_i = x_1], \ldots, \alpha_M = P[X_i = x_M]$ the priors and $\beta_1 = (m_1, \Gamma_1), \ldots, \beta_M = (m_M, \Gamma_M)$ the mean vectors and covariance matrix defining the distributions of Y_i conditioned on $X_i = x_1, \ldots, x_M$, respectively. Gaussian densities defined by $\beta_1 = (m_1, \Gamma_1), \ldots, \beta_M = (m_M, \Gamma_M)$ will be written $f_1, \ldots, f_M$. Thus $\alpha = (\alpha_1, \ldots, \alpha_M)$ defines priors and $\beta = (\beta_1, \ldots, \beta_M)$ defines the distributions of each Y_i conditioned on $X_i = x_1, \ldots, x_M$, respectively.

Finally, $\theta = (\alpha, \beta)$ defines the distribution of (X_V, Y_V), and the problem is to estimate it from $Y_1, \ldots, Y_n$.

Let temporarily assume that $X_1, \ldots, X_n$ are observable. The maximum likelihood estimator (MLE) $\hat{\theta} = \hat{\theta}(X, Y)$ is defined by

$$\hat{\theta}(X, Y) = \operatorname{Arg} \max_{\theta} \log[f_\theta(X, Y)], \quad (3)$$

which gives in our case

$$\hat{\alpha}_i(X) = \frac{\sum_{j=1}^{n} 1_{[X_j = x_i]}}{n} \quad (4)$$

$$\hat{m}_i(X, Y) = \frac{\sum_{j=1}^{n} Y_j 1_{[X_j = x_i]}}{\sum_{j=1}^{n} 1_{[X_j = x_i]}} \quad (5)$$

$$\hat{\Gamma}_i(X, Y) = \frac{\sum_{j=1}^{n} (Y_j - \hat{m}_i)^t (Y_j - \hat{m}_i) 1_{[X_j = x_i]}}{\sum_{j=1}^{n} 1_{[X_j = x_i]}}. \quad (6)$$

As X is not observable, EM allows one to define a sequence $(\theta^k)_{k \in N}$ using the principle

$$\hat{\theta}^{k+1}(Y) = \operatorname{Arg} \max_{\theta} \log[E_{\theta_k}[f_\theta(X, Y)/Y)]. \quad (7)$$

Let us denote by α_{ij}^k the θ^k-based probability of $X_j = x_i$ conditioned on Y_j:

$$\alpha_{ij}^k(Y_j) = \frac{\alpha_i^k f_i^k(Y_j)}{\sum_{q=1}^{M} \alpha_q^k f_q^k(Y_j)}. \quad (8)$$

In the case of interest, the sequence $(\theta^k)_{k \in N}$ is given by

$$\alpha_i^{k+1}(Y) = \frac{\sum_{j=1}^{n} \alpha_{ij}^k(Y_j)}{n} \tag{9}$$

$$m_i^{k+1}(Y) = \frac{\sum_{j=1}^{n} \alpha_{ij}^k(Y_j) Y_j}{\sum_{j=1}^{n} \alpha_{ij}^k(Y_j)} \tag{10}$$

$$\Gamma_i^{k+1}(Y) = \frac{\sum_{j=1}^{n} \alpha_{ij}^k(Y_j)[Y_j - m_i^{k+1}(Y)]^t[Y_j - m_i^{k+1}(Y)]}{\sum_{j=1}^{n} \alpha_{ij}^k(Y_j)}. \tag{11}$$

According to its principle, the ICE algorithm [22] produces a sequence $(\theta^k)_{k \in N}$ defined by

$$\hat{\theta}^{k+1}(Y) = E_{\theta_k}[\hat{\theta}(X, Y)/Y]. \tag{12}$$

Applying (12) to (4) we obtain (9): the reestimation of priors is exactly the same in the ICE case as in the EM case. The calculation of (12) applied to (5) and (6) is not possible and we have to resort to stochastic approximation. In fact, if $x^1, \ldots, x^N$ is a sequence of realizations of $X = (X_1, \ldots, X_n)$ according to the θ^k-based distribution of $X = (X_1, \ldots, X_n)$ conditioned on $Y = (Y_1, \ldots, Y_n)$ (we have $x^i = (x_1^i, \ldots, x_n^i)$), θ^{k+1} defined by (12) can be approached, by virtue of the law of large numbers, by

$$\theta^{k+1}(Y) = \frac{[\hat{\theta}(x^1, Y) + \cdots + \hat{\theta}(x^N, Y)]}{N}. \tag{13}$$

Simulation studies show that one can take, in this case, $N = 1$ without significant perturbation of the results.

Finally, the next values of m_i and Γ_i are obtained by simulating one realization $x^1 = (x_1^1, \ldots, x_n^1)$ of X according to its θ^k-based distribution conditioned on Y and applying to $x^1 = (x_1^1, \ldots, x_n^1)$ formulas (5) and (6).

The reestimation of priors by the SEM is obtained by applying (4) to $x^1 = (x_1^1, \ldots, x_n^1)$, while the reestimation of m_i and Γ_i is the same as that of ICE with $N = 1$ above.

