First International Conference on Multisource-Multisensor Information Fusion, July 6 - 9, 1998, Las Vegas, Nevada, USA , pp. 794-801.

Unsupervised Bayesian Fusion of Correlated Sensors

W. Pieczynski J. Bouvrais C. Michel
Département Signal et Image, Institut National des Télécommunications, 9, rue Charles Fourier, 91000 Evry, France.

Abstract - We address in this paper the problem of classifying of multidimensional data. We adopt the context of Bayesian unsupervised classification, so that our problem amounts to estimating a mixture of k components on R^m, where k is the number of classes and m is the number of sensors. When these components are Gaussian, one can use some general methods like Expectation-Maximization (EM) or Iterative Conditional Estimation (ICE). When the components are not Gaussian but the components of each of them are independent, one can still estimate such a mixture by the use of ICE or some stochastic variant of EM. We attack in this paper the more general problem of possibly correlated and non Gaussian sensors. A new method, called ICE-COR, of estimation of the corresponding mixture is presented and we provide some simulation results. The method proposed is inspired from a recent "generalized" mixture estimates, which means that we do not know, a priori, what the exact forms of the components are.

Key Words: Correlated sensor data classification, Bayesian classification, mixture estimation.

1. Introduction

The aim of this paper is to generalize some known solutions to the following problem. We are faced with m series of real data produced by m sensors. For each sensor $1 \le j \le m$ the data are denoted by $y_1^j, \ldots, y_n^j$, where n is the size of each series. We assume that for each point $1 \le s \le n$ the data $y_s^1, \ldots, y_s^m$ correspond to a certain class ω_i, among k classes $\omega_1, \ldots, \omega_k$, and the problem is to find which class it is. In other words, the problem is to classify each point $1 \le s \le n$ from the data available.
Solutions to this problem find many applications in economy, medicine, and signal or image processing, the latter being covered by the simulations we present below.

The probabilistic approach, which we shall adopt in this paper, consists in assuming that the class of the point $1 \le s \le n$ is a realization of a random variable X_s, and the data $y_s^1, \ldots, y_s^m$ produced by the m sensors are a realization of a random vector $Y_s = (Y_s^1, \ldots, Y_s^m)$. Finally, the problem is to estimate the unobservable realisations of a random process $X = (X_1, \ldots, X_n)$ from the observed realisation of a random process $Y = (Y_1, \ldots, Y_n)$. Different methods of such a statistical classification exist once the distribution $P_{(X,Y)}$ of (X,Y) is known: thus determinating the distribution $P_{(X,Y)}$ means doing the fusion of sensors.
In order to simplify things, let us temporarily assume that the random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent

and equidistributed, such that $P_{(X,Y)}$ is defined with $P_{(X_s,Y_s)}$, which is independent of $1 \le s \le n$.
Thus we treat here the case where $P_{(X_s,Y_s)}$ is not known and has to be estimated. $P_{(X_s,Y_s)}$ is generally assumed to be defined with the distribution of X_s, so called priors, which are $\pi_i = P[X_s = \omega_i]$, and the family of functions f_i, which are densities on R^m of the distribution of Y_s conditional to $X_s = \omega_i$. As the distribution of each Y_s is a mixture of k distributions on R^m, the problem of estimating $P_{(X_s,Y_s)}$ is sometimes called the "mixture estimation" problem. If the densities are assumed Gaussian, their parameters and the priors can be estimated by different methods like Expectation-Maximization (EM [7, 8, 12, 16, 18, 20,]), some stochastic approximations of EM [3, 5, 14, 16, 22], Iterative Conditional Estimation (ICE [2, 4, 6, 15, 16, 17, 21]), or stochastic gradient methods [23]. In the Gaussian case, sensors can be independent or not, which means that the random variables $Y_s^1,...,Y_s^m$ can be independent or not, the latter being considered conditional to X_s. Unfortunately, the f_i are not necessarily Gaussian in practice. Let us denote f_i^j the density of the distribution of the sensor j, conditional to the class ω_i. Considering independent sensors, which means that

