ESTIMATION OF MIXTURE AND UNSUPERVISED SEGMENTATION OF IMAGES

Nicole Marhic*   Wojciech Pieczynski **

*** Groupe Traitement d'Images**
Ecole Nationale Supérieure des Télécommunications
B.P. 832, 29285 BREST Cédex, FRANCE
****Groupe Images**
Institut National des Télécommunications
9, rue C. Fourier, 91011 EVRY Cédex, FRANCE

ABSTRACT

This paper presents a statistical approach to the Bayesian unsupervised segmentation of images. The main problem lies in the estimation of a distribution mixture. Iterative Conditional Estimation provides solutions to such a problem. After a brief recall about the general procedure two stochastic algorithms are described in the case of a finite Gaussian mixture. They are applied on synthetic images corrupted by Gaussian noise in order to estimate the parameters required when performing contextual Bayesian segmentation.

Keywords: Distributions mixture- conditional estimation- contextual segmentation- Bayesian rule- Gaussian noise.

INTRODUCTION

This work deals with a statistical approach to the unsupervised Bayesian segmentation of images. The Bayesian segmentation problem can be expressed as follows. Let S denote a set of pixels; an image is then modeled by a pair ($\zeta = (\zeta_s)_{s\in S}$, $X=(X_s)_{s\in S}$) of random fields. Each ζ_s takes its values in a finite set of classes Ω and each X_s is a real random variable. The distribution of (ζ,X) is defined by the distribution P_ζ of ζ and by the conditional distributions P_X^ε of X given $\zeta=\varepsilon$; so is defined the conditional distribution of ζ given X=x (*a posteriori* distribution). Bayesian segmentation consists in estimating the unobservable realization of ζ from the data X=x by ζ^* which maximizes the a posteriori distribution.

When sufficient information about P_ζ and P_X^ε is available there are several algorithms which can be used for Bayesian segmentation. When using the whole observation one can choose between the well known MAP and MPM methods which can be performed respectively by the Simulated Annealing algorithm described by Geman *et al.* in [2] and the algorithm proposed by Marroquin *et al.* in [3]. In both these approaches ζ is a Markov field and the random variables $(X_s)_{s\in S}$ are assumed to be independent conditionally to ζ.

Another point of view consists in classifying each ζ_s from observations made on a neighborhood V_s of the pixel s; such methods are said to be local while MAP and MPM are global methods.

When classifying real data, P_ζ and P_X^ε are unknown and, whatever the segmentation method retained, their estimation must be carried out in a previous step or simultaneously from the random variables X_s whose distribution is a mixture of distributions.

This paper focuses on local methods implying estimation in the case of images dominated by Gaussian noise.

ITERATIVE CONDITIONAL ESTIMATION

Let (ζ,X) be a pair of random variables, the distribution of which is assumed to be defined by the distribution P_ζ of ζ and the family $\{P_X^\varepsilon\}$ of distributions of X given $\zeta=\varepsilon$, these distributions depending respectively on parameters α and β that need to be estimated from X only. Let us suppose that estimators α^* and β^* for α and β can be defined from ζ and (ζ,X) respectively. When ζ is unobservable the Iterative Conditional Estimation procedure described in [6] defines α^+ and β^+ as conditional expectations of α^* and β^* given X=x computed according to the current values α^c and β^c. These are the best approximations of α^* and β^* in terms of mean square error when the current parameter tends towards (α,β).

From an initial value (α_0,β_0) the reestimation formulæ are:

$$\alpha^+ = E_c[\alpha^* | X = x] \quad (1.1)$$

$$\beta^+ = E_c[\beta^* | X = x] \quad (1.2)$$

where E_c denotes the conditional expectation using (α^c,β^c).

We will see in the following sections that, when the contextual segmentation in concerned, an explicit computation of α^+ using (1.1) is possible but the computation of β^+ has to be done through a "stochastic" approximation of (1.2).