2.3. *Adaptive Versions of EM, ICE, and SEM*

We present below adaptive versions of EM, ICE, and SEM, which will be denoted by AEM, AICE, and ASEM, respectively. The basic idea, which is exposed in the SEM case in [24, 25], is to make α depend on pixels and use the reestimation formulas for priors on a window containing s. This makes all unsupervised procedures valid in the nonstationary class field case; however, the field Y remains stationary conditionally to X. More precisely, the distribution of X_V depends on the position of V in S, but that the distributions of Y_V conditional to realizations of X_V are independent of this position.

Returning to the notations of the previous subsection, we associate with each pixel s a window W_s. The reestimation of each α_s by AEM, AICE, or ASEM runs as the reestimation of α by EM, ICE, or SEM, with the difference that W_s is the sample used. The reestimations of β by EM, ICE, SEM and AEM, AICE, ASEM, respectively, are the same.

3. REFERENCE DATA

We consider three images of different homogeneity: "Ring," "Gibbs," and "Letter B" (Fig. 1). Each is corrupted with four different Gaussian noises; two are white (W), and two correlated (C). One of the white noises is "means discriminating" (MD: $m_1 \neq m_2$, $\sigma_1^2 = \sigma_2^2$), while the other is "variances discriminating" (VD: $m_1 = m_2$, $\sigma_1^2 \neq \sigma_2^2$), and likewise for the correlated noises. The four noises so obtained will be denoted by WMD, WVD, CMD, and CVD, respectively. We take $m_1 = 1$, $m_2 = 2$ and $\sigma_1^2 = \sigma_2^2 = 1$ in the MD cases and $\sigma_1^2 = 1, \sigma_2^2 = 4$ in the VD ones. We give three examples (Fig. 2) which show that visual aspects of the corrupted images change with the noise nature. Furthermore, as we shall see in the following, the efficiency of different unsupervised segmentation methods also depends on the four noises considered.

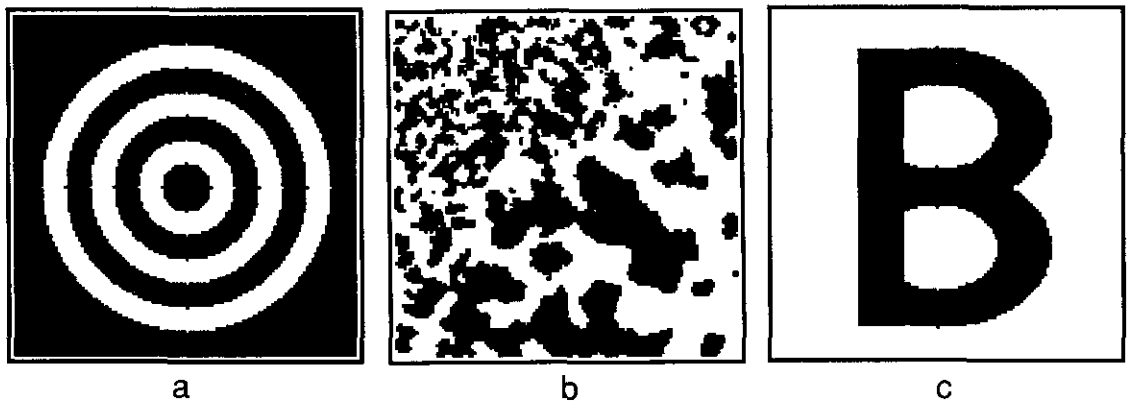


FIG. 1. (a) "Ring," (b) "Gibbs," and (c) "Letter B."

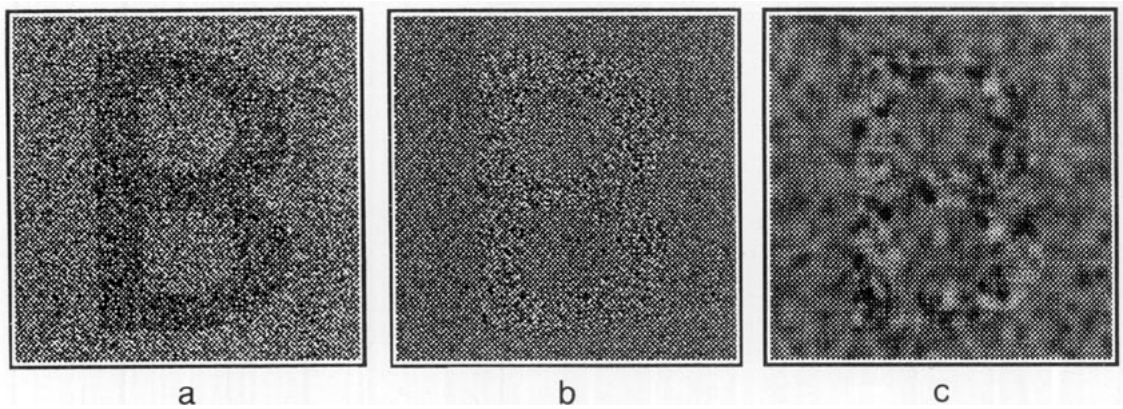


FIG. 2. (a) "Letter B" + WMD, (b) "Letter B" + WVD, and (c) "Letter B" + CVD.

4. UNSUPERVISED SEGMENTATION RESULTS

4.1. *General Results*

To each of the 12 noisy images we have applied 12 unsupervised segmentation methods: 6 are blind with respectively EM, ICE, SEM, and their adaptive versions for the estimation step, and likewise for the 6 unsupervised contextual segmentation methods. The context used contains the pixel considered and one neighbor.

