GLOBAL AND LOCAL METHODS OF UNSUPERVISED BAYESIAN SEGMENTATION OF IMAGES

Bjørn Braathen
Nordlysobservatoriet, University of Tromsø
N-9037 Tromsø, Norway
Wojciech Pieczyński
INT DSR, 9 Rue C. Fourier, 91011 Evry Cedex, France
Pascale Masson
Ecole Nationale Supérieure
des Télécommunications de Bretagne
BP 832, 29285 Brest Cedex, France.

Abstract. This paper deals with unsupervised Bayesian image segmentation and we focuse on the following two issues. First we introduce a new global method where the segmentation step is performed by MAP or MPM whereas at the parameter estimation step we apply the recently published ICE. We also show, in the case of hidden Markov fields, that ICE can be applied in a quite general frame. Then we compare the global unsupervised segmentation method so obtained with a contextual one, whose principle is radically different. A simulation study shows that the behaviour of the two methods considered strongly depends on the homogeneity of the images and the spatial correlation of the noise. Moreover, their efficiency is good in complementary situations. This leads us to propose a generalised unsupervised segmentation method containing a "choice algorithm" where the choice of the better suited method is automatically performed based on parameters estimated from the single image to be segmented. The unsupervised segmentation is then performed by the method chosen in the first step. Such a generalisation allows us to avoid the worst case segmentations and can be integrated in more complex automatical image processing systems.

1. Introduction

The aim of this article is two-fold: to present a new algorithm of unsupervised Bayesian image segmentation and to compare the efficiency of this new method with the efficiency of a fundamentally different method when some specific properties of the images change. In fact, the two methods are different in that the segmentation phases are respectively global (Maximum a Posteriori Probability – MAP, or Maximizer of Posterior Marginals –MPM) and local (contextual). The novelty of our algorithm lies at the parameter estimation level where Iterative Conditional Estimation (ICE), [17, 18], is applied. Our key result is that parameters such as image homogeneity or noise correlation can strongly alter the efficiency of a given unsupervised Bayesian segmentation method. We precise

Machine GRAPHICS and VISION vol. 2, no. 1, 1993, pp. 39–52

this dependency and propose an "unsupervised choice algorithm" which aim is to choose automatically, i.e. from the single image to be segmented, the better suited unsupervised segmentation method.

An analogous comparative study between global and local methods has been proposed by J.F. Yao, [9]. The important difference in comparison to our work is that the [9] does not tackle the parameter estimation problem and is only interested in the behaviour of the segmentation step. However, he proposes a study of the evolution of the efficiency of different methods when the value of the parameters used move from the true ones. The current study clearifies the importance of the estimation step and can be seen as an important step in the understanding of the behaviour of unsupervised segmentation methods.

The paper starts with a short description of the statistical modelling of images by random fields and the difference between the two segmentation methods considered is pointed out. Our method, ICE, is fairly unknown due to the fact that it has just appeared in a paper [17, 18] and was briefly presented at the GRETSI conference, [15], and we therefore present its description in the third section of this paper with appropriate detail. In particular we underline the frame of applicability of the ICE and show that it is much more general than the frame of applicability of the EM based estimation methods. ICE needs an estimator of priors from the unobserved field. The ICE based method we propose makes use of the method described by Derin et al., [7], which is exposed in the fourth section. We have chosen the Stochastic Estimation Maximisation (SEM), which is a stochastic variant of the EM algorithm, [2], for the treatment of the parameter estimation step in the second, contextual, unsupervised segmentation method. Our choice is intentional, in fact, some previous studies show the good behaviour of the SEM in this context, [14, 20]. A short description of the SEM makes up the fifth section. The numerical results take place in the sixth section whereas the seventh section is devoted to the concluding remarks, the principle of the new "unsupervised choice algorithm" that we propose and the corresponding "generalised unsupervised segmentation".