BAYESIAN CONTEXTUAL SEGMENTATION

For A and B such that $S \supseteq B \supseteq A$ let us denote by ζ_A and X_B the restriction of ζ and X to A and B respectively. Bayesian segmentation consists in estimating ζ_A by $\zeta_A{}^*$ that maximizes the distribution of ζ_A given $X_B=x_B$. In the case of contextual segmentation A is set equal to {s} and B is a neighborhood of the pixel s. ζ needs not to be assumed Markovian and "discriminating" functions can be used directly. These functions are derived from the distribution of (ζ_B,X_B) that is assumed to be independent of B.
For each ω_i in Ω:

$$FD(\omega_i, x_B) = \sum_{\omega \in \Omega^{|B|-1}} P\{\zeta_s = \omega_i, \zeta_{B-\{s\}} = \omega, X_B = x_B\} \quad (2)$$

where $|B|$ denotes the cardinal of B and Ω is the (finite) set of classes where ζ_s takes its values.
Then :

$$\zeta_s{}^* = \omega_i \Leftrightarrow FD(\omega_i, x_B) = \underset{\omega_j \in \Omega}{\mathrm{Max}}\ FD(\omega_j, x_B) \quad (3)$$

So, in order to perform segmentation, some knowledge about the distribution of (ζ_B,X_B) is required. In the most general case it has to be estimated from the data X=x.
Let us denote by F_i the distribution of X_B given $\zeta_B=\varepsilon_i$ and assume that ζ_B is equal to ε_i with probability α_i. The law of X_B can be written:

$$F = \sum_{i=1}^{m} \alpha_i F_i \quad (4)$$

where $m = (\mathrm{card}\Omega)^{|B|}$.
Then the problem is to estimate the components of a distribution mixture.

In the following, X is assumed to be Gaussian conditionally to ζ. In this case F_i depends on a parameter β_i composed of a mean vector and a covariance matrix and the problem lies in estimating $\theta=\{\alpha=(\alpha_1,...,\alpha_m);\beta=(\beta_1,...,\beta_m)\}$ from a realization x of X which provides a sample of X_B.

FINITE GAUSSIAN MIXTURE

Let $x=(x_1,...,x_n)$ be a sample of independent realizations of a random variable X taking its values in R^d.

The density of X is assumed to be:

$$f(t) = \sum_{i=1}^{m} \alpha_i f_i(t) \quad (5)$$

where f_i is normal with mean μ_i and covariance matrix Γ_i; each α_i belongs to]0,1[and $\Sigma\{\alpha_i;\ i=1,...,m\}=1$. In fact, $(\alpha_i)_{i=1,..m}$ corresponds to the *a priori* distribution described here above.
The problem lies in the estimation of $\alpha=(\alpha_1,...,\alpha_m)$ and $\beta=((\mu_1,\Gamma_1),...,(\mu_m,\Gamma_m))$ from the sample x.
x can be regarded as a sample of incomplete data by considering each x_k to be the "known" part of an observation $y_k=(x_k,\omega_k)$ where ω_k is an integer indicating the component population of origin. Then Iterative Conditional Estimation provides solutions to such a problem as soon as estimators for α and β can be defined and if the conditional expectations in (1) can be computed or approached.
If $\varepsilon=(\omega_1,...,\omega_n)$ were known, an estimator for α could be defined by the empirical distribution as follows:
For $i \in \{1,...,m\}$,

$$\alpha_i^* = \frac{1}{n} \sum_{k=1}^{n} \chi_{[\omega_k = i]} \quad (6)$$

where $\chi_{[\omega_k=i]}=1$ if $\omega_k=i$, 0 otherwise.
An alternative is to use the estimator proposed by Tilton in [8]; in this case:

$$\alpha^* = A^{-1}Y \quad (7)$$

where $A=[a_{ij}]$ with $a_{ij}=\int_{R^d} f_i f_j$ for $i,j \in \{1,...,m\}$ and $Y={}^t(Y_1,..., Y_m)$ with $Y_i=\Sigma\{f_i(x_k)/n;\ k=1,...,n\}$.
The components of β can be estimated by empirical means and covariance matrices respectively:

$$\mu_i^*(\varepsilon, x) = \frac{\sum_{k=1}^{n} x_k \chi_{[\omega_k = i]}}{\sum_{k=1}^{n} \chi_{[\omega_k = i]}} \quad (8.1)$$

$$\Gamma_i^*(\varepsilon, x) = \frac{\sum_{k=1}^{n} (x_k - \mu_i{}^*)(x_k - \mu_i{}^*)^T \chi_{[\omega_k = i]}}{\sum_{k=1}^{n} \chi_{[\omega_k = i]}} \quad (8.2)$$

Then by considering (6) and (8) or (7) and (8) two iterative algorithms denoted respectively by A1 and A2 can be defined whenever formulae (1) can be computed or approached.
When α^* is defined by (6), $\alpha^+ = E_c[\alpha^* \mid x]$ can be computed:

we find, for each α_i^+, the mean value of the posterior probabilities p_{ik}, computed from the current values α^c and β^c.

$$p_{ik}=\text{Prob}\{\omega_k=i \mid x_k\} \qquad (9)$$

According to the law of large numbers, β^* given by (1.2) can be approached by:

$$\frac{1}{N}\sum_{l=1}^{N} \beta^*(\varepsilon_l, x) \qquad (10)$$

where $\varepsilon_1,\ldots,\varepsilon_N$ are independent realizations of ε according to the posterior probabilities p_{ik} based on α^c and β^c.

Then from an initial value (α_0,β_0) taken not "too far" from the real parameter and if $\alpha^c=(\alpha_1{}^c,\ldots,\alpha_m{}^c)$ and $\beta^c=(\beta_1{}^c,\ldots,\beta_m{}^c)$ are the current approximate values of α^* and β^* the new values prescribed by the algorithm A1 satisfy:

$$\alpha_i^+ = \frac{1}{n}\sum_{k=1}^{n} \frac{\alpha_i^c f(x_k|\beta_i^c)}{f(x_k|\beta^c)} \qquad (11)$$

and:

$$\mu_i^+ = \frac{1}{N}\sum_{l=1}^{N} \frac{\sum_{k=1}^{n} x_k \chi_{[\omega_{l,k}=i]}}{\sum_{k=1}^{n} \chi_{[\omega_{l,k}=i]}} \qquad (12.\ 1)$$

$$\Gamma_i^+ = \frac{1}{N}\sum_{l=1}^{N} \frac{\sum_{k=1}^{n} (x_k-\mu_i^+)(x_k-\mu_i^+)^T \chi_{[\omega_{l,k}=i]}}{\sum_{k=1}^{n} \chi_{[\omega_{l,k}=i]}} \qquad (12.\ 2)$$

where $\omega_{l,k}$ is the k-th point of ε_l.

It can be noticed that α_i^+ given by (11) is the reestimated value prescribed by the M-step of the EM algorithm for α (cf. [7]).

Since Tilton's estimator dœs not depend on ε the reestimation formulæ of algorithm A2 are obtained directly by replacing (11) by:

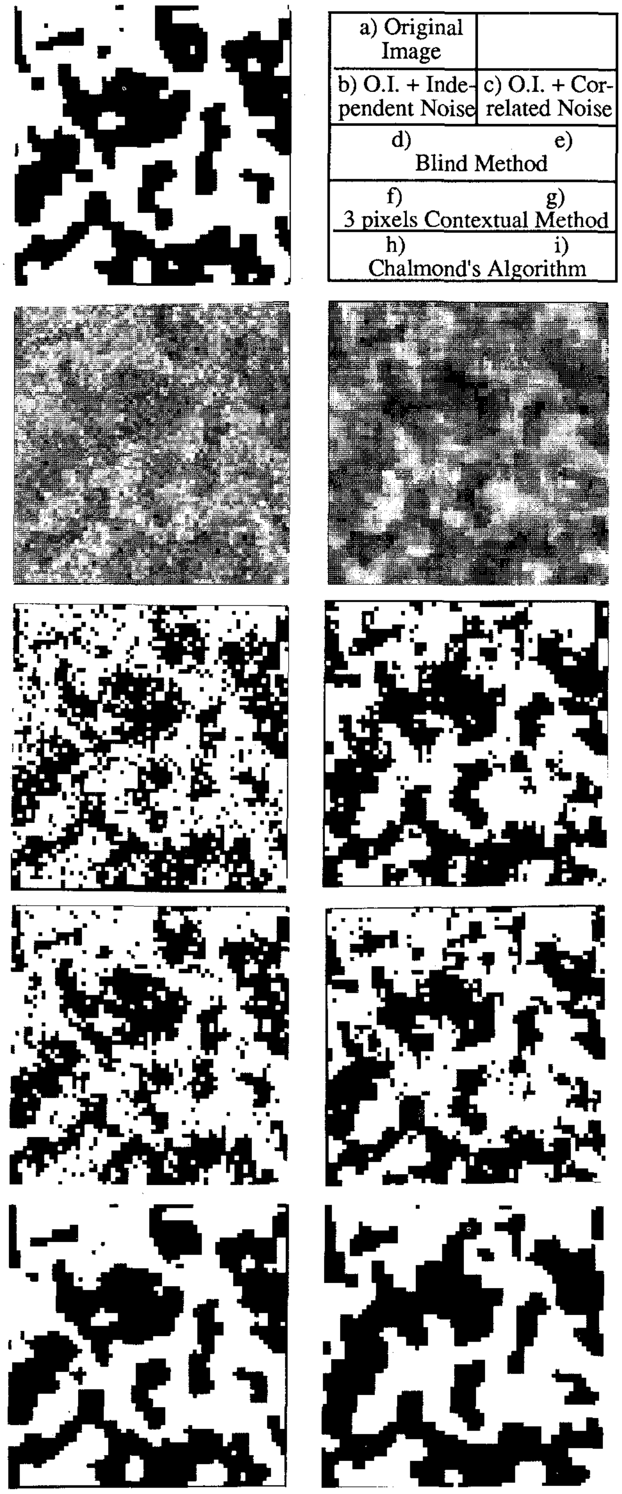
$$\alpha^+ = (A^c)^{-1} Y^c \qquad (13)$$

where A^c and Y^c are computed as in (7) according to the current value of β.

RESULTS

The algorithms have been first applied in order to perform contextual segmentation of two classes Markovian fields corrupted with a real Gaussian noise spatially independant and spatially correlated. The corresponding means are $m_1=1$, $m_2=3$ and dispersion $\sigma^2=1$.

Figures (1d), (1e), (1f) and (1g) are Bayesian restorations obtained when estimating the parameters by algorithm A1. Image (2c) is a restoration of (2b) by the three pixel contextual method involving algorithm A2. In each case we also present the restoration obtained by a global method that is the algorithm proposed by Chalmond in [1].

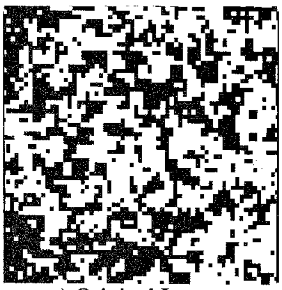


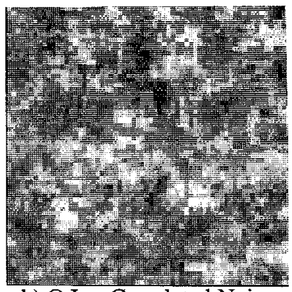
a) Original Image	
b) O.I. + Independent Noise	c) O.I. + Correlated Noise
d) Blind Method	e)
f) 3 pixels Contextual Method	g)
h) Chalmond's Algorithm	i)

