UNSUPERVISED BAYESIAN CLASSIFICATION OF SAR - IMAGES

Hans-Christoph QUELLE* Jean-Marc BOUCHER* Wojciech PIECZYNSKI**

* Groupe Traitement d'Images
Département Mathématiques et Systèmes de Communication
Ecole Nationale Supérieure des Télécommunications de Bretagne
Technopôle de Brest-Iroise, B.P. 832, 29285 BREST Cedex, France
** Groupe Images
Département Systèmes et Reseaux
Institut National des Télécommunications
9, rue Charles Fourier, 91011 EVRY Cedex, France

ABSTRACT

We present a non-supervised method of bayesian contextual radar image segmentation. We adopt the hierarchical image model (HIM) from Kelly and Derin. In contrast to their global segmentation method a local method is used in order to speed up the segmentation. The algorithm for parameter estimation is SEM, a recent variation of EM. The algorithm obtained is tested on synthetic images and also applied to the segmentation of real SEASAT-scenes.

Key Words: Unsupervised Bayesian local (contextual or blind) segmentation, Hierarchical Field Model.

INTRODUCTION

When considering a statistical segmentation of images authors generally suppose the existence of two random fields : the field of "classes" $\xi = \{\xi_s : s \in S\}$, and the field of "measurements" $X = \{X_s : s \in S\}$. Each ξ_s takes its value in a finite set $\Omega = \{\omega_1, \dots, \omega_K\}$ of classes and X_s in R. So the problem of segmentation is the problem of the estimation of an "ignored" realization of ξ from an "observed" realization of X.

We will suppose that the realization of X depends on the realization of ξ and a certain "noise". The distribution of (ξ, X) is defined by P_ξ , the distribution of ξ, and the family P_X^ε of distributions of X conditional to $\xi = \varepsilon$.

MODELING RADAR IMAGES

The hierarchical image model adopts this general proceeding, it consists of two random fields. One governs the grouping of pixels, called region process. The other consists of K random fields which represent the speckled appearance of the K types of nature, the classes.

We have chosen Kelly and Derin's image model for two reasons.

Firstly, this model points clearly out the idea that the "observed speckle" is not only a worthless disturbance, but carries information about the nature of ground.

Secondly, the hierarchical character fits well to Bayesian decision theory and this model allows quite easily the generation of a great number of images with real radar image statistics.

Region process

The region process is responsible for the distribution of pixels within the different classes.

The field ξ is said to be markovian, with :

$$P[\xi_i = \varepsilon_i | \xi_j = \varepsilon_j,\ (i) \neq (j)] = P[\xi_i = \varepsilon_i | \xi_j = \varepsilon_j,\ (i) \in \nu_i]$$

where ν_i designs a neighborhood of (i).

In fact it can be a very simple type of a MF, called multilogistic level (MLL) field [3] [5]. In this model one parameter exists for each clique type: α_k for the K differently colored singletons, β for the four types of clique-pairs. For the sake of simplicity all cliques consisting of more than two pixels are ignored.

Under a positivity condition the field ξ has the Gibbs distribution :

$$P[\xi = \varepsilon] = \frac{1}{Z}\exp\left\{-\sum_{c \in C} V_c(\varepsilon)\right\} \qquad (2)$$

Where the potential function $V_c(\varepsilon)$ is defined as :

- for c a clique of pairs :

$$V_c(\varepsilon) = \begin{cases} \beta & \text{if all } \xi_i \text{ are equal} \\ -\beta & \text{otherwise} \end{cases} \qquad (3)$$

- for c a single pixel clique :

$$V_c(\varepsilon) = \alpha_k \quad \text{if } \varepsilon_i = \omega_k \quad \text{for } (i) \in c \qquad (4)$$

Speckle process

The speckle consists, in analogy with the physical model of a field, of complex, gaussian, zero-mean random variables $Z = \{Z_i : i \in S\}$.

At each site i, the intensity w and the phase φ are defined by :

$$w_i = |z_i|^2, \text{ and } \varphi_i = \text{arctg}\frac{Im[z_i]}{Re[z_i]} \quad (5)$$

The density of the intensity of one pixel is then :

$$P[w] = \int_0^{2\pi} P[w, \varphi] d\varphi = \frac{1}{2\sigma^2} e^{-\frac{w}{2\sigma^2}} \quad (6)$$

The joint probability $P[\underline{w}]$ may be expressed explicitly [5] with :

$$\underline{W} = [W_{k,l}, W_{k-1,l}, W_{k+1,l}, W_{k,l-1}, W_{k,l+1},] \qquad (k,l) \in S$$

For multi-look images with

$$V = \frac{1}{M}\sum_1^M W_m \quad (7)$$

the joint density can be expressed by a (M - 1) fold convolution of either $P[w]$ or $P[\underline{w}]$:

$$P[v] = \frac{v^{(M-1)}}{\Gamma_{(M)}\mu^M} e^{-\frac{v}{\mu}}; \text{ with } \mu = \frac{E[v]}{M} \quad (8)$$

Expressing $P[\underline{v}]$ analytically fails at the evaluation of the convolution integral.