2. Random fields and statistical segmentation

We will consider a couple of random fields $(\xi, X) = (\xi_{ij}, X_{ij})$, $ij \in I$ where I is the set of pixels. Each ξ_{ij} takes its values in a finite set $\Omega = \{\omega_m\}, m = 1, 2$ and X_{ij} in $\mathcal{R}$, and we will assume that the distribution of (ξ, X) is defined by, firstly, $P(\xi)$, the distribution of ξ and, secondly, the family $P(X \mid \epsilon)$ of distributions of X conditioned on $\xi = \epsilon$, and we assume that the distributions are of known form. The field X is the only observable and the problem is that of estimating two parameters (α, β) which define $P(\xi)$ and $P(X \mid \epsilon)$ respectively, and, when these parameters have been estimated, that of segmenting the image based on these estimates. This segmentation will be an estimate

of the unobservable variable ξ based on the observable X. Using a Bayesian model with a $0-1$ loss function, and by letting $\xi_{\mathcal{A}}$ and $X_{\mathcal{B}}$ be restrictions of respectively ξ and X, $\mathcal{A} \subset I$, $\mathcal{B} \subset I$, ξ_A is estimated by $\hat{\xi_A}$ whose "a posteriori", i.e. conditional to X_B, probability is maximal. There are several possible methods of solving the problem. They may be summarised as follows:

- MAP: $\mathcal{A} = \mathcal{B} = I$,
- MPM: $\mathcal{A} = \{ij\}$, $\mathcal{B} = I$,
- Blind: $\mathcal{A} = \{ij\}$, $\mathcal{B} = \{ij\}$,
- Contextual: $\mathcal{A} = \{ij\}$, $\mathcal{B} = \mathcal{N}(ij)$,

where $\mathcal{N}$ denotes some neighborhood of pixel ij. We may say that the two first methods are global, while the two latter are local. The field ξ is supposed stationary and Markovian with $P[\xi = \epsilon] \neq 0$ for each ϵ and, thus, the distribution $P(\xi)$ is Gibbsian,

$$P(\xi = \epsilon) = \mathcal{K} \exp^{-\mathcal{U}_\alpha(\epsilon)},$$

and the distribution $P(X \mid \xi)$, that in fact describes the noise added, will be supposed Gaussian, so marginal conditional distributions are

$$P(X_{ij} = x_{ij} \mid \xi_{ij} = \omega_m) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[(x_{ij} - \mu_m)^2 / 2\sigma_m^2\right],$$

where $\mu = (\mu_1, \mu_2)$ and $\sigma = (\sigma_1, \sigma_2)$ are the parameter β. When considering contextual methods β also includes spatial correlations. The function $\mathcal{U}_\alpha(\epsilon)$ is a sum over the potentials associated with the cliques, [5, 7], in the neighbourhood system. We will consider noise of two different kinds that will profoundly influence the performance of the methods considered. The observed field X will be regarded as a sum of two fields, $X_{ij} = f(\xi_{ij}) + N_{ij}$ where the function f serves solely to distingush the two classes in the treatment and the random variables N_{ij} are respectively independent and correlated.

3. Iterative conditional estimation

The use of MAP or MPM, which constitutes the segmentation phase in our method, suppose the independence, conditional to ξ, of the random variables X_{ij}. So β defined above defines all distributions $P(X \mid \xi)$. Finally, knowing (α, β) is sufficient to apply MAP or MPM. We can estimate them using ICE which necessitates that we have at our disposal two estimators, namely $\hat{\alpha} = \hat{\alpha}(\xi)$ and $\hat{\beta} = \hat{\beta}(\xi, X)$, defined from ξ and X. The direct use of $\hat{\alpha}$ and $\hat{\beta}$ being impossible since we ignore realizations of ξ, the idea of ICE consists of approaching them by a function of X, the only observable. The best approximation, as far as the least square error is concerned, is the conditional expectation. Let $u(X)$ be any function of X. Then it is well known that the least square error

$$\mathcal{S}(X) = E[(\hat{\alpha} - u(X))^2]$$

is minimized by choosing $u(X) = E[\hat{\alpha} \mid X]$, [3]. This implies, in fact, that we will have to condition $\hat{\alpha}$ by X. In doing this we loose the quality of $\hat{\alpha}$ as an estimator. Although $\hat{\alpha}$ is independent of the parameters α and β, we have that $E[\hat{\alpha} \mid X]$, which for each $X = x$ is the expectation of $\hat{\alpha}$ according to the distribution of ξ conditioned by $X = x$, depends on the parameters α and β. The same goes for $E[\hat{\beta} \mid X]$.

Hence, if we want to compute $E[\hat{\alpha} \mid X]$ and $E[\hat{\beta} \mid X]$ we are obliged to make use of two previously obtained parameters α_n and β_n.