$$f_i(y_s^1,...,y_s^m) = f_i^1(y_s^1)... f_i^m(y_s^m) \qquad (1.1)$$

we have proposed in [10] a quite general method allowing one to find the form of the km functions f_i^j, and estimate their parameters, once we know that the form of each f_i^j belongs to a given set of forms. We called such a mixture a "generalized" mixture, because there are numerous possibilities of classical mixtures and one has to determinate what case the data come from [6, 10]. For instance, the case of three classes and two independent sensors, in which each component can be exponential or Gaussian, leads to sixty-four possibilities of “classical” mixtures. Estimating such a mixture entails a supplementary difficulty: one must label, for each class and each sensor, the exact nature of the corresponding distribution. Thus the method proposed in [10] allows one to (i) identify the conditional distribution for each class and each sensor, (ii) estimate the unknown parameters in this distribution, and (iii) estimate priors.
In this paper we generalize the method proposed in [10] to the case of possibly non Gaussian or independent sensors (let us insist on the fact that the independence or the dependence of the sensors is always considered conditionally to the random process of classes X). The organization of the paper is as follows.
The assumptions needed and the general method proposed are presented in the next section. Section 3 is devoted to the particular case of hidden Markov fields and a simulation is provided. Section 4 contains some concluding remarks and perspectives for further work.

2. Generalized correlated sensors mixture estimation

We consider a stochastic process $X = (X_s)_{s \in S}$, with each X_s taking its values in a finite set of classes $\Omega = \{\omega_1,...,\omega_k\}$, and whose distribution P_X depends on a parameter α. The process X is not observed and one observes realizations of a process $Y = (Y_s)_{s \in S}$, such that each $Y_s = (Y_s^1,...,Y_s^m)$ takes its values in R^m. The random variables $(Y_s)_{s \in S}$ are assumed to be independent conditional to X, and the distribution of each Y_s conditional to X is equal to its distribution conditional to X_s. The random variables $Y_s^1,...,Y_s^m$ are not necessarily independent conditionally to X_s; however, we assume that there exist k triangular matrices $A_1,...,A_k$ such that for each $1 \le i \le k$, the components of $Z_s = A_i Y_s$ are independent conditionally to $X_s = \omega_i$. Furthermore, the form of each of

components of $Z_s = A_i Y_s$ is not known, but necessarily belongs to a family of forms $\Psi = \{F_1,...,F_M\}$.
We will admit the following hypotheses:

(A_1) An estimator $\hat{\alpha} = \hat{\alpha}(X)$ of α from X is available;
(A_2) One may simulate realizations of X according to its distribution conditional to Y;
(A_3) Each family F_j of $\Psi = \{F_1,...,F_M\}$ is characterized by a parameter β^j, i.e., $F_j = \{g_{\beta^j}\}_{\beta^j \in B^j}$. In practice B^j is a subset of R^{n_j} with n_j depending on F_j: for instance $n_j = 2$ if the F_j are Gaussian;
(A_4) M estimators $\hat{\beta}^1,...,\hat{\beta}^M$ are available such that if a sample $z = (z_1,...,z_r)$ is generated by a distribution g_{β^j} in F_j, then $\hat{\beta}^j = \hat{\beta}^j(z)$ estimates β^j;
(A_5) A decision rule D is available, such that for any sample $z = (z_1,...,z_r)$ and any $(g_1,...,g_M) \in F_1 \times ... \times F_M$, the rule D associates to z the "best suited", according to some criterion, density $g_1,...,g_M$.

Roughly speaking, the method we propose resembles the method proposed in [10], except that we use, at each iteration, some estimates of the matrices $A_1,...,A_k$ in order to "decorrelate" the sensors. Of course, the matrices $A_1,...,A_k$ are not known and thus they represent the additional parameters to be estimated, with respect to our previous work.
The method we propose, called ICE-COR (COR for "correlated") is an iterative method and runs as follows. After having initialized the procedure by some algorithm well suited to a given particular situation, we have to calculate, at each iteration q, the next value α^{q+1} of the parameter α, and the next probability densities $f_1^{q+1},..., f_k^{q+1}$ from the observation $Y = y$ and current value α^q and current densities $f_1^q,..., f_k^q$.