Before analyzing the results, let us note the following point. For each statistical image segmentation method there exists a theoretical error, obtained with the true parameter values, which depends on the signal-to-noise ratio. When the noise is too dominant for a given method, the method becomes unreliable and it is pointless to proceed to a parameter estimation algorithm. If we want to compare different unsupervised segmentation methods we have to deal with limit cases: if the noise is not dominant they will all be good, whereas if it is too dominant they will all be bad. In our examples we have chosen the cases in which the Bayesian theoretical error ratio is about 30%. As we are essentially interested in the efficiency of different segmentation methods, we do not give here parameter estimation results which form an intermediary step. All results concerning parameter estimation can be seen in [20].

The rates of wrongly classified pixels by different unsupervised methods must be compared with "theoretical" rates, i.e., obtained from real and noisy images, which are given in Table 2.

Let us note that the theoretical rate above is the rate obtained using parameters estimated from both real and noisy images and is not exactly the theoretical Bayesian rate. In fact, priors in each sample used are not necessarily stationary, and the problem of compatibility of the model used can arise. To be more precise, let us consider a sample of 100 points where the first 50 points are distributed according to priors $\alpha_1 = 0.3$, $\alpha_2 = 0.7$ and the final 50 points are distributed according to priors $\alpha_1 = 0.7$, $\alpha_2 = 0.3$. If we assume sample stationarity, the estimation of priors will give approximately $\hat{\alpha}_1 = 0.5$, $\hat{\alpha}_2 = 0.5$, which will be used in the discriminating functions. Thus the first 50 noisy points will be wrongly classified, as with the final 50 points, and the theoretical rate will be superior to the Bayesian rate which corresponds to the case where the whole sample is distributed according to priors $\alpha_1 = 0.5$, $\alpha_2 = 0.5$. In particular, this explains why the theoretical rate varies with images. In Fig. 3 we give all the results that we have obtained.

TABLE 2
Theoretical Rate of Wrongly Classified Pixels

Ring						Gibbs						Letter B					
Blind		Contextual				Blind		Contextual				Blind		Contextual			
MD	VD	MD		VD		MD	VD	MD		VD		MD	VD	MD		VD	
		WH	CR	WH	CR			WH	CR	WH	CR			WH	CR	WH	CR
27,0	25,2	22,4	26,5	21,4	32,4	30,5	35,3	27,0	30,2	30,4	45,9	25,1	23,4	20,2	24,9	19,6	29,9

Note. MD, means discriminating noise; VD, variances discriminating noise; WH, white noise; CR, correlated noise; contextual, context limited to one pixel.