Iterative Conditional Estimation (ICE) is defined as follows, [17, 18]:

- α_0 and β_0, i.e. initial values of α and β are chosen,
- α_{n+1} and β_{n+1} are computed from α_n and β_n and $X = x$ according to

$$\alpha_{n+1} = E_n[\hat{\alpha} \mid X = x], \tag{1}$$

$$\beta_{n+1} = E_n[\hat{\beta} \mid X = x], \tag{2}$$

with E_n denoting the expectation at iteration n using α_n and β_n. Let $P_n(\xi \mid X)$ be the distribution of ξ conditioned by X computed from α_n and β_n, then 1 and 2 may be written:

$$\alpha_{n+1} = \sum_{\epsilon} \hat{\alpha}(\epsilon) P_n(\epsilon \mid X = x), \tag{3}$$

$$\beta_{n+1} = \sum_{\epsilon} \hat{\beta}(\epsilon, x) P_n(\epsilon \mid X = x). \tag{4}$$

The computation of these equations may be impossible in practice, but we can approach 3 and 4, in accordance with the law of large numbers by:

$$\alpha_{n+1} = \frac{1}{K} \sum_{k=1}^{K} \hat{\alpha}(\epsilon_k), \tag{5}$$

$$\beta_{n+1} = \frac{1}{K} \sum_{k=1}^{K} \hat{\beta}(\epsilon_k, x), \tag{6}$$

where ϵ_k, $k = 1, \ldots, K$ are independent realizations of ξ according to the distribution $P_n(\xi \mid x)$.

Finally, we conclude this section by noting that the use of the ICE depends on the availability of the following:

1. two estimators, $\hat{\alpha} = \hat{\alpha}(\xi)$ of α and $\hat{\beta} = \hat{\beta}(\xi, x)$ of β,
2. two initial values α_0 and β_0, not "too far" from the true parameters α and β,
3. the obtainability of $E_n(\hat{\alpha} \mid X)$ and $E_n(\hat{\beta} \mid X)$ or simulated realizations of ξ according to the posterior distribution $P_n(\xi \mid X)$.

The principle of ICE is quite different from the well-known EM-algorithm– in particular, the notion of likelihood is not used and Dirac distributions can be considered. To be more precise let us denote by $\varphi_{\omega_1}, \varphi_{\omega_2}$ the distribution densities of each X_{ij} conditional

to $\xi_{ij} = \omega_1, \omega_2$, respectively. The sampling of ξ according to the posterior distribution $P_n[\xi|X]$, and thus the application of formulae 5 and 6, are possible for every $\varphi_{\omega_1}, \varphi_{\omega_2}$. In fact, the energy of the posterior distribution is

$$U^{X=x}(\epsilon) = U(\epsilon) + \sum_{ij} ln\varphi_{\epsilon_{ij}}(x_{ij}),$$

and the Gibbs sampler can be used as soon as the $\varphi_{\epsilon_{ij}}(x_{ij})$ are known. In particular, $\varphi_{\omega_1}, \varphi_{\omega_2}$ can be of different kinds and one of them, or both, can be of a more limited kind. This kind of situations may occur in the practical segmentation problems, in particular, the distribution of the speckle in radar images is not known exactly and may vary with the class.

Thus the use of ICE is not limited by the form of the conditional distribution of the noise and the EM principle, based on the maximisation of the likelihood, turns out to be powerless in such a general context.

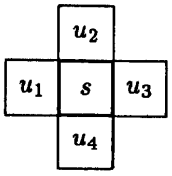
Let us note that in spite of the fairly corresponding design the ICE method is quite different from the ICM of J. Besag, [5]. ICE is a parameter estimation algorithm while ICM occupies itself with segmentation. It is then possible to perform a non-supervised segmentation by using ICE for the parameter estimation step and ICM, based on the estimated parameters, for the final segmentation step.

In the simple case of a Gaussian mixture the EM-reestimating formulae is obtained from (1) and (2), [17, 18].

4. Gibbs distribution parameter estimation

The problem of estimating the parameters of the Gibbs energy function has recently received some attention, [11]. The coding method, [1], has been used extensively in the literature, but requires the solution of a set of non-linear equations. Derin & Elliott, [7], proposed in 1987 a novel estimator that consists of simple histogramming and standard, linear, least-squares estimation. See also [13] for an interesting stochastic gradient method.