The run of each iteration is:

(a) Simulate x^q, a realization of X, according to its α^q and $f_1^q,..., f_k^q$ based distribution conditional to $Y = y$;
(b) Calculate $\alpha^{q+1} = E_q[\hat{\alpha}(X) | Y = y]$, where $E_q[\,.\,| Y = y]$ denotes the conditional expectation given $\alpha = \alpha^q$ and $(f_1,..., f_k) = (f_1^q,..., f_k^q)$. If this calculation is impossible, calculate $\alpha^{q+1} = \hat{\alpha}(x^q)$;
(c) Consider $S_i^q = \{s \in S \,/\, x_s^q = \omega_i\}$ for each $i = 1,...,k$. Let $y_i^q = (y_s)_{s \in S_i^q} = (y_s^1,...,y_s^m)_{s \in S_i^q}$ and $y_i^{q,r} = (y_s^r)_{s \in S_i^q}$. For each $i = 1,...,k$ calculate, from $y_i^q = (y_s)_{s \in S_i^q}$, the empirical covariance matrix $\hat{\Gamma}_i^q$ and consider a triangular matrix A_i^q such that $A_i^q \hat{\Gamma}_i^q (A_i^q)^T$ is diagonal (we assume all $\hat{\Gamma}_i^q$ are diagonalizable). For each $s \in S_i^q$, put $z_s = A_i^q y_s$ and consider $z_i^q = (z_s)_{s \in S_i^q}$. Remember that $z_i^q = (z_i^{q,1},...,z_i^{q,m})$;
(d) For each $r = 1,...,m$ and each class $i = 1,...,k$, calculate M parameters $\beta_i^{1,r} = \hat{\beta}^1(z_i^{q,r})$, ..., $\beta_i^{M,r} = \hat{\beta}^M(z_i^{q,r})$, which give the densities $g_i^{1,r,q+1},...,g_i^{M,r,q+1}$. Use the decision rule D to determinate $g_i^{j,r,q+1}$, the best suited, among the densities $g_i^{1,r,q+1},...,g_i^{M,r,q+1}$, to the sample $(z_i^{q,r})$.
(e) Put
$$f_i^{q+1}(y_s^1,...,y_s^m) = g_i^{j,1,q+1}(z_s^1) \times ... \times g_i^{j,m,q+1}(z_s^m).$$
(recall that $z_s = A_i^q y_s$).

Finally, the algorithm above allows us to estimate the parameters which define the prior distribution of X and choose the k distributions (k densities $f_1,..., f_k$ on R^m) in the set of all distributions of random vectors which are linear combinations of random vectors having independent components and such that the form of each component is in a known set of forms. Concerning the estimation of priors, the method above can be applied in a wide range of situations; in particular, it covers the modelling by hidden Markov chains and hidden Markov fields.

Remark

Let us consider the following problem of image segmentation: we have two sensors and two classes and we do not wish use a Markovian model. We desire to classify each pixel s from the observation of Y_s and of Y_t, where t is a neighbor of s. Such a segmentation is called "local" segmentation. It becomes feasible once the distribution of (X_s, Y_s, Y_t) is known, and the latter distribution is given by the distribution of (X_s, X_t, Y_s, Y_t). Recalling that there are two sensors, we have to determine the distribution of $(X_s, X_t, Y_s^1, Y_s^2, Y_t^1, Y_t^2)$. Putting $X_s^* = (X_s, X_t)$ and $Y_s^* = (Y_s^1, Y_s^2, Y_t^1, Y_t^2)$, we see that the distribution of Y_s^* is a mixture distribution of four components on R^4. So this is mathematically equivalent to having four classes and four sensors, and thus the whole procedure above can be applied. Of course, this can be generalized provided there are not too many neighbors considered. We have compared in [2] such local methods with Markovian methods and it turns out that in certain particular situations local methods are competitive. Thus the study of local methods with ICE-COR, which would generalize the Gaussian case study described in [14], could undoubtedly be of interest in some special situations.