Hierarchical model

At each pixel s the value of $\xi_s = \omega_k$ determines the marginal distribution of the speckle process. The color of the pixel s from speckle process, $W_s^{(k)}$ gives the final value X_s .

$$X_s = W_s^{(k)} \quad \text{if} \quad \xi_s = \omega_k \quad (9)$$

BAYESIAN APPROACH

Global

We denote by P_ξ the distribution of ξ and by f_X^ε the density of the distribution of X conditional to $\xi = \varepsilon$. This defines the distribution of (ξ, X) and therefore the conditional distribution of ξ knowing that $X = x$, which we denote P^x. The Bayesian rule r_g is then defined by :

$$r_g(x) : \hat{\xi} \Leftrightarrow P^x[\hat{\xi}] = \sup_\varepsilon P^x[\varepsilon] \quad (10)$$

it can also be expressed as follows :

$$r_g(x) : \hat{\xi} \Leftrightarrow P_{\hat{\xi}} f_X^{\hat{\xi}} = \sup_\varepsilon P_\varepsilon f_X^\varepsilon \quad (11)$$

The functions $g_\varepsilon = P_\varepsilon f^\varepsilon$ are called "discriminating". From the basis of (11) it is "impossible" to find $\hat{\xi}$ directly, due to the high number of possible ε - realizations (equal to K^n, n = card (S)).

S. Geman et al. propose in [4] an iterative procedure, called "simulated annealing", allowing the approximation of $\hat{\xi}$ given by (10) and (11). This method is used by Kelly and Derin [5]

Contextual

The contextual method consists in estimating the realization of each ξ_s from $X_v = \{X_t : t \in v_s\}$. The method corresponding to $v = \{s\}$ will be called "blind". The Bayesian rule is expressed as follows :

$$r_v(x) : \hat{\xi}_s = \omega_m \Leftrightarrow g_m(x_v) = \sup_{1 \le q \le K} g_q(x_v) \quad (12)$$

Let us denote n = card (v), $v^* = v - (s)$. For each $\varepsilon^* \in \Omega^{n-1}$, $P_{m,\varepsilon^*} = P[\xi_s = \omega_m, \xi_{v^*} = \varepsilon^*]$ and f^ε the density of X_v knowing $\zeta_v = \varepsilon$. The discriminating functions g_q in (12) may be expressed as follows :

$$g_q(x_v) = \sum_{\varepsilon^*} P_{q,\varepsilon} f^{(w_q, \varepsilon^*)} \quad (13)$$

They can be calculated as soon as the distribution P_ε of ξ_v and the conditional densities f^ε knowing $\xi_v = \varepsilon$ are determined.

So the previous problem is that of estimation the parameters of P_v and ξ^ε. This is the problem of estimation of a mixture of densities : we propose the use of a recent variation of EM, called the SEM [1] [2] [6].

ALGORITHM SEM

Let $v_1, \ldots, v_n$ be a sequence of contexts in S, we denote further by $E = (\varepsilon_1, \ldots, \varepsilon_q)$ with $q = K^n$ and n = card(v) the set of possible realizations of ξ_v and f_i conditional distributions of X_v which will be supposed exponentional (gamma). Each f_i is defined by its mean μ_i , by its covariance matrix R_i and by M, the number of averaged independent looks. Let Π be the distribution of ξ_s , $\Pi_i = P[\xi_v = \varepsilon_i]$.

Initialization :

$\Pi^{t=0}$ and parameters defining $f^{t=0}$ are chosen at random. They can also be estimated by easier methods.

Step S :

For each x_i we draw a realization $e^t(x_i)$ from the set E with regard to the conditional distribution knowing $X_i = x_i$ with the help of the couple $(\Pi_k, f_k)^t$. We obtain a partitioning $Q_1^t, \ldots, Q_K^t$ with :

$$[x_i \in Q_j^t] \Leftrightarrow [e^t(x_i) = \omega_j] \quad (14)$$

Step M :

We estimate Π_k^{t+1} by :

$$\Pi_k^{t+1} = \frac{\text{card}(Q_k)}{\text{card}(S)} \quad (15)$$

and parameters defining f_k^{t+1} are estimated in the selection Q_k^t by classical estimators (empirical mean and covariances).

Step E :

We compute the distribution $P_{X_i}^{t+1}(\omega_j)$ and return to step S.

Experimental Results

Simulated image

We consider a two class image modelled by a HIM, with $\alpha_1 = \alpha_2 = 0$, and $\beta = -0.3$. We processed 50 iterations, using the Gibbs sampler.