We will briefly describe the estimator in terms of our model, i.e. the four nearest neighbourhood system, [7]. The figure to the right shows a pixel belonging to class s with the 4 nearest neighbours defined $t = \{u_1, u_2, u_3, u_4\}$.



Two indicator functions $\mathcal{I}$ and $\mathcal{J}$ are defined,

$$\mathcal{I}(y_1, \ldots, y_M) = \begin{cases} -1 & \text{if } y_1 = \cdots = y_M \\ 1 & \text{otherwise} \end{cases}$$

M being the number of classes (=2 in this work)

and

$$\mathcal{J}_k(s) = \begin{cases} 1 & \text{if } s = \omega_k \\ 0 & \text{otherwise} \end{cases} .$$

Define

$$\phi(s,t) = \begin{bmatrix} \mathcal{J}_1(s) \\ \mathcal{J}_2(s) \\ \mathcal{I}(s,u_1) + \mathcal{I}(s,u_3) \\ \mathcal{I}(s,u_2) + \mathcal{I}(s,u_4) \end{bmatrix} ,$$

and let

$$V(s,t',\alpha) = \phi^T(s,t') \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} .$$

By manipulating the distributions of the different possible configurations we obtain the following expression for two different values of s but identical neighbourhood t':

$$e^{-V(j,t',\alpha)+V(k,t',\alpha)} = \frac{P(j,t')}{P(k,t')} ,$$

which leads to

$$(\phi(k,t') - \phi(j,t'))^T \alpha = \ln\left(\frac{P(j,t')}{P(k,t')}\right) , \tag{7}$$

where $P(s,t)$ is the joint distribution of s and t. As α is the unknown parameter vector to be estimated and the RHS of 7 may be estimated using simple histogramming, 7 reduces to a set of linear equations. As the set of linear equations is overdetermined the problem of choosing the equations that gives the best estimate poses a problem. In addition, attention must be payed to the RHS of 7.

We have chosen to make use of all the equations obtained, i.e. the equations that correspond to configurations that do exist in the field, and use the mean of the estimates for each parameter and, in accordance with the observation in [7], we have had no problem whatsoever.

5. Stochastic estimation and maximisation

Our method is compared with a contextual method, based on the four nearest neighbours, in which the useful parameters are estimated by the SEM, [6, 14, 20]. We note that the contextual method allows taking into account the spatial correlation of the noise, which is impossible when considering a global method. The SEM algorithm may be regarded as an EM one with a probabilistic learning stage, and is briefly described, as follows:

Initialization. The number (initial) of components K is set and a threshold c is chosen with value beetween 0 and 1. The initial probabilities for each class k, i.e.

$P^0(\xi_{ij} = \omega_k \mid X_{ij} = x_{ij})$, are randomly chosen satisfying the two conditions

$$0 < P^0(\omega_k \mid x_{ij}) < 1,$$
$$\sum_k P^0(\omega_k \mid x_{ij}) = 1.$$

Stage S (stochastic). At each iteration n the chosen sample x is partitioned, based on $P(\omega_k \mid x_{ij})$, into K groups. The number of elements in each partition P_k^n is then the number of pixels assigned to class k. If, for a certain k and with $c_n^k = \text{card}(P_k^n)$, c_n^k is smaller than $c \times \text{card}(I)$, the algorithm is re-initialized with $K - 1$ components.

Stage M (maximisation). The maximum likelihood estimates of the parameters are computed, in the case of the mean of a distribution we get

$$\mu_k^{n+1} = \frac{1}{c_n^k} \sum P_k^n,$$

where the sum goes over all the elements contained in P_k^n. Variances and a priori probabilities are eventually estimated in the same manner.

Stage E (estimation). $P^{n+1}(\omega_k \mid x_{ij})$ is then computed based on the parameters obtained in the former stage.

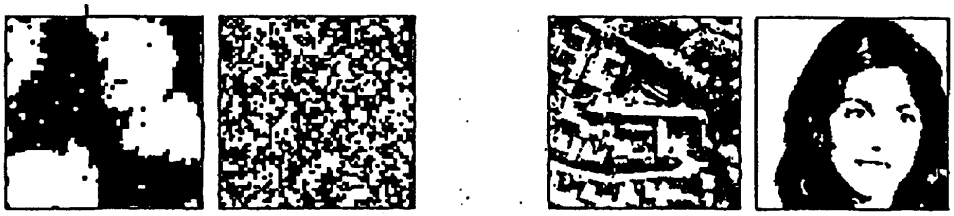
Note that our presentation of the SEM is as it has been presented in the literature, but it can also be seen as a particular case of the ICE procedure, [17, 18].