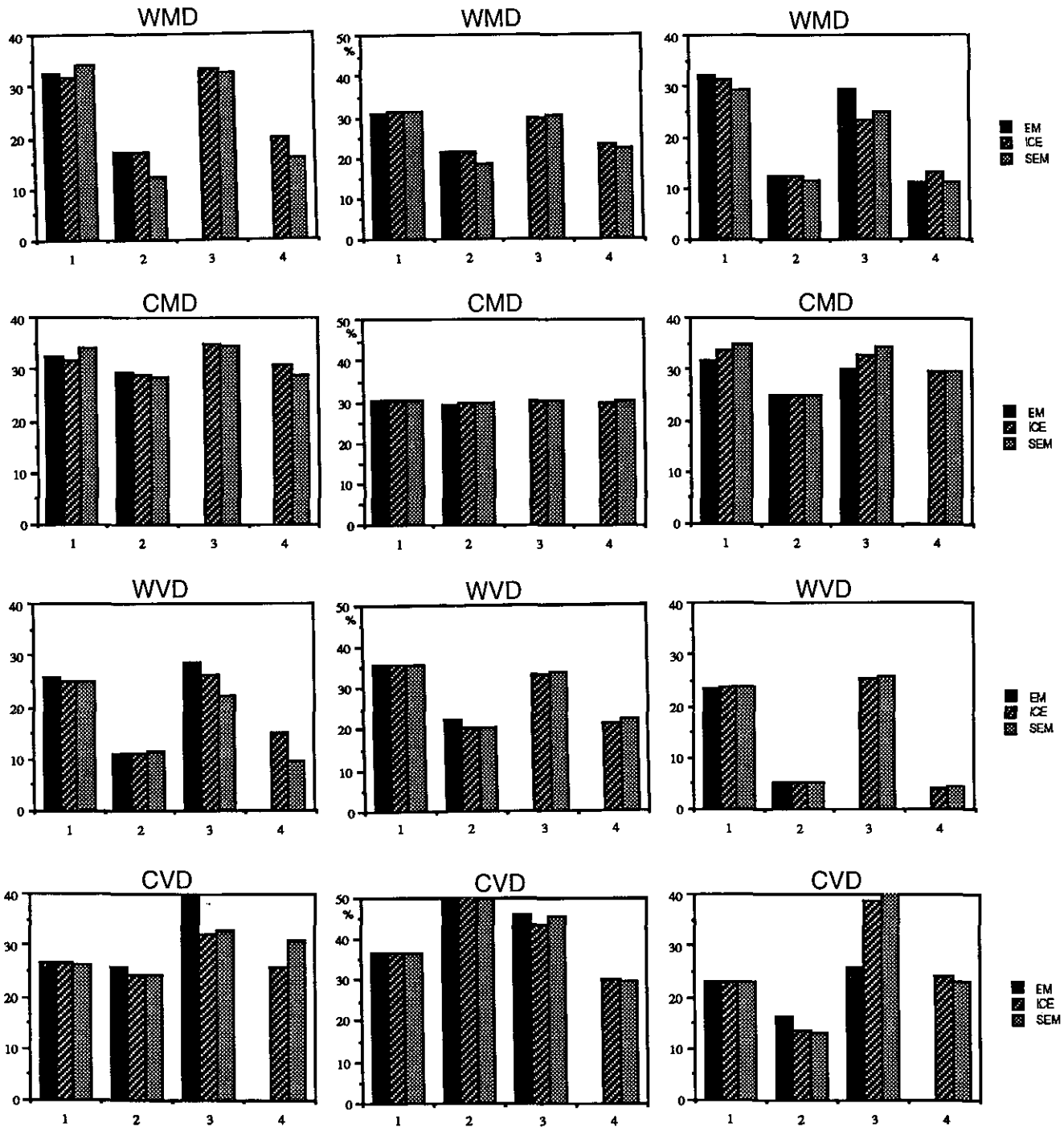


FIG. 3. Percentage of wrongly classified pixels in (left) "Ring," (middle) "Gibbs," and (right) "Letter B." (1) Blind segmentation, (2) blind adaptive segmentaton, (3) contextual segmentation, and (4) contextual adaptive segmentation.

The most striking result is the great efficiency, in some situations, of the adaptive methods. They prove extremely successful in the case of the homogeneous image (letter B) corrupted with a VD white noise. The contextual adaptive SEM- and ICE-based methods give an error of 4.15 and 4.38%, respectively, when the theoretical error is 19.6%. The blind adaptive EM-, SEM-, and ICE-based methods give an error of 5.21, 5.20, and 5.13%, respectively, when the theoretical error is 23.4%. We note that in the VD case its use is always effective, except in the case of very nonhomogeneous images (image "ring") and correlated noise. In the case of letter B corrupted with MD white

noise, the unsupervised contextual ASEM-based method gives 11% as an error rate when the theoretical error is 20.2%. Similar remarks apply to blind segmentation: AICE- and ASEM-based methods give about 12% when the theoretical error is 25.1%. This confirms the good behavior of ASEM tested in the unsupervised segmentation of radar images [24, 25].

To be more precise, the results obtained allow us to put forward the following:

4.2. *Comparison between Three Mixture Estimation Algorithms and Their Adaptive Versions*

The calculation of the mean errors from all data available for three algorithms gives

EM	AEM	ICE	AICE	SEM	ASEM	τ
30.08	20.23	29.59	20.98	30.04	20.08	25.74

If we take into account data available for ICE, AICE, SEM, and ASEM, we obtain

ICE	AICE	SEM	ASEM
31.08	22.00	31.43	21.17

Thus we can say that the behavior of EM, ICE, and SEM are comparable, and similarly for AEM, AICE, and ASEM. Roughly speaking, the choice of EM, ICE, or SEM has little importance in the context studied. The estimation step does not strongly degrade the theoretical error and the use of adaptive versions strongly improves it, which is very interesting. Furthermore, the adaptive versions are significantly more efficient that the classical ones. However, this advantage can vary strongly with image characteristics. We have seen above situations where it is very important. According to Fig. 3 it is less important in the case of nonhomogeneous images and correlated noise.

4.3. *Choice between Blind and Contextual Methods*

For ICE, AICE, SEM, and ASEM, we can separate mean errors between their blind (BICE, BAICE, BSEM, BASEM) and contextual (CICE, CAICE, CSEM, CASEM) versions. The results obtained are

BICE	CICE	BAICE	CAICE	BSEM	CSEM	BASEM	CASEM
30.09	32.07	21.66	22.34	30.45	32.41	20.84	21.50

We can see that in our context blind methods are preferable to the contextual ones. This is rather surprising from a theoretical point of view: the theoretical Bayesian error is smaller in the contextual case than in the blind one. However, this can be explained by two factors. First, we use just one neighbor as a context. Thus the small improvement of the theoretical rate can be offset by the greater difficulties in the parameter estimation step. Second, the compatibility of the model at the level of priors stationarity can be worse in the contextual case than in the blind one. Let us note that this conclusion remains valid only for the situations studied: when the noise is less dominant or the context used larger, results can be significantly different [18].

4.4. *Influence of the Noise Correlation*

According to Fig. 3, we note that the noise correlation strongly degrades the efficiency of adaptive segmentation methods. The mean errors computed from Fig. 3 are

	ICE	AICE	SEM	ASEM
WN	29.25	15.37	29.11	13.72
CN	32.92	28.65	33.74	28.63

We can see that adaptive methods are very effective with white noise, and remain of interest with correlated noise.

4.5. *Influence of MD and VD Aspects*

As above, we compute different mean errors in the MD and VD cases:

	ICE	AICE	SEM	ASEM
MD	31.40	23.50	31.92	22.05
VD	30.77	20.51	30.93	20.30

We observe that the improvement in results obtained with adaptive algorithms is independent of the MD or VD aspects of noise. On the other hand, the methods remain stable with respect to an MD or VD nature of the noise.

