UNSUPERVISED BAYESIAN SEGMENTATION USING HIDDEN MARKOVIAN FIELDS

F. Salzenstein and W. Pieczynski

Département Signal et Image
Institut National des Télécommunications
9, rue Charles Fourier
91011 EVRY Cédex FRANCE

ABSTRACT

The aim of our paper is to present a new unsupervised Bayesian image segmentation method using a recent model by Hidden Fuzzy Markov Fields. The main problem of parameter estimation is solved using a recent general method of estimation regarding hidden data, called Iterative Conditional Estimation (ICE, [4]). This has been successfully applied in classical Hidden Markov Fields based segmentations ([8], [9]). The first part of our work involves estimating the parameters defining the Markovian distribution of the fuzzy picture without noise. We then combine this algorithm with the ICE method in order to estimate all the parameters of the noisy picture.

1. FUZZY MARKOVIAN MODEL

1.1 Distribution of the non noisy Markovian field.

S being the set of pixels, we will consider two random fields $X=(X_s)_{s\in S}$ and $Y=(Y_s)_{s\in S}$. The image to be segmented is the realisation $Y=y$ of Y and the picture looked for is the realisation $X=x$ of the field X. Hence, the values of Y are real numbers. We consider the case of two classes. In the usual case, which we will be called "hard" in what follows, X_s has the value of $\Omega=\{0,1\}$, where the numbers "0" and "1" correspond to the "hard" classes (for example the classes pure "wood" and pure "city"). In the fuzzy model presented in [3] we take $\Omega_f=[0,1]$, where the numbers "0" and "1" correspond to the "hard" classes and $]0,1[$ to the "fuzzy" classes. Otherwise, if $X_s=x_s\in]0,1[$, then x_s indicates the percentage of the class "1" in the value of the pixel. Therefore, $1-x_s$ is the percentage of the class "0". According to the model proposed in [3] the distribution of each X_s is represented by a density h according to the measure $\nu=\delta_0+\delta_1+\mu$. This includes a "hard" component (Dirac functions δ_0,δ_1 on $\{0,1\}$), and a "fuzzy" component which is the Lebesgue measure μ. on $]0,1[$ This "pixel by pixel" model, has been recently generalised by introducing the "Fuzzy Markov Random Fields" described in [1]. The distribution of X is represented by

$$P_X[x]=Ke^{-U_f(x)}.\nu^N$$

where $N=$ Card(S) represents the number of pixels and $h_f(x)=K.e^{-U_f(x)}$ is a density of an analogous form as in the usual hard case. The difference is that this density is with respect to the measure $\nu^N=(\delta_0+\delta_1+\mu)^N$. It is possible to show, as in the hard case, that this law defines a Markovian field.

$U_f(X)$ defines the Markovian energy of the field X. We consider the spatial markovianity of X relative to the eight nearest neighbours. We write this energy as a sum of functions defined on the single and double cliques:

$$U_f(X)=\sum_{x,\ \text{single}}\Phi(x_s)+\sum_{i=1}^{4}\sum_{x_s,x_t\ \text{neighbourghs}}\Psi_i(x_s,x_t)$$

The index i indicates the kind of the clique : we consider four kinds of two order cliques. $i=1,2,3,4$ correspond respectively to the "horizontal" neighbours, "vertical" neighbours, "east-north" and "west-north" neighbours for the diagonal directions. The function Φ indicates the proportions of the hard and fuzzy pixels:

$$\Phi(x_s)=\eta_0 \text{ if } x_s=0\ ;\ \Phi(x_s)=\eta_1 \text{ if } x_s=1.$$
$$\Phi(x_s)=\lambda \text{ if } x_s\in]0,1[\ \text{(fuzzy pixel)}$$

Otherwise, for instance, for i=1 we get the function:

for $(x_s,x_t)\in\{0,1\}^2$ (both pixels hard) :

$$\Psi_1(x_s,x_t)=\begin{cases}-\alpha^h_{hor} \text{ if } x_s=x_t\\ +\alpha^h_{hor} \text{ if } x_s\neq x_t\end{cases}$$

for $(x_s,x_t)\in[0,1]^2-\{0,1\}^2$ (one pixel is fuzzy) :

$$\Psi_1(x_s,x_t)=-\alpha^f_{hor}\left(1-2.|x_s-x_t|\right)$$

The formulas are similar for the other directions using the parameters $\alpha^h_{ver},\alpha^f_{ver},\alpha^h_{en},\alpha^f_{en},\alpha^h_{wn},\alpha^f_{wn}$.