6. Experimental results

In order to test the algorithms we have created a range of different images. Of these the two that in the best way display the differences from method to method have been chosen together with two "natural" images. We are solely discussing the results of the segmentation, not the performance of the parameter estimation methods. This is not less interesting or important, but demands an independent treatment that surpasses the intention of this work.

6.1. Image data

The simulations are based on four different images of size 64 × 64 which we will call $\mathcal{IM}_1$,$\mathcal{IM}_2$, $\mathcal{IM}_3$ and $\mathcal{IM}_4$ (fig. 1). The two first were created by the method of Ising that allows us to simulate binary fields (images) with two parameters specified that control the degree of intra-class and inter-class attraction and repulsion. The following parameters were used: •$\mathcal{IM}_1$: $\theta_1 = 5.0$, $\theta_2 = -2.5$ •$\mathcal{IM}_2$: $\theta_1 = 2.0$, $\theta_2 = -1.0$. The last two images are "natural", although binarized in such a way that the number of pixels belonging to the value 0 equals that of the other. Gaussian noise of two different kinds were then



(a). The two synthetic images (b). The two "natural" images

Fig. 1. The four binary images considered

added to all four images and a total of eight images were obtained. The different binary images are shown in fig. 1.

We see that $\mathcal{IM}_1$ is an image with rather large grouping of pixels belonging to the same class whereas $\mathcal{IM}_2$ has a considerably more spotty appearance.

The means and the variances of the different images are all equal: $\bullet m_1$: 1.0 m_2: 3.0 σ^2: 1.0, and in the case of dependent noise the intra-class correlation is 0.72 and the inter-class correlation 0.64. Then the images were stretched and digitized with values in the range $[0, 255]$ in order to facilitate the computation. As the images were stretched based on the minimum and maximum values within the images, the corresponding mean values and variances differ (see tab. 1). Fig. 2 and 3 show the resulting images.

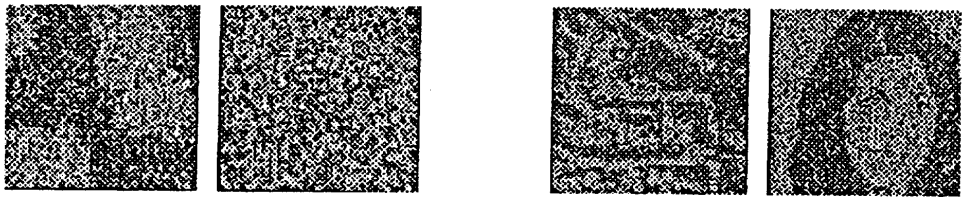


Fig. 2. Independent noise added

The ICE algorithm was used with the classical empirical estimator of mean and variance and with the estimator of Derin & Elliot, [7], using a 4-neighbourhood system, and was eventually combined with two different algorithms of segmentation, namely the MAP, [4], and the MPM, [8], whereas the SEM was applied with a 4-neighbourhood system.

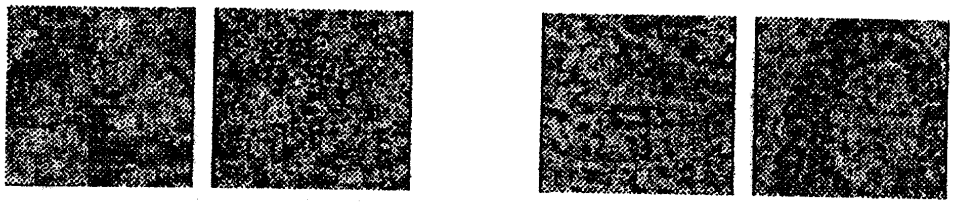


Fig. 3. Dependent noise added

6.2. Segmentation results

The MAP which is here approached by the simulated annealing algorithm, was in each case run with a large number of iterations (800) and the constant c was set to 3.5. The MPM was run with 80 iterations.