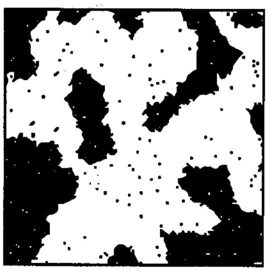

Fig. 1: The field X, the partition of classes.

The speckle processes for the two classes have the parameters $\mu_1 = 2\sigma_1^2 = 20$, $\mu_2 = 2\sigma_2^2 = 40$ and $\rho_{k_{1,2}} = \rho_{v_{1,2}} = 0,1$.

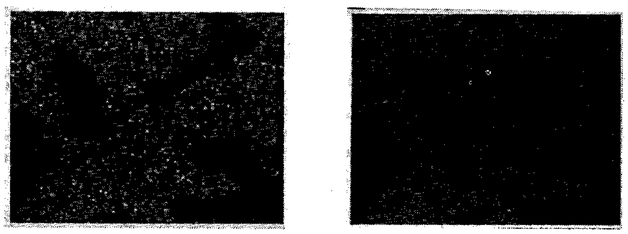

Fig. 2: The speckled images for N = 4 (left), N = 16 (right).

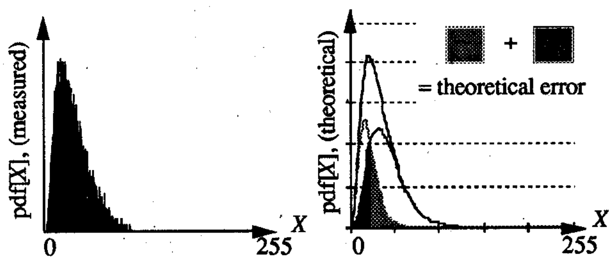

Fig. 3 : N = 4 : The density of X mesured (left) and theoretical (right) and the minimum error of blind classification $e_{th} = 0.24$.

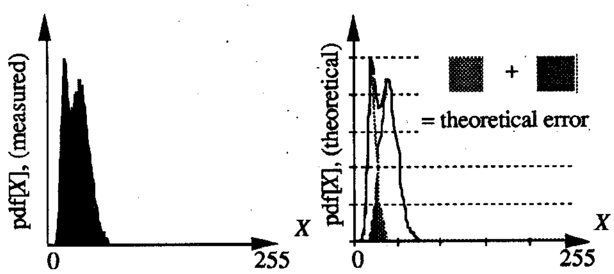

Fig. 4 : N = 16 : The density of X mesured (left) and theoretical (right) and the theoretical error of blind classification $e_{th} = 0.14$.

<u>Test between two different versions of the algorithm.</u>

- Contextual method, the density of one pixel is gaussian.
  ==> $v_s = (s)$ + two neighbors, gaussian f^{ε}.
- Blind method, one pixel has a gamma density.
  ==> $v_s = (s)$, gamma f^{ε}.

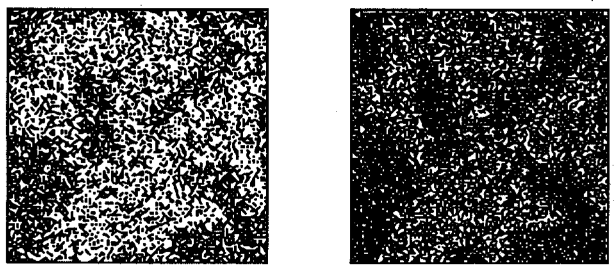

Fig. 5 :

left :	right:
$v_s = (s)$	$v_s = (s)$ + two neighbors
f^{ε} : gamma.	f^{ε} : gaussian.

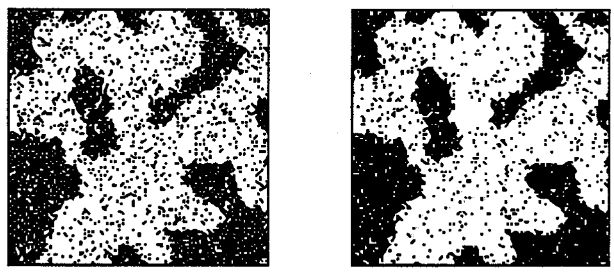

Fig. 6 :

left :	right:
$v_s = (s)$	$v_s = (s)$ + two neighbors
f^{ε} : gamma.	f^{ε} : gaussian.

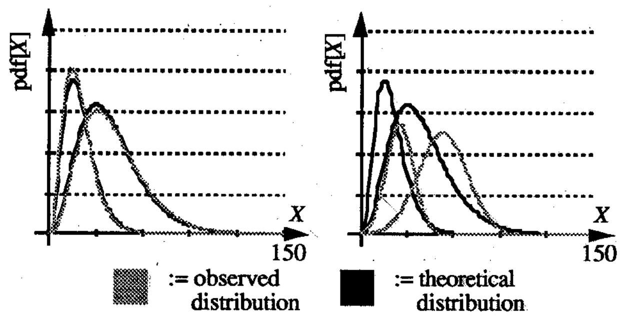

Fig. 7 : observed and theoretical distributions for $N = 4$, blind method (left) and contextual method (right).