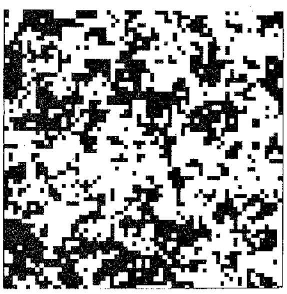
Figure 1

From (4) it can be easily seen that the complexity in estimation increases with the context size. The blind method corresponds to a context size equal to one; the "three pixels contextual" method restores a pixel s from observations made on a neighborhood V_s of s:

t'	
s	t


a) Original Image


b) O.I. + Correlated Noise


c) 3 pixels contextual method

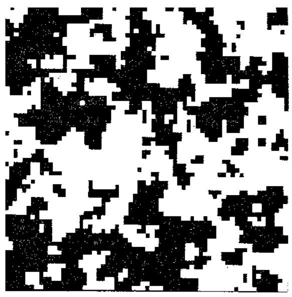

d) Chalmond's Algorithm

Figure 2

Percentages of misclassified pixels are given in Table 1 and 2.

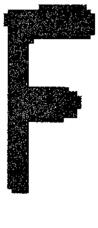
fig 1d: 16.4	fig 1e: 16.3
fig 1f: 12.5	fig 1g: 14.3
fig 1h: 4.9	fig 1i: 15.7

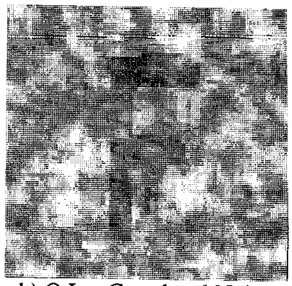
Table 1: error rates relatives to figure 1

fig 2c: 12.3	fig 2d: 21.1

Table 2: error rates relatives to figure 2

In figure 3 the original image (3a) is a handdrawn letter digitalized and binarized. It has been corrupted with a spatially correlated noise to obtain image (3b). Figure (3c) (respectively (3d)) is a Bayesian segmentation in two classes (respectively 3 classes) of (3b) by the blind method when the parameters are estimated by algorithm A1.


a) Original Image


b) O.I. + Correlated Noise


c) A1 with 2 classes

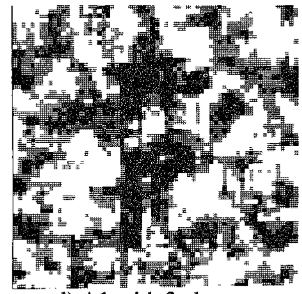

d) A1 with 3 classes

Figure 3

CONCLUSION

Non supervised image segmentation appears as an important problem because, in practice, the parameters required are unknown. When Bayesian segmentation is concerned the algorithms A1 and A2, which are relatively simple to implement, seem to be well suited to carry out their estimation in the Gaussian case.

Numerous simulations allow the following conclusions:

- In both cases (independent or correlated noise) the "3 pixels contextual" method is more efficient than the blind method.
- But, as spatial correlation of the noise increases the contextual method is less efficient. However it seems to stay more efficient than Chalmond's algorithm in the case of correlated noise, especially when the original image is not homogeneous.

REFERENCES

[1] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognition, Vol. 22*, No.6, pp. 747-761, 1989.

[2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6*, pp.721-741, 1984.

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

[4] P. Masson and W. Pieczynski, "Segmentation of SPOT images by contextual SEM," in *Proceedings of EUSIPCO*, Barcelone, Spain, Sept. 1990.

[5] W. Pieczynski, "Estimation of context in random fields," *Journal of Applied Statistics, Vol. 6*, No.2, 1989.

[6] W. Pieczynski, "Mixture of distributions, Markov random fields and unsupervised segmentation of images," Technical report, No.122, LSTA, Université de Paris 6, 4 place Jussieu, PARIS, Nov.1990.

[7] R.A. Redner and H.F. Walker, "Mixture densities, maximum likehood and the EM algorithm," *Society for Industrial and Applied Mathematics Review,Vol. 26*, No.2, pp. 195-239, Apr. 1984.

[8] J. Tilton, S. Vardeman and P. Swain, "Estimation of context for statistical classification of multispectral image data," *IEEE TGMS, GE-20*, pp.445-452, Oct. 1982.