The measure of error used is the simple pixel to pixel error rate (root mean square). In addition we measured the error by extending the concept of rms to a 2×2 and a 3×3 window, but the only information of interest supplied by these measures were, as they did not change the final rating of methods, that the pixels wrongly classified were somewhat randomly distributed throughout the images or that they occured in (approximately) the same regions.

We see that the efficiency of both global and local methods strongly depend on two factors: the homogeneity and the spatial correlation of the noise. This point will be precised in the next section. Some examples of the resulting images are shown in figs. 4 and 5. The values in the column Init. are the initial segmentation results obtained in the first blind (not taking the neighbors into account) maximum likelihood classification, a classification that is performed in order to obtain a starting point for the parameter estimation. Note that some of the results obtained by the global methods are worse than this blind segmentation showing that in these cases the global methods fall through.

(a) SEM. (b) MPM

Fig. 4. A selection of the results, independent noise case

(a) SEM. (b) MPM

Fig. 5. A selection of the results, dependent noise case

7. Conclusion, choice algorithm and generalised unsupervised segmentation

As we noted in the introduction the purpose of this paper is two-fold. We first described how a new general method of estimation in the case of hidden data, called ICE, [17, 18], can provide a family of unsupervised statistical image segmentation methods. We showed its superiority, at the level of the generality of the models treated, over the EM-based ones. We showed in some previous works, [15], that in the classical case of the conditionally Gaussian noise the efficiency of our methods is comparable to the efficiency of some known algorithms, [10, 19]. Thus we adopted this frame here and considered that our methods are representative of the whole family of global methods. Then we considered a local unsupervised segmentation method in which the parameter estimation step is performed by SEM. Once more we considered that this algorithm is representative of the whole family of local methods. In fact, in the context considered the local methods seem to be relatively stable with respect to the parameter estimation algorithm used, [16]. Nevertheless, both hypotheses are approximations which stay valid in the case of the relatively simple images that we considered and undoubtedly the behaviour of different algorithms within each family is different when the images become more complex, in particular when the number of classes exceeds two.

However, our simulation study shows that the behaviour of the two methods is very different and they can give really bad results. This implies that the choice of a suitable method, even in the case of relatively simple binary images, is of the utmost importance. To be more precise, we considered four images: $\mathcal{IM}_1$,$\mathcal{IM}_2$, $\mathcal{IM}_3$ and $\mathcal{IM}_4$. Let us denote by H_i the homogeneity of $\mathcal{IM}_i$. We can say, according to the visual impression: $H_4 = H_1 > H_3 > H_2$. That is to say $\mathcal{IM}_4$ is as homogeneous as $\mathcal{IM}_1$, $\mathcal{IM}_1$ is more homogeneous than $\mathcal{IM}_3$ and so on. Let us denote by LOC_i the error rate of the local SEM based method applied to $\mathcal{IM}_i$ and by GLO_i the mean of the error rates of our two global ICE based methods. According to tab. 1 we will then have as shown in tab. 2.

		SEM	ICE		
		4 N	MAP	MPM	Init.
$\mathcal{IM}_1$	ind.	3.75	3.6	2.4	15.5
	dep.	13.4	9.2	11.5	16.0
$\mathcal{IM}_2$	ind.	13.7	16.9	14.0	15.5
	dep.	10.0	30.4	27.7	17.0
$\mathcal{IM}_3$	ind.	6.3	6.7	5.4	16.5
	dep.	13.0	15.6	15.1	16.5
$\mathcal{IM}_4$	ind.	3.2	2.0	1.4	15.4
	dep.	11.6	5.4	6.8	13.9

Tab. 1. Results of the different algorithms

		LOC	GLO
$\mathcal{IM}_4$	ind.	3.2	1.7
	dep.	11.6	6.1
$\mathcal{IM}_1$	ind.	3.8	3.0
	dep.	13.4	10.4
$\mathcal{IM}_3$	ind.	6.3	6.1
	dep.	13.0	15.4
$\mathcal{IM}_2$	ind.	13.7	15.5
	dep.	10.0	29.1

Tab. 2. Conclusive results

The results in the tabs. 1 and 2 leads us to the following conclusions:

1) Independent noise. The efficiency of both local and global methods is degraded when the homogeneity decreases. In fact we have $LOC_4 > LOC_1 > LOC_3 > LOC_2$, $GLO_4 > GLO_1 > GLO_3 > GLO_2$, where $>$ means "is more efficient than". Furthermore, $GLO_4 > LOC_4$ but $GLO_2 < LOC_2$: global methods are more sensitive to the homogeneity of the images.
2) Dependent noise. The same hierarchy is respected when the global methods are concerned: $GLO_4 > GLO_1 > GLO_3 > GLO_2$. Furthermore, when the image is slightly homogeneous and the noise strongly correlated we conclude that global methods are to be avoided absolutely. The behaviour of local methods is somewhat surprising and could turn out to be very interesting in practical applications. The degradation in efficiency, when the homogeneity decreases, is stopped and the tendency is even inversed. In fact, we have $LOC_4 > LOC_1 < LOC_3 < LOC_2$. This is surprising because the local method considered contains a parameter estimation step and, in a general manner, the efficiency of estimation procedures is degraded when the correlation of data increases. Now, as we pointed out in [20], SEM seems to be robust with respect to the noise correlation and the loss at the parameter estimation level is here less important than the benefit due to the better efficiency of the segmentation step.
3) With the exception the interesting case LOC_2, the correlation of the noise always degrades the efficiency of the methods considered.

This leads us to propose the following "choice algorithm":

i) if the image is homogeneous, use a global method, whatever the correlation of the noise,
ii) if the image is very little homogeneous, use a local method, whatever the correlation of the noise,

iii) if the image is moderately homogeneous use a local method if the noise is strongly correlated and a global one if it is next to independence.

Our choice algorithm is not very precise and in real situations it may be difficult to decide if a given noisy image is homogeneous or not. Now, the distribution of $(\xi_s, \xi_t, X_s X_t)$ for s,t neighbours, contains some objective information about the homogeneity and noise correlation. In fact, if an image is homogeneous the distribution of (ξ_s, ξ_t) charges $\{(0,0),(1,1)\}$, and if not, it charges $\{(0,1),(1,0)\}$. The noise correlations are given by the distributions of (X_s, X_t) conditional to (ξ_s, ξ_t). The distribution of $(\xi_s, \xi_t, X_s X_t)$ can be estimated from X by the SEM, which can constitute an early step in order to make the choice algorithm above unsupervised.

We propose the following:

a) estimate the distribution of $(\xi_s, \xi_t, X_s X_t)$, for s,t neighbours by the SEM,
b) decide in what case i), i), iii) of the choice algorithm you are,
c) perform the corresponding unsupervised segmentation.

This unsupervised choice algorithm would be entirely automatic if the decision b) is taken automatically from the data $X = x$. We can imagine that the homogeneity is given by the parameter $h = P[(\xi_s,\xi_t) = (0,0)] + P[(\xi_s,\xi_t) = (1,1)]$ and the degree of noise correlation by the coefficient of correlation ρ, both parameters h, ρ being estimated in step a). h and ρ take their values in $[0,1]$, thus make the step automatic means to know two regions GLO, LOC in $[0,1]^2$: if $(\hat{h},\hat{\rho}) \in GLO$ we use a global method and if $(\hat{h},\hat{\rho}) \in LOC$ we use a local one.

Our choice algorithm allows us to have a general idea about the curve C defining the two regions as in fig. 6a.

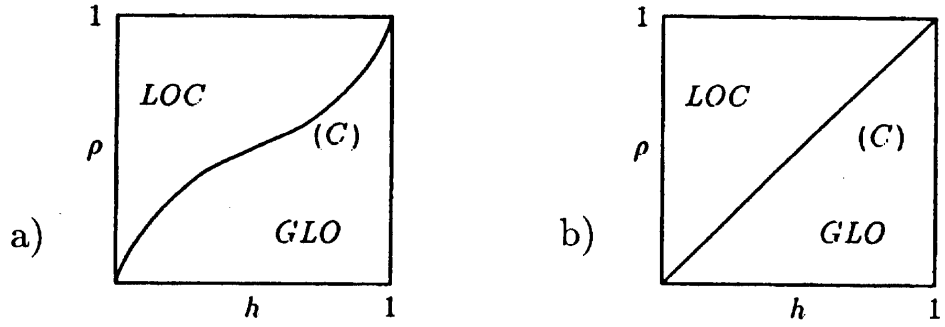


Fig. 6. Curve C defining two regions

The empirical knowledge of C with better precision should demand more complete simulations which we hope to perform in our future works. At this stage we can propose the shape of the curve C as in fig. 6b.