Finally the distribution of X is defined by the set :

0-7803-2431-5/95 $4.00

$$\alpha = \left(\eta_0, \eta_1, \lambda, \alpha^h_{hor}, \alpha^f_{hor}, \alpha^h_{ver}, \alpha^f_{ver}, \alpha^h_{en}, \alpha^f_{en}, \alpha^h_{wn}, \alpha^f_{wn}\right)$$

1.2 Distribution of the noisy Markovian field.

The global segmentation of the noisy image Y requires a knowledge of α and the set of parameters denoted by β which is defines all conditional distributions $P[Y|X]$. In our study, Y is assumed to be Gaussian conditionally on X. As in the hard case, we assume :
(i) random variables Y_s are independent conditionally to X.
(ii) distribution of each Y_s conditional on X is equal to its distribution conditional on X_s.
The parameters of the noise on a fuzzy pixel depend linearly on the parameter of both "hard" classes. We denote $N(m, \sigma^2)$ the Gaussian law with the mean m and variance σ^2. Then:

$$P\left[Y_s = y_s | X_s = x_s\right] = N\left((1-x_s)m_0 + x_s m_1, (1-x_s)\sigma_0^2 + x_s \sigma_1^2\right)$$

The means (m_0, m_1) and the standard deviations (σ_0, σ_1), correspond to the class "0" and "1". So $\beta = (m_0, m_1, \sigma_0, \sigma_1)$. If we denote by f_{x_s} the previous density, the density f of (X,Y) with respect to the measure $\nu^N \otimes \mu^N$ is written :

$$f(x,y) = K.e^{-U_f(x)} \prod_{s \in S} f_{x_s}(y_s)$$

According to this result, it is possible to simulate realisations of X according to P_X^y, its distribution conditioned to $Y = y$ (called posterior distribution, [1]).

2. ICE PROCEDURE

ICE is a recent general method of estimation in the case of incomplete data [4] whose principle is as follows: we consider $\hat{\theta} = \hat{\theta}(X,Y)$ an estimation of $\theta = (\alpha, \beta)$ from the complete data (X,Y). Data X being unknown, we have to approach $\hat{\theta} = \hat{\theta}(X,Y)$ by a function of Y. The best approximation, as far as the mean square error is concerned, is the conditional expectation. By denoting E_n the conditional expectation based on the current parameter θ_n, ICE procedure is written:

(i) $\theta = \theta_0$ given

(ii) $\theta_{n+1} = E_n\left[\hat{\theta} | Y = y\right]$

The relation (ii) is not workable. In accordance with the law of large numbers, we use an approximate ICE:

$$\theta_{n+1} = \frac{1}{N_{ICE}} \sum_{i=1}^{N_{ICE}} \hat{\theta}(x_i, y)$$

where $x_1, \ldots, x_{N_{ICE}}$ are independent realisations of X according to P_X^y, with the current parameter θ_n, and y the realisation of Y (image to be segmented). We take for $\hat{\theta} = \hat{\theta}(X,Y)$ the estimation $\hat{\theta} = (\hat{\alpha}(X), \hat{\beta}(X,Y))$, where $\hat{\alpha}(X)$ is the stochastic gradient algorithm (see section 2.1) and $\hat{\beta}(X_i, Y)$ is given by the empirical means and variances (where $j = 0,1$) :

$$\hat{m}_j(x_i, y) = \frac{\sum_{s \in S} y_s . 1_{[x_{s,i} = \omega_j]}}{\sum_{s \in S} 1_{[x_{s,i} = \omega_j]}}$$

$$\hat{\sigma}_j^2(x_i, y) = \frac{\sum_{s \in S} \left(y_s - \hat{m}_j\right)^2 . 1_{[x_{s,i} = \omega_j]}}{\sum_{s \in S} 1_{[x_{s,i} = \omega_j]}}$$

2.1 Estimation of the parameter (α, β)

2.1.1 Estimation of α from the field X.

The first part of our work consists of evaluating of α, when the picture does not contain any noise. We take X a fuzzy field without noise. In this way, we adapted the stochastic gradient algorithm [7] to the fuzzy case. We denote by $U_f'(X)$ the gradient of $U_f(X)$ with respect to θ. The method is defined by:

(i) $\alpha_0, X = x_0$ given

(ii) $\alpha_{n+1} = \alpha_n + \frac{c}{(n+1)}\left[U_f'(x_{n+1}) - U_f'(x_0)\right]$

where x_{n+1} is a realisation of X simulated by the "fuzzy" Gibbs Sampler, using the current parameter α_n. The parameter c > 0 is a constant ensuring the convergence of the algorithm. If c is too small, we have to initialise α near the real value. So c must be large enough, if we want to ensure a fast convergence. In the first steps of the algorithm we must maintain constant the value of $c/(n+1)$, in order to readjust the initial value of α. We select $c = 1/N^2$.