4.6. *Influence of Class Image Homogeneity*

Calculation of mean errors gives

	ICE	AICE	SEM	ASEM
Ring	30.34	21.74	30.30	20.36
Gibbs	33.77	28.35	34.24	27.80
Letter B	29.14	15.94	29.73	15.36

We note that homogeneity plays an important role, especially in the behavior of the adaptive estimation algorithms-based methods.

4.7. *Examples*

We give in Fig. 4 some examples of visual aspects of normal and adaptive unsupervised segmentations.

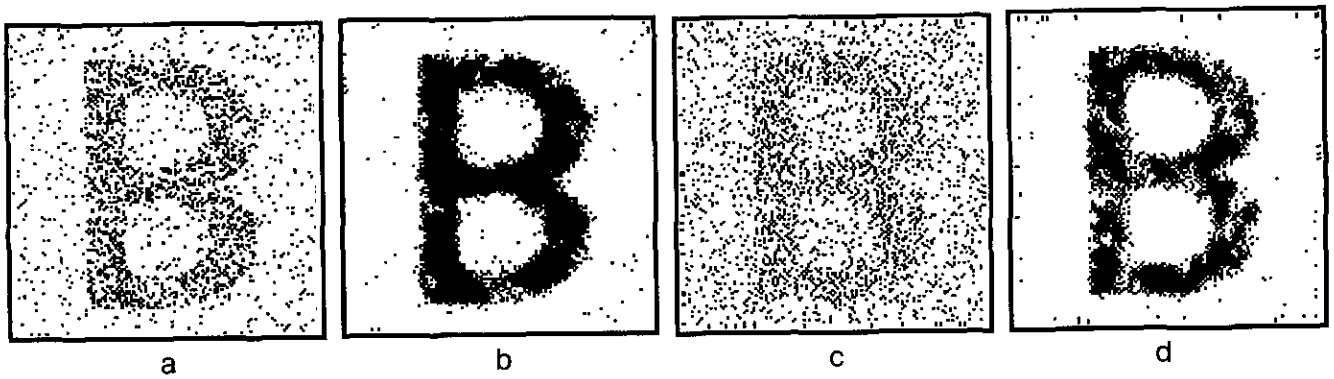


FIG. 4. "Letter B" + WVD (Fig. 2b). (a) segmented with blind EM and (b) segmented with blind AEM. "Letter B" + WMD (Fig. 2a). (c) Segmented with contextual EM, and (d) segmented with contextual AEM.

We can observe a real visual improvement of results when applying an adaptive version of EM.

5. CONCLUSIONS

We have presented in this paper some results of a simulation study of the behavior of different unsupervised local Bayesian segmentation methods in different situations. Situations differed in the homogeneity of images, independence or correlation of the noise, and its means discriminating or variances discriminating nature. According to the results obtained, all these factors have some influence on the behavior of the different methods studied. Unsupervised segmentation methods were distinguished only by the estimation step. Using Gaussian mixture estimation, we have studied three different algorithms EM, ICE, and SEM along with their adaptive versions AEM, AICE, and ASEM, where AEM and AICE are new. The results obtained allow us to put forth the following four conclusions:

1. Choice of EM, ICE, or SEM has little importance;

2. Adaptive versions of EM, ICE, or SEM should be used;

3. When the context used is reduced to one neighbor, blind methods are preferable to contextual ones;

4. The methods considered remain stable when image characteristics change.

In spite of the first conclusion, let us note that EM, ICE, and SEM are three different general mixture estimation methods, and their effectiveness can differ in other situations. The equivalence of their behavior here may be due to the robustness of the segmentation methods considered: although the estimated parameters are different, their differences are too small to influence the effectiveness of the segmentation method used. Some robustness curves we present in [20] tend to confirm such an explanation. The second conclusion confirms the results presented in [24, 25]. The problem of automatically choosing the best window W_j remains open. Results exposed in [24, 25], though, show that the adaptive SEM-based unsupervised segmentation method is relatively "robust" with respect to the window size considered. The third conclusion should be interpreted with care. As we showed in [18], in many situations the use of larger context, containing four nearest neighbors, can be beneficial. Furthermore, as stated in Section 4.1, we studied only the limit cases and we cannot say under which circumstances regarding the signal-to-noise ratio this conclusion remains valid. The fourth conclusion means that different characteristics of images do not disturb significantly the effectiveness of unsupervised methods. More precisely, the rates of Fig. 3 follow the theoretical rates of Table 2. This confirms results of some of our previous work [3, 16]. The general behavior of local methods is different from that of global methods which are extremely effective in certain situatons, but can also give bad results when images are nonhomogeneous and the noise is correlated.

Finally, the general conclusion is that when the noise is pronounced we have to use blind adaptive unsupervised segmentation methods and the choice of mixture estimation algorithms EM, ICE, or SEM has little importance.

These results give a partial answer to a general question: which methods for which images? A good answer to this question will undoubtedly take greater importance in future automated image processing systems.