As a general conclusion we can say that the choice between a global and a local unsupervised statistical segmentation method can turn out to be crucial. When integrating the segmentation step in more complex automatical image processing systems the automatical nature, i.e. by only using $X = x$, of this choice is important. The processes we propose, based on C of a simple shape above, should allow to avoid the really bad segmentations.

References

1974

[1] Besag J.: Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc. B., 36.

1977

[2] Dempster M.M., Laird N.M., Rubin D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, B-39, 1–38.

1981

[3] Priestley M. B.: Spectral analysis and time series. AP.

1984

[4] Geman S., Geman D.: Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. IEEE Trans. PAMI, 6.

1986

[5] Besag J.: On the statistical analysis of dirty pictures. J. Roy. Statist. Soc., B-48, 259-302.

[6] Celeux G. and Diebolt J.: L'agorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités. Rev. Stat. Appli., 34 (2).

1987

[7] Derin H., Elliott H.: Modeling and segmentation of noisy and textured images using Gibbs random fields. IEEE Trans. PAMI, 9.

[8] Marroquin J., Mitter S., Poggio T.: Probabilistic solution of ill-posed problems in computer vision. J. ASA, 82, 397.

1988

[9] Yao J.F.: Méthodes bayesiennes en segmentation d'images. Bayesian Statistics, Proc. 8th Franco-Belgian Meeting of Statisticians (CORE, Université Catholique de Louvain, Belgique)

1989

[10] Chalmond B.: An iterative Gibbsian technique for reconstruction of m-ary images. PR, 22 (6), 747–761.

[11] Dubes R. C., Jain A. K.: Random field models in image analysis. Journal of Applied Statistics, 16 (2).

[12] Mardia K.V.: Markov models and Bayesian methods in image analysis. Journal of Applied Statistics, 16 (2).

[13] Younes L.: Parametric inference for imperfectly observed Gibbsian fields. Probability Theory and Related Fields, 82, 625–645.

1990

[14] Masson P., Pieczyński W.: Segmentation of SPOT images by contextual SEM. Proc. Eusipco, Barcelona, Spain.

1991

[15] Braathen B., Marhic N., Masson P., Pieczyński W.: Sur une nouvelle approche de segmentation bayesienne non supervisée d'images. Treizième Colloque GRETSI 91, Juans-les-Pins, France, September 91.

[16] Marhic N., Masson P., Pieczyński W.: Mélange de lois et segmentation non supervisée des données SPOT. Statistique et Analyse des Données., 16 (2).

1992

[17] Pieczyński W.: Parameter estimation in the case of hidden data. Proceedings of 16th Biennal symposium on Communications, Kingston, Canada.

[18] Pieczyński W.: Statistical image segmentation. Machine Graphics and Vision, 1 (1/2), 261–268 (Proc. of 2nd GKPO'92 Conference, Nałęczów, Poland).

[19] Devijver P. A.: Real time modelling of image sequences. Submitted to IEEE Trans. PAMI.

[20] Masson P., Pieczyński W. SEM algorithm and unsupervised segmentation of satellite images. To appear in IEEE Transactions on Geoscience and remote sensing.

Pascale Masson was born in France in 1963, received his Ph.D. degree in Electronics and Image Processing from the Université de Bretagne Occidentale in 1991. She was a visiting researcher at the University of Sherbrooke, Le Quebec, Canada, from 1991 to 1992. Her research iterest include statistical image processing and pattern recognition.

Bjorn Braathen was born in Norway in 1957, received his M.S. degree in Physics from the University of Tromso, Norway, in 1988. He was a visiting researcher at the Ecole Nationale Supérieure des Télécomunications de Bretagne in Brest, France, from 1990 to 1991. In 1992 he joined the Institute of Natural Sciences at the University of Tromso and is currently finishing his Ph.D. degree in Physics. His research interests include pattern recognition, image processing, signal analysis and control.

Wojciech Pieczyński was born in Poland in 1955, received the Doctorat d'Etat degree from Pierre et Marie Curie University, Paris, in 1986. He is currently Professor and Head of Group Image at the Institut National des Télécomunications, Evry, France. His research interests include mathematical statistic, stochastic processes and statistical image processing.