Hidden Markov Fields and Unsupervised Segmentation of Images

Olivier Allagnat * Jean-Marc Boucher * Dong-Chen He ** Wojciech Pieczynski ***

* Ecole Nationale Supérieure des Télécommunications de Bretagne
** Université de Sherbrooke
*** Institut National des Télécommunications

Abstract

This paper deals with unsupervised Bayesian segmentation of images. We introduce a new algorithm based on a recent general method of estimation in the case of incomplete data (iterative conditional estimation). The efficiency of our method is compared with a recent algorithm based on the stochastic gradient by L.Younes. We give results of numerous simulations and an application to a real radar image is also derived.

1. Introduction.

The statistical approaches of the important problem of image segmentation have been developed by several authors in recent years. The Bayesian method, which is generally the adopted one, demands the modelling by random fields: S being a set of pixels one considers two collections of random variables $\zeta=(\zeta_s)_{s\in S}$, $X=(X_s)_{s\in S}$ called "random fields". Each ζ_s takes its values in a finite set $\Omega=\{\omega_1, \omega_2,..., \omega_m\}$ of classes and X_s is in R^d, with $d\in N$. So the problem of segmentation is the problem of estimating an "unknown" realization of ζ from an "observed" realization of X. The distribution of (ζ,X) is defined by P_ζ, distribution of ζ, and the family P^ε_X of distributions of X conditional to $\zeta=\varepsilon$. In this work we are interested by the MAP segmentation, which will be performed by simulated annealing ([9]), so we adopt the following model:

The field ζ is assumed Markovian with all realizations possible; so its distribution is a Gibbs distribution:

$$P_\zeta[\varepsilon] = k\, e^{-U_\alpha(\varepsilon)} \qquad (1)$$

where k is unknown and α is a parameter which, once known, makes it possible to compute the distribution of each ζ_s conditional to $(\zeta_t)_{t\neq s}$. The set of parameters defining all conditional distributions P^ε_X will be denoted by β. On the other hand we will suppose that the random variables (X_s) are independent conditionally to each realization of ζ : so, if we have m classes it is sufficient to define m distributions of each X_s conditional to $\zeta_s=\omega_1,..., \omega_m$ respectively. Thus β will have m components $\beta_1,...., \beta_m$. In this paper the field X will be assumed Gaussian conditionally to ζ and each X_s real, so we have $\beta_i = (\mu_i, \sigma^2_i)$, where μ_i is the mean and σ^2_i the variance of the Gaussian distribution of each X_s conditional to $\zeta_s=\omega_i$.

When the parameter $\theta=(\alpha, \beta)$ is known simulated annealing ([9]) gives excellent results, there also exists some other efficient algorithms : ICM of Besag ([1]) or MPM of Marroquin and al. ([13]). When θ is not known the problem becomes much more difficult: in the absence of any information, which is the more general case, one has to estimate it from X. The statistical problem is that of "incomplete data": parameters are estimated from random variables whose distributions are a mixture of distributions.

Several solutions to this problem have been proposed in recent years. Some algorithms (Chalmond ([3]), Devijver and al. ([6]), Masson and Pieczynski ([14]), Qian and Titterington ([18])) use variations of the EM algorithm ([4]), adapted to the models considered. One alternative technique (Besag ([1]), Lakshmanan and Derin ([10])) consists of a re-estimation of the parameters based on the segmentations obtained with the current parameters.

In this work we present a new algorithm for estimating θ. We compare its efficiency with that of Younes's method ([20]). We examine the behaviour - by studing real and simulated images - of two unsupervised segmentation methods obtained from our estimation algorithm and from Younes's by adding simulated annealing ([9]).

The organization of the paper is as follows:

In the next section we present our new method and briefly review the operation of the Younes algorithm.

The third section is devoted to different simulations and contains some comments. Concluding remarks constitute the fourth and last section.