2.1.2 Initialisation of α.

This method has a stochastic and a deterministic character. We have nerely to ensure at the initialisation that the order of size of α is the same as the real parameter. To initialise

α, we use the algorithm of Derin and Eliott [5] adapted to the fuzzy case.

2.2 ICE procedure using gradient algorithm and empirical moments.

We compute the ICE procedure applied to $\hat{\theta} = \hat{\theta}(X, Y)$ defined above, as follows :
(i) take θ_0 as an initial value of θ with the methods described in the previous sections. To initialise the means m_0 and m_1, we use the empirical method called the "cumulated histogram".
(ii) compute $\theta_{n+1} = (\alpha_{n+1}, \beta_{n+1})$ from $\theta_n = (\alpha_n, \beta_n)$ and $Y = y$ in the follwing way
simulate N_{ICE} realisations of X according to the posterior distribution corresponding to θ_n and $Y = y$. For each realisation x_i, we estimate the parameter $\alpha^i_{n+1}(x_i)$ and $\beta^i_{n+1}(x_i, y)$ by the empirical means and variances. Then taking the average of these values, perform the new parameter θ_{n+1} :

$$\alpha_{n+1} = \frac{1}{N_{ICE}} \sum_{i=1}^{N_{ICE}} \alpha^i_{n+1}(x_i)$$

$$\beta_{n+1} = \frac{1}{N_{ICE}} \sum_{i=1}^{N_{ICE}} \beta^i_{n+1}(x_i, y)$$

(iii) when the sequence $\theta_n = (\alpha_n, \beta_n)$ becomes steady, stop the estimation step and proceed to the segmentation.

3. SEGMENTATION

We use a segmentation method, which is an adaptation of the Maximum Posterior Mode (MPM, [6]), proposed in [1]. The case of the blind fuzzy segmentation was studied in [3]. We are interested in the laws of X_s conditional on Y which are the marginal posterior laws of the field X. For each s in S, we choose the class whose probability is maximum. The choice of the strategy depends on the loss function. Denote by $L(x_s, \hat{x}_s)$ this function with $\hat{x}_s$ the estimated pixel of x_s. If it is possible to define L, the Bayesian methods should minimise the expectation of $L(X_s, \hat{X}_s)$ conditional to $Y = y$,. This expectation can be estimated by the law of large numbers (this gives us the error of estimation):

$$E_y\left[L(X_s, \hat{X}_s)|Y\right] \approx \frac{1}{N} \sum_{s \in S} L(x_s, \hat{x}_s)$$

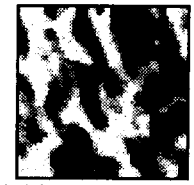
In any cases, we have to minimise the Bayesian risk, conditioned on $Y = y$. The minimisation of the global Bayesian risk corresponds to the minimisation of the "blind" Bayesian risk. With respect to the density $\mu \otimes \nu$. tihs risk is written:

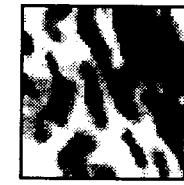
$$R(y,s) = L(0,s).h_y(0) + L(1,s).h_y(1) + \int_0^1 L(t,s).h_y(t,s).dt$$

with respect to the notations of the previous sections. We dispose of the three following method:
(i) $L(x_s, \hat{x}_s) = |x_s - \hat{x}_s|$: we compute the integral formulas by an approximation of $\int_0^1 L(t,s).h_y(t).dt$, taking $(s_0 = 0, s_1 = 1/n, \ldots, s_{n-1} = (n-1)/n, s_n = 1)$ for the discretisation.
(ii) $L(x_s, \hat{x}_s) = \begin{cases} 0 \text{ if } \hat{x}_s = x_s \\ 1 \text{ if } \hat{x}_s \neq x_s \end{cases}$: here we generalise the (0,1) discreet loss function of the hard case. this gives us the maximum likelihood method.
(iii) we dispose of a strategy, described in two steps called "adapted maximum likelihood", which is efficient comparable to the two others methods:
first step: we choose in the set $\{0, 1, F\}$ ("F" means fuzzy) according to the classical Bayesian law. The chosen element has to maximise the probability $h^y_s(0), h^y_s(1), 1 - h^y_s(0) - h^y_s(1)$.
second step: if this element is in $\{0,1\}$, we stop. Otherwise we choose in $]0,1[$ the element which maximise the restriction of h^y_s in this set.
This method takes better the fuzzy part of the field into account than the others ones. The problem is to find its corresponding loss function, because of the two steps. A loss function which could be an approach to measure the error of estimation is similar to the (0,1) function of the hard case, applied to the set $\{0, 1, F\}$, where we consider the fuzzy set as a class.

4. EXPERIMENTS AND RESULTS

Our first example shows the estimation of α using the fuzzy iterative method applied to a non noisy Markovian field, simulated by the Gibbs Sampler. Using the estimation of this parameter, we reconstruct the field. Visually the final field looks like the initial one.