3. Simulation results

We present in this section some results concerning the case of two classes ($k = 2$) and two sensors ($m = 2$). We consider a hidden Markov field, with application to unsupervised image segmentation. We focus on the interest of taking the sensor correlation into account in unsupervised image segmentation. Although the estimation of the parameter α which defines the energy of the Markov field we use does not pose a problem, we keep it fixed, to better specify the interest we focus on. According to the general modelling described on the previous section, an observation is thus a realization of a random process $Y = \begin{bmatrix} Y^1 \\ Y^2 \end{bmatrix}$, whose distribution conditionally to X by the two distributions of $Y_s = \begin{bmatrix} Y_s^1 \\ Y_s^2 \end{bmatrix}$ conditional on $X_s = \omega_1$ and $X_s = \omega_2$, respectively. Furthermore, there exist two matrices $A_1 = \begin{bmatrix} 1 & 0 \\ a_1 & 1 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & 0 \\ a_2 & 1 \end{bmatrix}$, such that the components of the random vector $Z_s = A_1 Y_s$ are independent conditionally to $X_s = \omega_1$, and the components of the random vector $Z_s = A_2 Y_s$ are independent conditionally to $X_s = \omega_2$.

Let us denote by

$f_1(z) = f_1(z^1, z^2) = f_1^1(z^1) f_1^2(z^2)$

and

$f_2(z) = f_2(z^1, z^2) = f_2^1(z^1) f_2^2(z^2)$

the densities of the distribution of $Z_s = A_1 Y_s$ conditional to $X_s = \omega_1$, and $Z_s = A_2 Y_s$ conditional to $Z_s = A_2 Y_s$, respectively. The densities of the distribution of Y_s conditional to $X_s = \omega_1, \omega_2$, respectively, are then $f_1(A_1 y)$, $f_2(A_2 y)$. We consider the case where each of the densities f_1^1, f_1^2, f_2^1, f_2^2 can be exponential or Gaussian. An exponential density is of the form $f(z) = be^{-b(z-a)} 1_{[a,+\infty[}(z)$, and thus depends on two parameters, which can easily be determined from the mean and the variance. Finally, we have a sample $(y_1, y_2, ..., y_n) = (\begin{pmatrix} y_1^2 \\ y_1^1 \end{pmatrix}, \begin{pmatrix} y_2^2 \\ y_2^1 \end{pmatrix}, ..., \begin{pmatrix} y_n^2 \\ y_n^1 \end{pmatrix})$ and we must:

(i) Identify the forms of f_1^1, f_1^2, f_2^1, f_2^2 and estimate their parameters;,

(ii) Estimate a_1, a_2

The algorithm is as follows:

1. Initialization:
Assume the sensors independent (matrices A_1 and A_2 are identity) and all densities Gaussian. Calculate, from $y_1^1, y_2^1, ..., y_n^1$, the empirical mean and variance M_0^1, Σ_0^1 of first sensor, and M_0^2, Σ_0^2 the empirical mean and variance of the second one. Put $m_1^1 = M_0^1 - \frac{\Sigma_0^1}{2}$ and $m_2^1 = M_0^1 + \frac{\Sigma_0^1}{2}$ for means of f_1^1, f_2^1, and $(\sigma_1^1)^2 = (\sigma_2^1)^2 = \Sigma_0^1$ for their variances. Proceed in the same way to calculate $m_1^2, m_2^2, (\sigma_1^2)^2, (\sigma_2^2)^2$, the means and variances of f_1^2, f_2^2.

2. At each iteration

2. a) Simulate a realization x^q of X according to its distribution conditional to $Y = y$, by the use of the Gibbs sampler.