0-8186-2920-7/92 $3.00

2. Ice procedure and Younes algorithm.

2.1 Ice procedure.

ICE procedure is a recent general method of estimation in the case of incomplete data ([16], [17]), stemming from the following idea: if X were observable one could generally estimate α, β by some efficient estimators $\hat{\alpha}$, $\hat{\beta}$ defined from ζ, X. ζ being unknown we have to approach $\hat{\alpha}$, $\hat{\beta}$ by some functions of X, the only one observable. The best approximation, as far as the mean square error is concerned, is the conditional expectation. This in turn depends on the parameters, so we have to take a "current" value of the parameter defined by some way. This leads to the following iterative method: let $\hat{\alpha}$, $\hat{\beta}$ be two estimators of α, β defined from ζ, X. By denoting E_n the conditional expectation using (α_n, β_n) the procedure is:
(i) one takes an initial value (α_0, β_0)
(ii) $(\alpha_{n+1}, \beta_{n+1})$ is computed from (α_n, β_n) and X=x by:

$$\alpha_{n+1} = E_n [\, \hat{\alpha} / X = x \,] \qquad (2)$$

$$\beta_{n+1} = E_n [\, \hat{\beta} / X = x \,] \qquad (3)$$

Let us denote by P^x_n the distribution of ζ conditioned on X=x (posterior distribution) computed from (α_n, β_n). (2) and (3) are not workables, but it is possible to simulate realizations of ζ according to P^x_n, using the Gibbs sampler for example. So we can use an "approximative" ICE : (2) and (3) can be approached, in accordance with the law of large numbers, by:

$$\alpha_{n+1} = \frac{1}{N} \sum_{i=1}^{N} \hat{\alpha}(\varepsilon_i) \qquad (4)$$

$$\beta_{n+1} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}(\varepsilon_i, x) \qquad (5)$$

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ are independent realizations of ζ according to P^x_n.
Our method is an ICE procedure, and thus we have to choose an estimator of α from ζ and another one of β from ζ, X. We opt for $\hat{\alpha}=\hat{\alpha}(\zeta)$ another Younes algorithm ([19]) (defined from ζ) and $\hat{\beta}=\hat{\beta}(\zeta,X)$ given by empirical means and variances. The choice of $\hat{\alpha}$ is justified by its good behaviour; furthermore the author proposes a rigorous study of its asymptotical behaviour. More precisely $\hat{\alpha}$, which is an iterative method, runs as follows: let U_α denote the energy of the distribution of ζ. For each $\zeta=\varepsilon$ $U'_\alpha(\varepsilon)$ will denote the gradient of U_α with respect to α.

(i) one takes an initial value α_0
(ii) α_{n+1} is calculated from α_n and $\zeta=\varepsilon_0$ by:

$$\alpha_{n+1}=\alpha_n+(1/n+1)[U'_{\alpha_n}(\varepsilon_{n+1})-U'_{\alpha_n}(\varepsilon_0)] \qquad (6)$$

where ε_{n+1} is a realization of ζ simulated by Gibbs sampler based on the current parameter α_n.
Empirical means and variances are:

$$\hat{\mu}_i = \frac{\sum_{s \in S} X_s \, 1_{[\zeta_s = \omega_i]}}{\sum_{s \in S} 1_{[\zeta_s = \omega_i]}} \qquad (7)$$

$$\hat{\sigma}^2_i = \frac{\sum_{s \in S} (X_s - \hat{\mu}_i)^2 \, 1_{[\zeta_s = \omega_i]}}{\sum_{s \in S} 1_{[\zeta_s = \omega_i]}} \qquad (8)$$

Finally, our new parameter estimation method runs as follows:
(i) we take an initial value of the parameter θ_0
(ii) θ_{n+1} is obtained from θ_n and X=x by:
- we simulate N realisations $\varepsilon_1,\ldots,\varepsilon_N$ of ζ according to the posterior distribution based on θ_n.
- θ_{n+1} is defined from $\varepsilon_1,\ldots, \varepsilon_N$ and X=x by (4) and (5), where α and β are defined by (6), (7) and (8) respectively.
Some other applications of ICE to the problem of unsupervised segmentation are exposed in [2] and [12].