APPENDIX

Real Images Segmentation Results

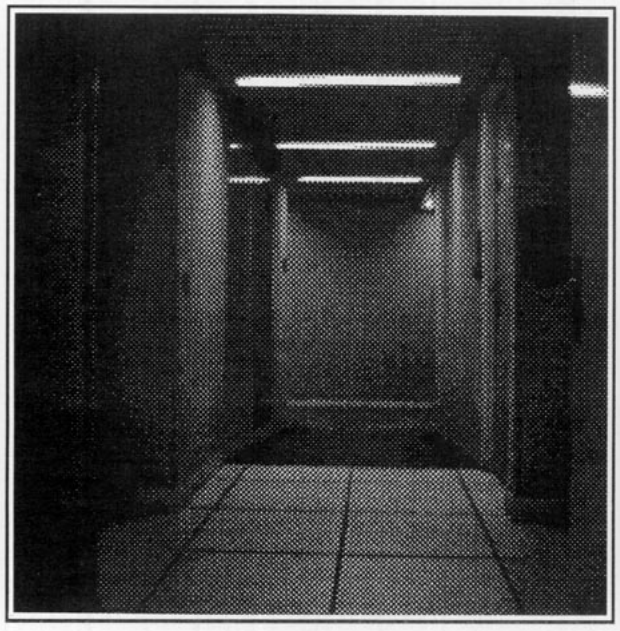


Image A

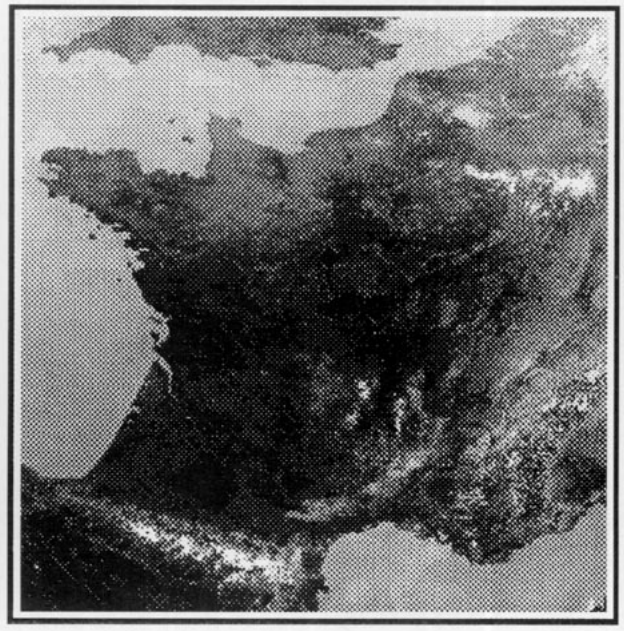


Image B

Image C

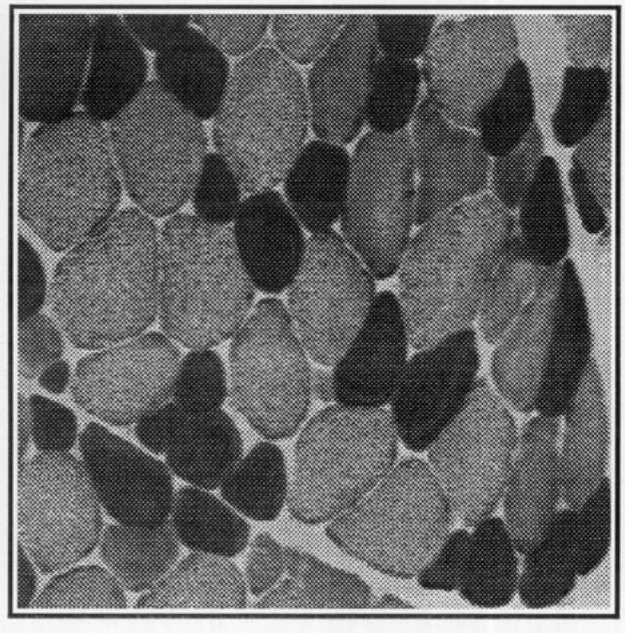
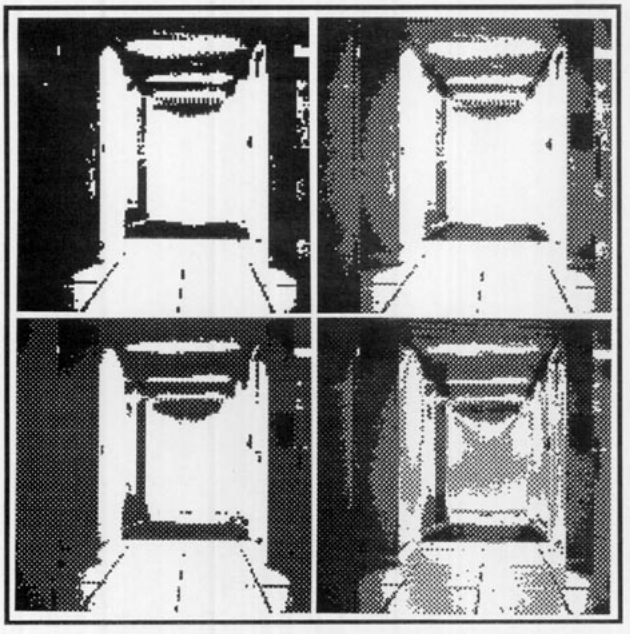
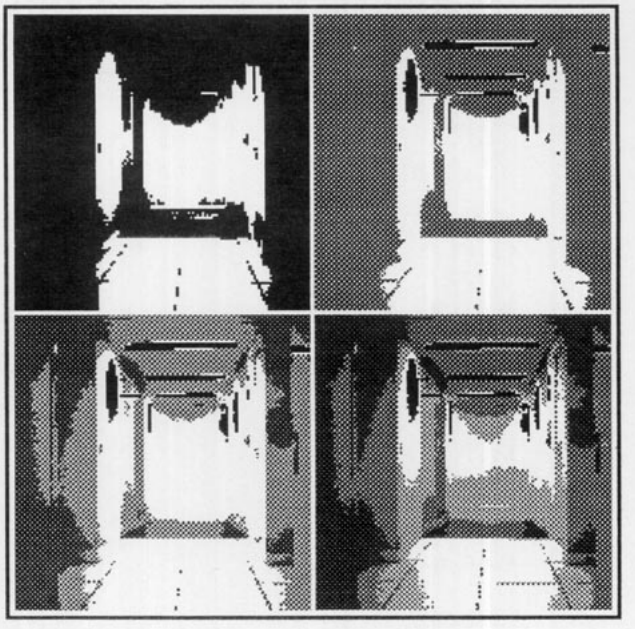