2. b) Calculate, from S_1^q et S_2^q, the empirical covariance matrices $\hat{\Gamma}_1^q$, $\hat{\Gamma}_2^q$. Calculate A_1^q, A_2^q (take $A = \begin{bmatrix} 1 & 0 \\ -\frac{\rho}{\sigma_1^2} & 1 \end{bmatrix}$ for $\Gamma = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}$);

2. c) Consider

$$z_1 = \begin{pmatrix} z_s^2 \\ z_s^1 \end{pmatrix}_{s \in S_1^q} = A_1^q \begin{pmatrix} y_s^2 \\ y_s^1 \end{pmatrix}_{s \in S_1^q},$$

$$z_2 = \begin{pmatrix} z_s^2 \\ z_s^1 \end{pmatrix}_{s \in S_2^q} = A_2^q \begin{pmatrix} y_s^2 \\ y_s^1 \end{pmatrix}_{s \in S_2^q}$$

and use the samples $(z_s^1)_{s \in S_1^q}$, $(z_s^2)_{s \in S_1^q}$, $(z_s^1)_{s \in S_2^q}$, et $(z_s^2)_{s \in S_2^q}$ to identify the forms of $g_1^1, g_1^2, g_2^1, g_2^2$ and estimate their parameters. The latter is done as follows: estimate the mean and the variance from $(z_s^1)_{s \in S_1^q}$, which gives a Gaussian density h^1 on the one hand, and an exponential density h^2 on the other. Calculate the histogram $\hat{h}$ from $(z_s^1)_{s \in S_1^q}$ and consider $d_i = \int_R [h^i(z) - \hat{h}(z)]^2 dz$ for $i = 1, 2$. Put $g_1^1 = h^1$ if $d_1 \leq d_2$ and $g_1^1 = h^2$ if $d_1 \geq d_2$. Proceed in the same way for g_1^2, g_2^1, and g_2^2.

2. d) Determine the densities f_1, f_2 (recall that for $g_i(z) = g_i^1(z^1) g_i^2(z^2)$ we have $f_i(y) = g_i(A_i z)$ Calculate the posterior distribution.

We present in Table 1 the results of two cases studied. In the first one, we consider two Gaussian densities and two exponential densities, and in the second one we consider three Gaussian densities and one exponential density. When applying the MPM method based on the real parameters we obtain the error ratio $\tau = 0.65\%$, which means that the image is not very noisy. However, the parameter restimation is not so easy; in particular, the error ratio obtained after estimation of the correlated generalized mixture is of $\tau_1 = 25.25\%$. Now, if we do not take the correlation into account, the result $\tau_2 = 38.00\%$ is still worse, and could well be described as disastrous. Thus taking the correlation into account can be presented as keeping some interest in the case studied. The second case is not very noisy either, as the real parameter segmentation error ration is $\tau = 1{,}00\%$. When applying the correlated sensors generalized mixture estimation method we propose, this ratio becomes $\tau_1 = 3.03\%$, which can be seen as a good result. Not taking the correlation into account leads to the error ratio of $\tau_2 = 46.27\%$, which is really quite poor compared to $\tau_1 = 3.03\%$. The different images corresponding to the case 2 are presented in Figure 1, and, in fact, one can notice that it is very important, visually, to take the correlation into account.
These two simple examples show that taking the sensor correlation into account is of interest. In some situations the improvement can be moderate, but, in some others, it can be quite impressive.

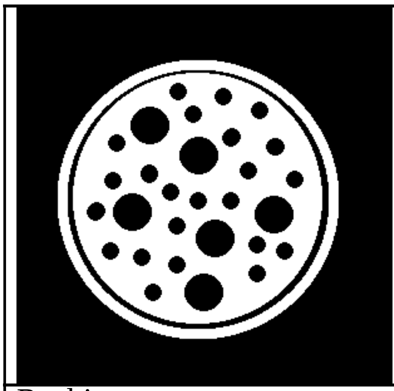
Table 1

GEEG: g_1^1 and g_2^2 Gaussian, g_1^2 and g_2^1 Exponential. ρ_1, ρ_2: correlation's in Γ_1, Γ_2. m_1^1, m_1^2, m_2^1, and m_2^2: means of g_1^1, g_1^2, g_2^1 and g_2^2, respectively. Variance of the four distribution equal to 1. τ_1 and τ_2: error rates of Bayesian classification without (τ_1) and with (τ_2) taking the correlation into account. τ : real parameters based segmentation.