2.2 Younes algorithm.

This iterative method runs as follows: let V_θ denote the energy of the distribution of (ζ,X), which also is, under hypotheses above, a Gibbs distribution. For each $(\zeta,X)=(\varepsilon,x)$ $V'_\theta(\varepsilon,x)$ will denote the gradient of V_θ with respect to θ.

(i) one takes an initial value θ_0
(ii) θ_{n+1} is calculated from θ_n and X=x by:

$$\theta_{n+1}=\theta_n+(k/n+1)[V'_{\theta_n}(\varepsilon_{n+1},x_{n+1})-V'_{\theta_n}(\varepsilon^*_{n+1},x)] \qquad (9)$$

where $(\varepsilon_{n+1},x_{n+1})$ is a realization of (ζ,X) simulated by Gibbs sampler (using θ_n), ε^*_{n+1} a realization of ζ according to posterior distribution (simulated by Gibbs sampler based on θ_n) and k a constant. This method is presented as a generalisation of the estimator defined by (6) to the case of hidden data, in fact, when the noise vanishes, i.e. when the variances of its conditional

distributions tend to 0, the formula (9) gives (6). Unfortunately its theoretical study becomes much more difficult and is not immediately generalisable to the hidden data case.

2.3. Unsupervised ICE and Stochastic Gradient based segmentation methods.

Let us denote by A1 our new parameter estimation ICE based algorithm and by A2 the method of Younes ([20]). Both algorithms A1 and A2 are iterative. When the sequence of parameters becomes steady the estimation step is finished and we apply simulated annealing ([9]). Let us denote by S1, S2 the methods so obtained of unsupervised segmentation.

Finally:

A1= ICE based on the algorithm of Younes ([19]) and empirical moments given by (6) and (7).

A2=Algorithm of Younes ([20])

S1=A1+MAP (simulated annealing [9])

S2=A2+MAP (simulated annealing [9])

As pointed out above the aim of our work is to compare A1 with A2 on the one hand and S1 with S2 on the other. Incidentally our study gives some information about the robustness of the simulated annealing with respect to the considered parameters.

We suppose ζ to be a Markov random field relative to four nearest neighbours, so the energy (function U in (1)) is defined by:

$$U(\varepsilon) = \sum_{s \in S} \phi(\varepsilon_s) + \sum_{s,\, t \text{ neighbours}} \gamma(\varepsilon_s, \varepsilon_t)$$

For m classes, ϕ is then defined by m values $\alpha_1=\phi(\omega_1)$, .., $\alpha_m=\phi(\omega_m)$ and we take γ of the shape $\gamma(\varepsilon_s,\varepsilon_t)=\delta$ if $\varepsilon_s=\varepsilon_t$, $-\delta$ otherwise. As the noise is Gaussian and independent conditionally to ζ, the parameter β is given by m means $\mu_1,\ldots,\mu_m$ and m standard deviations $\sigma_1,\ldots,\sigma_m$.

Numerous simulations (we consider two, three, four or five classes, different means, variances equal or different) allow us to put forth the following conclusions:

1. A1 is always much more efficient that A2, as far as the estimation of noise variances is concerned. The difference is particularily significant when they are different. The qualities of estimation of all other parameters are equivalent.

2. S1 is always more efficient that S2. The efficiency, measured by the rate of well classified pixels, can reach 15 percent when the noise is significant.

Our first study concerns the behaviour of the three methods (S1,S2 and the initialisation, which is a standard method based on histogramme and which will be denoted INIT) when the number of classes (NC) and the signal to noise ratio (SNR) vary. We consider a common standard deviation equal to 1, and SNR is then defined by: SNR=inf $|\mu_i-\mu_j|$. The results, expressed in rates of well classified pixels, are:

SNR	NC	2	3	4	5
1.5	INIT	77.3	68.1	63.6	64.0
	S1	86.9	84.9	88.2	88.2
	S2	82.4	75.7	72.9	74.8
2	INIT	83.9	77.2	74.5	73.8
	S1	96.5	96.4	97.2	97.7
	S2	91.8	88.4	89.1	89.7

Tab. 1.