Image D



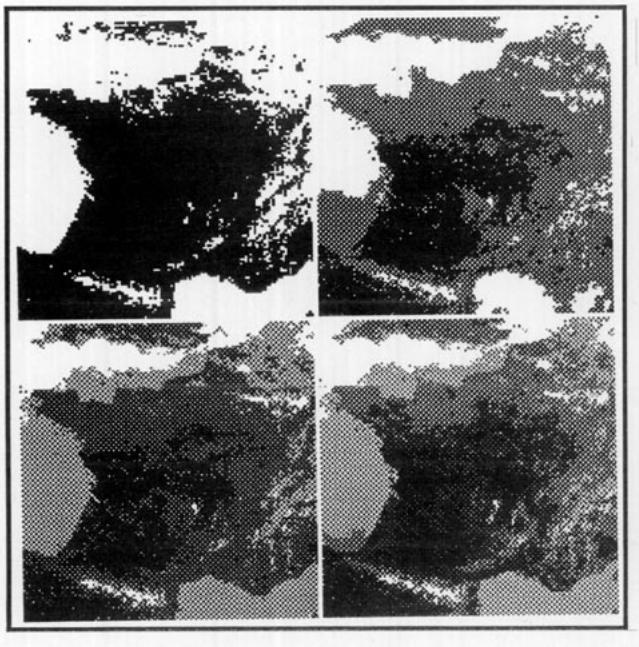
2 classes	3 classes
4 classes	5 classes

Segmentation of A with Blind ICE



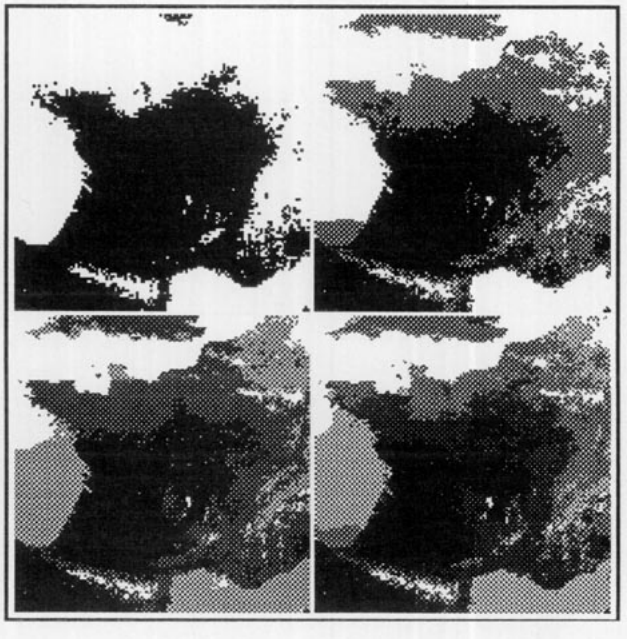
2 classes	3 classes
4 classes	5 classes

Segmentation of A with Blind AICE



2 classes	3 classes
4 classes	5 classes

Segmentation of B with Blind ICE

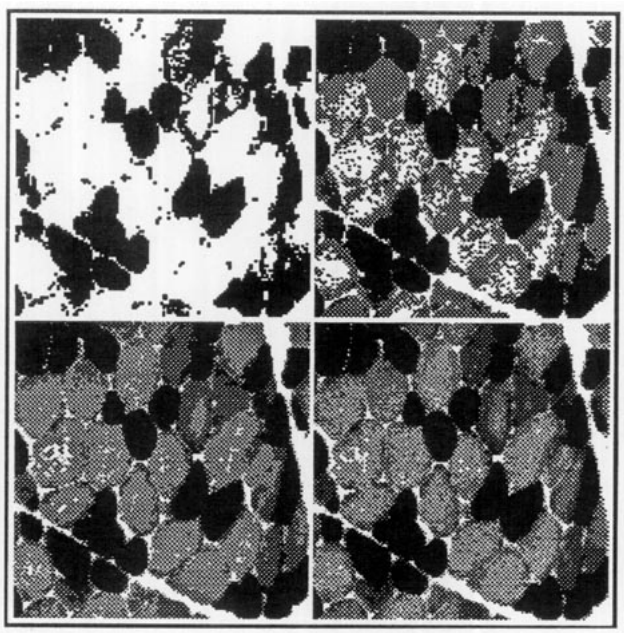


2 classes	3 classes
4 classes	5 classes

Segmentation of B with Blind AICE

2 classes	3 classes
4 classes	5 classes

Segmentation of C with Blind EM



2 classes	3 classes
4 classes	5 classes

Segmentation of D with Blind EM

REFERENCES

1. C. O. Acuna, Texture modeling using Gibbs distribution, *CVGIP: Graphical Models Image Process.* **54,** 1992, 210–222.
2. J. Besag, On the statistical analysis of dirty pictures, *J. R. Statist. Soc. B* **48,** 1986, 259–302.
3. B. Braathen, W. Pieczynski, and P. Masson, Global and local methods of unsupervised Bayesian segmentation of images, *Mach. Graphics Vision* **2**(1) 1993, 39–52.
4. G. Celeux and J. Diebolt, L'algorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités, *Rev. Statist. Appl.* **34**(2), 1986.
5. B. Chalmond, An iterative Gibbsian technique for reconstruction of m-ary images, *Pattern Recognit.* **22**(6) 1989, 747–761.
6. R. Chellapa and R. L. Kashyap, Digital image restoration using spatial interaction model, *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-30**(3), 1982, 461–472.
7. R. Chellapa and A. Jain (Eds.), *Markov Random Fields,* Academic Press, San Diego, 1993.
8. H. Derin and H. Elliot, Modeling and segmentation of noisy and textured images using Gibbs random fields, *IEEE Trans. Pattern Anal. Mach. Intell.* **9,** 1987, 39–55.
9. M. M. Dempster, N. M. Laird, and D. B. Rubain, Maximum likelihood from incomplete data via the EM algorithm, *J. R. Statist. Soc. B* **39,** 1977, 1–38.
10. R. C. Dubes and A. K. Jain, Random field models in image analysis, *J. Appl. Statist.* **16**(2), 1989.
11. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**(6) 1984, 721–741.
12. R. Haralick and J. Hyonam, A context classifier, *IEEE Trans. GRS* **GE-24,** 1986, 997–1007.
13. R. L. Kashyap and R. Chellapa, Estimation and choice of neighbors in spatial-interaction models of images, *IEEE Trans. Inform. Theory* **29,** 1983, 60–72.
14. P. A. Kelly, H. Derin, and K. D. Hartt, Adaptive segmentation of speckled images using a hierarchical random field model, *IEEE Trans. Acoust. Speech Signal Process.* **36**(10), 1988, 1628–1641.