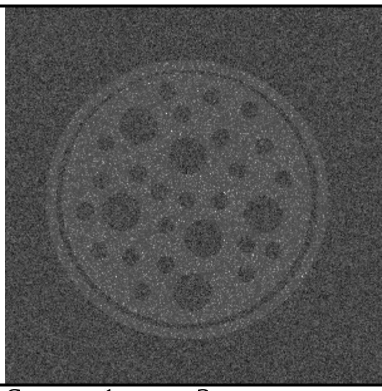
Case	Laws	ρ_1	ρ_2	m_1^1	m_2^1	m_1^2	m_2^2	τ_1	τ_2	τ
1	GEEG	0.8	0.0	-0.5	0.5	0.7	-0.7	38.00%	25.25%	0.65%
2	GEGG	0.8	0.8	-0.5	0.5	0.7	-0.7	46.27%	3.03%	1.00%

Figure 1

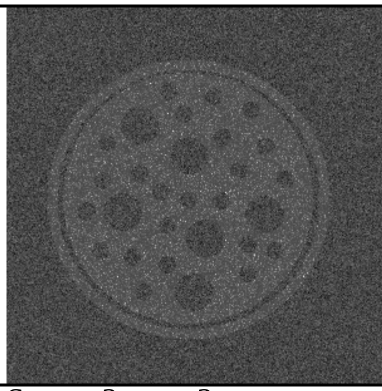
Visual aspect of the segmentations corresponding to the case 2, Table 1.



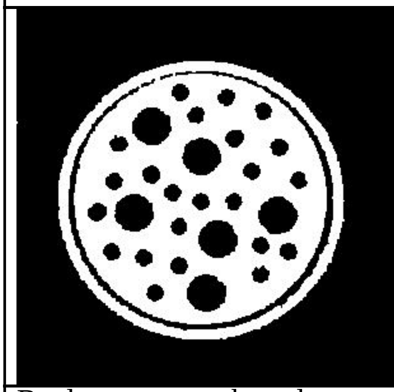
Real image



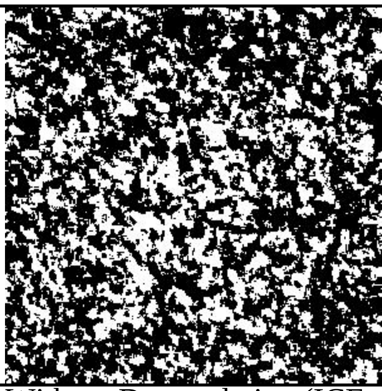
Sensor 1, case 2



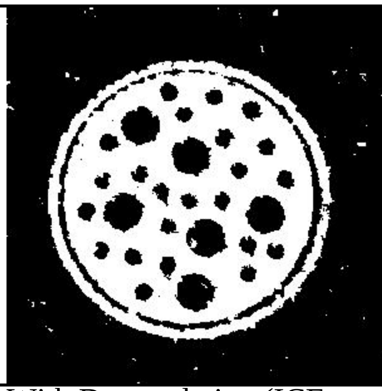
Sensor 2, case 2



Real parameters based segmentation, case 2, $\tau = 1.00\%$



Without Decorrelation (ICE-GEMI): $\tau = 46.27\%$



With Decorrelation (ICE-COR): $\tau = 3.03\%$

4. Conclusions

In this paper we presented a new unsupervised Bayesian fusion of correlated sensors, called ICE-COR. Unsupervised Bayesian fusion means that the joint distribution of the observed and the hidden data is previously estimated in some way. Focusing on the Bayesian multisensor classification in this paper, the latter estimation problem is a mixture estimation problem. Thus we have presented a new method of multisensor mixture estimation, whose originality is that the sensors are not necessarily Gaussian, and the form of the noise can vary with the class and the sensor.
The method presented generalizes the method proposed in [10], in which the form of the noise was allowed to vary with the class and the sensor, but was only valid in the case of independent sensors.
The method proposed is valid in a rather general setting; in particular, hidden Markov fields or hidden Markov chains can be treated, with known applications to image or signal restoration.
Application of the method proposed to the restoration of real processes is a natural direction for further work. In particular, following the simple simulations presented above, we hope to develop some applications of multisensor hidden Markov field models to unsupervised real world image segmentation.