We can note, according to Tab. 1., a real superiority of S1 over S2. It is particularly visible in the case of small SNR and four or five classes. This fact seems to be basically due to the superiority of A1 over A2 when the estimation of the variance of the noise is concerned. To be more precise let us consider two following examples:

1. Let us consider a four class (m=4) random field with $\alpha_1=\alpha_2=\alpha_3=\alpha_4=0$, $\delta=-5$. The corresponding image (IM1) is corrupted with a white Gaussian noise of means $\mu_1=32$, $\mu_2=96$, $\mu_3=160$, $\mu_4=224$ (we consider 256 grey levels) and a common standard deviation equal to 30 (IM2). The estimation of parameters gives:

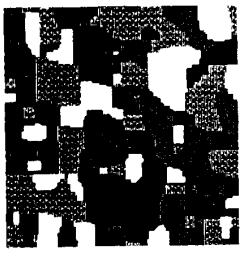
A1:
$\alpha_1=-0.002$ $\alpha_2=0.050$ $\alpha_3=-0.019$ $\alpha_4=-0.029$ $\delta=-0.65$
$\mu_1=33.19$ $\mu_2=96.37$ $\mu_3=158.28$ $\mu_4=220.81$
$\sigma_1=24.66$ $\sigma_2=26.23$ $\sigma_3=26.91$ $\sigma_4=24.72$
A2:
$\alpha_1=-0.001$ $\alpha_2=0.049$ $\alpha_3=-0.026$ $\alpha_4=-0.022$ $\delta=-0.62$
$\mu_1=34.34$ $\mu_2=96.02$ $\mu_3=157.68$ $\mu_4=219.72$
$\sigma_1=17.89$ $\sigma_2=18.38$ $\sigma_3=18.81$ $\sigma_4=17.83$

Both algorithms S1, S2 have been initialised by a classical method based on the histogramme (which gives 0.77 as a rate of well classified pixels: IM3). The corresponding rates (IM4, IM5) are 0.98 for S1 and 0.92 for S2.

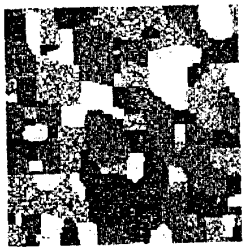
2. We consider a binary (m=2) image with $\alpha_1=\alpha_2=0$, $\delta=5$ (IM6) corrupted with white Gaussian noise $\mu_1=64$, $\mu_2=192$, $\sigma1=\sigma2=50$ (IM7). Estimations of parameters give:

A1 :
$\alpha_1=0.045$ $\alpha_2=-0.045$ $\delta=-0.573$
$\mu_1=66.65$ $\mu_2=185.70$ $\sigma_1=42.83$ $\sigma_2=43.62$
A2 :
$\alpha_1=0.008$ $\alpha_2=-0.008$ $\delta=-0.528$
$\mu_1=64.05$ $\mu_2=189.42$ $\sigma_1=33.99$ $\sigma_2=34.25$

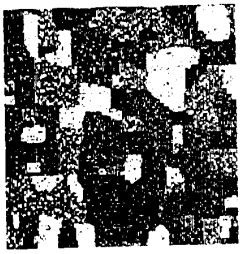
Both algorithms S1, S2 have been initialised by the same classical method based on the histogramme (which gives 0.90 as a rate of well classified pixels: IM8). The corresponding rates are 0.986 for S1 and 0.975 for S2 (IM9, IM10).