15. S. Lakshmanan and H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing, *IEEE Trans Pattern Anal. Mach. Intell.* **11**(8), 1989.
16. N. Marhic, P. Masson, and W. Pieczynski, Mélange de lois et segmentation non supervisée des données SPOT, *Stat. Anal. Donnéer* **16**(2), 1991, 59–79.
17. J. L. Marroquin, S. Mittle, and T. Poggio, Probabiliste solution of ill-posed problems in computational vision, *J. Am. Statist. Assoc.* **82,** 1987, 76–89.
18. P. Masson and W. Pieczynski, SEM algorithm and unsupervised statistical segmentation of satellite images, *IEEE Trans. GRS* **34**(3), 1993, 618–633.
19. E. Mohn, N. Hjort, and G. Storvic, A simulation study of some contextual classification methods for remotely sensed data, *IEEE Trans. GRS* **GE 25,** 1987, 796–804.
20. A. Peng, Segmentation statistique non supervisée d'images et détection de contours par filtrage, thesis, Université de Technologie de Compiègne, 1992.
21. W. Pieczynski, Estimation of context in random fields, *J. Appl. Statist.* **16**(2), 1989, 283–290.
22. W. Pieczynski, Statistical image segmentation, *Mach. Graphics Vision* **1**(1/2), 1992, 261–268.
23. W. Qian and D. M. Titterington, On the use of Gibbs Markov chain

models in the analysis of images based on second-order pairwise interactive distributions, *J. Appl. Statist.* **16**(2), 1989, 267–281.

24. H. C. Quelle, J. M. Boucher, and W. Pieczynski, Local parameter estimation and unsupervised segmentation of SAR images, in *Proceedings of IGARSS. Houston, TX, May 1992.*
25. H. C. Quelle, Segmentation Bayésienne non Supervisée en Imagerie Radar, thesis, Université de Rennes 1, 1993.
26. R. A. Redner and H. F. Walker, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Rev.* **26,** 1984, 195–239.
27. A. Rosenfeld (Ed.), Image Modeling, *Academic Press,* New York, 1981.
28. J. Tilton, S. Vardeman, and P. Swain, Estimation of context for statistical classification of multispectral image data, *IEEE Trans. GRS,* **GE-20,** 1982, 445–452.
29. A. Veijanen, A simulation-based estimator for hidden Markov random fields, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(8), 1991, 825–830.
30. C. S. Won and H. Derin, Unsupervised segmentation of noisy and textured images using Markov random fields, *CVGIP: Graphical Models Image Process.* **54**(4), 1992.
31. L. Younes, Parametric inference for imperfectly observed Gibbsian fields, *Probab. Theory Relat. Fields* **82,** 1989, 625–645.