The error rates in figure 5 and 6 are :

Method	Nb.of looks	Estimated means		Error rate
gamma, blind	4	19,4	41,1	0,314
gauss, contextual	4	25,3	53,0	0,403
gamma, blind	16	19,4	39,9	0,123
gauss, contextual	16	19,7	40,1	0,067

The experiment shows, that if the form chosen a priori for the densites f^{ε} deviates too much from the theoretical form, the spatial information from the neighbors is worthless.

For $N = 4$, the hypothesis, that the gamma density can be written as a gaussian, is not valid.

For $N = 16$, the gamma and the gaussian form are sufficiently similar so that the contextual method is able to achieve lower error rates.

Figure 8 shows, that the blind method converges in less iterations than the contextual method. Indeed one iteration of the contextual method is about 50 times longer than one iteration of the blind method.

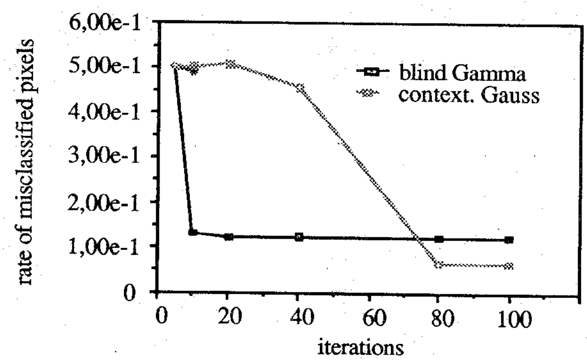

Fig. 8 : Convergence as a function of iteration, ($N = 16$).

Real image (SEASAT-SAR)

Fig. 9 : Real SEASAT-SAR scene from La Rochelle (France).

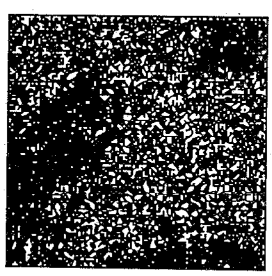

Fig. 10 :

left:

$v_s = (s)$

f^{ε} : gamma.

$\mu_1 = 17{,}2$ $\mu_2 = 56{,}7$

right:

$v_s = (s)$ + two neighbors

f^{ε} : gaussian.

$\mu_1 = 34{,}4$ $\mu_2 = 88{,}0$

CONCLUSION

a) We have compared the accuracy and the rapidity of the blind method and the contextual method using many different synthesized images. The results show that the blind method converges much faster but the contextual method is much more accurate for $N > 4$. So we combined these two methods, using the blind method as a pre-processing (pre-segmentation) for the contextual method. The moment to switch between the two methods is implemented in such a way that the final segmentation is found as fast as possible.

b) The contextual SEM has shown to be a robust method with regard to the variability of spatial dependence. Also when the dependency increases, the accuracy decreases, but remains always better than the blind method. For an image (N = 16 - look) which consists of two classes (means 20,40), the blind method achieves an error rate of 12%, whereas the contextual method achieves 8% for strong spatial dependency and for light dependency 6%.

REFERENCES

[1] G. Celeux, J. Diebolt, "L'algorithme SEM: Un algorithme d'apprentissage probabiliste pour la reconnaisance de mélange de dénsités," *Revue de Statisiques Appliquées*, **vol. 34**, (no. 2), 1986.

[2] G. Celeux, J. Diebolt, "Comportement asymptotique d'un algorithme d'apprentissage probabiliste pour les mélanges de lois de probabilités," Rapport de la Recherche, no. 563, 1986.

[3] H. Elliot, H. Derin, R. Christi and D. Geman, "Application to the Gibbs Distribution to image segmentation," in *Proceedings of the Int. Conf. Acoust., Speech, Signal Processing. ICASSP '84*, San Diego, CA, March 1984, pp. 32.5.1 - 32.5.4.

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

[5] P.A. Kelly, H. Derin, K.D. Hartt, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Transactions on Acoustics, Speech and Signal Processing*, **vol. 36**, (no. 10), pp. 1628 - 1641, 1988.

[6] P. Masson, W. Pieczynski, "Segmentation of SPOT images by contextual SEM," *Proceedings of EUSIPCO*, Barcelona, Spain, Sept. 1990.

[7] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, **vol. 21**, pp. 1087 - 1091, 1953.

[8] W. Pieczynski, "Estimation of context in random fields," *Journal of Applied Statistics*, **vol. 6**, (no. 2), 1989.