References

[1] J. Besag, On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society*, Series B, 48, pp. 259-302, 1986.

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

[3] M. Broniatowski, G. Celeux, J. Diebolt, Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabiliste, Data Analysis an Informatics 3, E. Diday (Ed.), North Holland, Amsterdam, 1983.

[4] H. Caillol, W. Pieczynski, and A. Hillon, Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation, *IEEE Transactions on Image Processing*, Vol. 6, No. 3, pp. 425-440, 1997.

[5] R. Chellapa, A. Jain Ed., Markov Random Fields, Theory and Application, *Academic Press, San Diego*, 1993.

[6] Y. Delignon, A. Marzouki, and W. Pieczynski, Estimation of Generalized Mixture and Its Application in Image Segmentation, *IEEE Transactions on Image Processing*, Vol. 6, No. 10, pp. 1364-1375, 1997.

[7] J. P. Delmas, An extension to the EM algorithm for exponential family, *IEEE Transactions on Signal Processing*, Vol. 45, No. 10, pp. 2613-2615, 1997.

[8] M.M. Dempster, N.M. Laird, and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, Series B, 39, pp. 1-38, 1977.

[9] S. Geman, G. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on PAMI*, Vol. 6, No. 6, pp. 721-741, 1984.

[10] N. Giordana and W. Pieczynski, Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 465-475, 1997.

[11] S. Lakshmanan, H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields, *IEEE Transactions on PAMI*, Vol. 11, pp. 799-813, 1989.

[12] T. Krishnan, EM algorithm in tomographiy: a review and a bibliography, Bulletin of Informatics and Cybernetics, Vol. 27, No. 1, pp. 5-22, 1995.

[13] J. Marroquin, S. Mitter, T. Poggio, Probabilistic solution of ill-posed problems in

computational vision, *Journal of the American Statistical Association*, 82, pp. 76-89, 1987.

[14] P. Masson et W. Pieczynski, SEM algorithm and unsupervised statistical segmentation of satellite images, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 34, No. 3, pp. 618-633, 1993.

[15] M. Mignotte, C. Collet, P. Pérez, et P. Bouthemy, Unsupervised segmentation applied on sonar images, Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science : 1223, Springer-Verlag, Berlin, pp.491-506, 1997.

[16] A. Peng, W. Pieczynski, Adaptive Mixture Estimation and Unsupervised Local Bayesian Image Segmentation, *Graphical Models and Image Processing*, Vol. 57, No. 5, pp. 389-399, 1995.

[17] W. Pieczynski, Statistical image segmentation, *Machine Graphics and Vision*, Vol. 1, No. 1/2, pp. 261-268, 1992.

[18] W. Qian, and D. M. Titterington, Stochastic relaxations and EM algorithms for Markov random fields, J. Statist. Comput. Simulation, 40, pp. 55-69, 1992.

[19] L.R. Rabiner (1989). A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of IEEE*, Vol. 77, No. 2, pp. 257-286.

[20] R.A. Redner and H.F. Walker, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, 26, pp. 195-239, 1984.

[21] F. Salzenstein and W. Pieczynski, Parameter Estimation in Hidden Fuzzy Markov Random Fields and Image Segmentation, *Graphical Models and Image Processing*, Vol. 59, No. 4, pp. 205-220, 1997.

[22] M. A. Tanner, Tools for statistical inference: observed data and data augmentation methods, Lecture Notes in Statistics67, Springer-Verlag, New York, 1991.

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

[24] M. C. Zhang, R. M. Haralick, and J. B. Campbell, Multispectral image context classification using stochastic relaxation, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 20, No. 1, pp. 128-140, 1990.