IM1



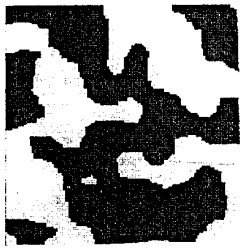
IM2



IM3

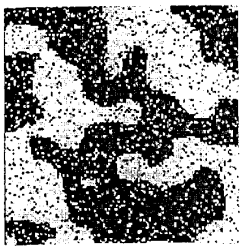
IM4

IM5

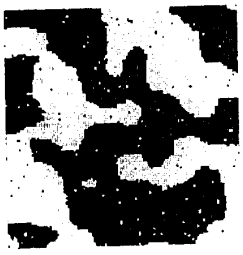


IM6

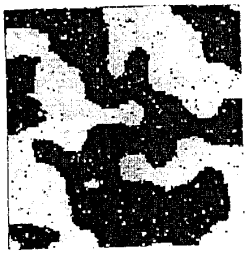
IM7



IM8



IM9

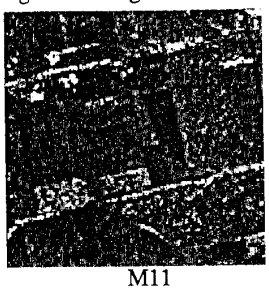


IM10

3. Segmentation of a real image.

We consider a real SAR image (IM11) which represents an agricol region of Ontario, Canada. It is well known that the conditional distributions of X are not Gaussian but, when the number of looks is large (IM11 is taken with 7 looks) they can be approached by Gaussian distributions. IM12 is the result of segmentation by the initialising method and IM13, IM14 are the results of segmentation by S1 and S2 respectively.

A rigourous comparison between the two methods is difficult because we ignore the real nature of the ground. However one could have the impression that the S1 segmentation gives more detail.



M11

IM12

IM13

IM14

4. Conclusion.

We presented a new algorithm of unsupervised Bayesian image segmentation and showed its superiority over a recent method based on the stochastic gradient. This superiority is verified in quite different situations: two, three, four or five classes, different means, variances equal or different. In some cases the difference of well classified pixels can reach 15 percent. Both unsupervised methods use the same segmentation algorithm (simulated annealing) and the difference is due to the difference of efficiencies of the estimators employed. Furthermore, the performances of two estimators used are comparable for every type of parameter except the noise variances, which are moderately well (A1) or badly (A2) estimated. Furhermore, they are quite sufficient when means and $\alpha_1, \alpha_2, .., \alpha_m$ are concerned, but not efficient in estimating δ.

This leads to two following conclusions (as far as the present framework is concerned):

- simulated annealing is sensitive to the noise variance; consequently a good estimation of this parameter is important.

- neither A1 nor A2 is efficient in estimating δ but the simulating annealing does not seem to be overly sensitive to this parameter.

So the collaboration of $\hat{\alpha}$, $\hat{\beta}$, ICE and simulated annealing turns out to be efficient in numerous situations.The good properties of $\hat{\alpha}$ ([19]), $\hat{\beta}$ and simulated annealing ([9]) being well known, we can say that ICE "saves them" in exposed cases.

The association of ICE and simulated annealing was already studied in a different context ([2]) where the estimator $\hat{\alpha}$ used is the algorithm of Derin and Elliot ([5]). The results obtained-although having to be interpreted with care, the simulation study being not rich enough-are good. In particular one of the corresponding unsupervised methods turns out to be more efficient, in the case of an homogeneous image, than the algorithms of Chalmond ([3]) and Devijver and al. ([6]).

In the general manner the use of ICE seems to be especially interesting in unsupervised segmentation of images where it appears as an alternative to the EM algorithm whose use is rather tedious and requires strong hypotheses ([18]): discrete or Gaussian noise, the same variance for all classes.The ICE procedure can be applied in a much more general framework. When modelling the distribution of ζ by a Gibbs distribution and supposing that the random variables (X_S) are conditionally independent, the conditional densities $f_1, f_2, ..., f_m$ of each X_S (for m classes) can be of any kind, in particular of nonlimited kinds. Moreover, they can be of distinct kinds. In fact, simulations of realizations of ζ according to the a posteriori distribution (by means of the Gibbs sampler for example) are possible once the $f_i(x_S)$ $(1 \le i \le m,\ s \in S)$ are known. The only condition is to estimate, for each $1 \le i \le m$, a parameter β_i (that defines f_i) based on a sample issued by class ω_i, that in general is possible using the empirical moments.