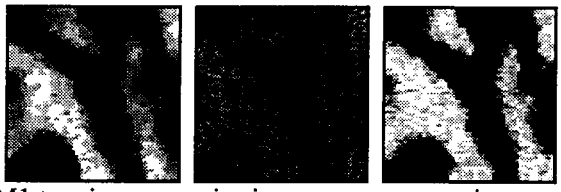

initial picture


reconstruction

We present below some examples which show the visual aspect of fuzzy Markov fields and the effectiveness of the unsupervised segmentation method proposed. We chosen the relative maximum posterior method. This algorithm gives visually a good aspect of the fuzzy segmented image. We present our results in the table below, which contains the initial and the estimated values of (α,β). Furthermore, we compare these segmentations with the supervised segmentations, according to the rates of errors (last line of the table). The components of the estimated α respect the directions of the picture. The method of segmentation is robust with respect to this parameter.

	IMAGE IM1		IMAGE IM2	
(α,β)	real val.	estim.	real val.	estim.
$\eta_0;\eta_1$	0; 0	−0. 07; −0; 05	0; 0	0. 02; 0. 02
λ	0	0. 12	0	−0. 04
$\alpha^h_{hor};\alpha^f_{ho}$	2. 0; 2. 5	0. 32; 0. 57	4. 0; 4. 5	0. 55; 0; 93
$\alpha^h_{ver};\alpha^f_{ver}$	2. 0; 2. 5	0. 67; 0. 91	4. 0; 4. 5	0; 95; 1; 2
$\alpha^h_{en};\alpha^f_{en}$	2. 0; 2. 5	0. 48; 0. 99	4. 0; 4. 5	0. 77; 1. 39
$\alpha^h_{wn};\alpha^f_{wn}$	2. 0; 2. 5	0. 13; 0. 48	4. 0; 4. 5	0. 69; 1. 16
$m_0;m_1$	1; 3	1. 14; 2. 82	1; 3	1. 05; 2. 97
$\sigma_0;\sigma_1$	1; 1	1. 06; 1. 02	1; 1	1. 06; 1. 04
error %	12.81	12.90	18.4	9.

table above : estimation of the noisy parameter.



IM1:true image noisy image segmentation

IM2:true image noisy image segmentation

5. CONCLUSION

We presented a new algorithm of unsupervised fuzzy image segmentation based on recent fuzzy hidden Markov field model, the iterative conditional estimation (ICE, [4]) and the stochastic gradient method [7]. This type of unsupervised segmentation was already studied in the case of the hard classes [9]. The estimation method gives us the relative component values of vector of the parameter. We studied other possible cases of noise (same means and different variances). Our segmentation method does not seem to be overly sensitive to the non noisy field parameters. However, the resulting pictures can be mainly degraded by a wrong estimation of the parameter β.. We were interested in the two class model. A model of fuzzy Markovian field concerning more than two classes would undoubtedly improve the fuzzy segmentation. In this way, to permit an easy estimation, we should add restrictive but realistic hypothesis: for instance to decide that the fuzzy pixels belong to maximum two classes.

6. REFERENCES

[1] Pieczynski W. - Cahen J.M. (1994), "Champs de Markov cachés flous et segmentation d'images," Revue de Statistique Appliquée, Vol. 42, No. 2, pp. 13-31.

[2] Kent J.T. - Mardia K.V. (1988), "Spatial Classification Using fuzzy Membership Models," IEEE Trans. on Patten Analysis and Machine Intelligence, Vol. 10, No. 5.

[3] Caillol H. - Hillion A. - Pieczynski W. (1993), "Fuzzy RandomFields and Unsupervised Image Segmentation," IEEE Trans. on Geoscience and Remote Sensing, Vol 31, No. 4, pp. 801-810.

[4] Pieczynski W. (1994), "Champs de Markov cachés et estimation conditionelle itérative," Traitement du Signal, Vol. 11, No. 2, pp. 141-153.

[5] Elliott H. - Derin H. (1987), "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 9.

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

[7] Younes L. (1988), "Estimation and annealing for Gibbsian fields," Annales de l'Institut Henri Poincaré, Vol. 24, No. 2, pp. 269-294.

[8] Braathen B. - Pieczynski W. - Masson P. (1993), "Global and local methods of unsupervised Bayesian segmentation of images," Machine Graphics & Vision, Vol. 2, No. 1, pp. 39-52.

[9] Allagnat O. - Boucher J.-M. - He D.C. - Pieczynski W. (1992), "Hidden Markov fields and unsupervised segmentation of images," IAPR 11th International Conference on Pattern Recognition (ICPR 92), Delft, Netherlands, september 92.