The possible applications of ICE go beyond image segmentation: we expose in [17] some examples of estimation in the case of hidden data. In the case of a simple Gaussian mixture the well known EM re-estimation formulae can be obtained from ICE in a different way, without any reference to the likelihood. ICE can also be used to estimate all parameters in a noisy auto-regressive process and in a hidden Markov chain. This last possibility affords the construction of a fast unsupervised image segmentation algorithm ([8]).

The great diversity of applications of the ICE procedure is due to the great freedom in choosing the estimator $\hat{\theta}$. In particular, the notion of likelihood is not necessary and it is possible to use modeling with Dirac's measure.

References

[1] J. Besag, "On the statistical analysis of dirty pictures," J. Roy. Stat. Soc., Ser. B, vol. 48, pp. 259-302, 1986.

[2] B. Braathen, N. Marhic, P. Masson, W. Pieczynski, "Sur une nouvelle approche de segmentation bayesienne non supervisée d'images," in Proc. GRETSI, Juan-Les-Pins, France, Sept., 1991.

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

[4] M. M.Dempster, N. M. Laird, D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Roy. Stat. Soc., Ser. B, vol. 39, pp. 1-38, 1977.

[5] H. Derin, H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields" IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-9, pp. 39-55, 1987.

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

[7] R.C. Dubes, A.K. Jain, "Random field models in image analysis," Journal of Applied Statistics, vol. 6, no 2, 1989.

[8] M. Emsalem, H. Caillol, P. Olivie, G. Carnat, W. Pieczynski, "Fast unsupervised statistical image segmentation," in Proc. 1992 Int. Geosci. Remote Sensing Symp., IGARSS 92, Houston, Texas, May 92.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-6, pp. 721-741, 1984.

[10] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," IEEE Trans. Pattern Anal. and Machine Intell., vol. PAMI-11, no 8, 1989.

[11] N. Marhic, P. Masson , W. Pieczynski, "Mélange de lois et segmentation non supervisée des données SPOT," Statistique et Analyse des Données, vol. 16, no 2, 1991.

[12] N. Marhic, W. Pieczynski, "Estimation of mixture and unsupervised segmentation of images," in Proc. 1991 Int. Geosci. Remote Sensing Symp., IGARSS 91, Helsinki, Finland, May 1991.

[13] J. Marroquin, S. Mitter, T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," Journal of the American Statistical Association, vol. 82, pp. 76-89, 1987.

[14] P. Masson and W. Pieczynski, "Segmentation of SPOT images by contextual SEM,", in Proc. of EUSIPCO 90, Barcelona, Spain, Oct. 90.

[15] W. Pieczynski, "Estimation of context in random fields," Journal of Applied Statistics, vol. 16, no 2, 1989.

[16] W. Pieczynski, "Mixture of distributions, Markov random fields and unsupervised segmentation of images," Technical report no 122, L.S.T.A., Université Paris VI, 1990.

[17] W. Pieczynski, "Parameter estimation in the case of hidden data," in Proc. of 16th Biennal Symp. on Communications, Kingston, Canada, May 1992.

[18] W. Qian and D. M. Titterington, "On the use of Gibbs Markov chain models in the analysis of images based on second-order pairwise interactive distributions," Journal of Applied Statistics, vol. 16, no 2, 1989.

[19] L. Younes, "Estimation and annealing for Gibbsian fields," Annales de l'Institut Henri Poincaré , 24, pp. 269-294, 1988.

[20] L. Younes, "Parametric inference for imperfectly observed Gibbsian fields," Probability Theory and Related Fields, 82, pp. 625-645, 1989.