Advanced Concepts for Intelligent Vision Systems (ACIVS 04), Aug. 31-Sept. 3, Brussels, Belgium, 2004

PARAMETER ESTIMATION IN PAIRWISE MARKOV FIELDS

Dalila Benboudjema and Wojciech Pieczynski

Dalila.Benboudjema@int-evry.fr

GET/INT, Département. CITI, CNRS UMR 5157, 9, rue Charles Fourier, 91000 Evry, France

ABSTRACT

Hidden Markov fields (HMF), which are widely applied in various problems arising in image processing, have recently been generalized to Pairwise Markov Fields (PMF). Although the hidden process is no longer necessarily a Markov one in PMF models, they still allow one to recover it from observed data. We propose in this paper two original methods of parameter estimation in PMF, based on general Stochastic Gradient (SG) and Iterative Conditional Estimation (ICE) principles, respectively. Some experiments concerning unsupervised image segmentation based on Bayesian Maximum Posterior Mode (MPM) are also presented.

1. INTRODUCTION

Hidden Markov fields (HMF) are widely used in various problems comprising two stochastic processes $X=(X_s)_{s\in S}$ and $Y=(Y_s)_{s\in S}$, in which $X=x$ is unobservable and must be estimated from the observed $Y=y$. The interest of HMF is based on the following property. When X is a Markov field and when the distributions $p(y|x)$ of Y conditional on $X=x$ are simple enough, the pair (X,Y) retains the Markovian structure, and likewise for the distribution $p(x|y)$ of X conditional on $Y=y$. Then the Markovianity of $p(x|y)$ allows one to estimate the unobservable $X=x$ from the observed $Y=y$, even in the case of very rich sets S. However, the simplicity of $p(y|x)$ required to ensure the Markovianity of $p(x|y)$ can pose problems, in particular in textured images segmentation [10]. To remedy this, Pairwise Markov fields (PMF), in which one directly assumes the Markovianity of (X,Y), have been proposed [12]. Both $p(y|x)$ and $p(x|y)$ are then Markovian, the former ensuring possibilities to model textures without approximations, and the latter allowing Bayesian processing. The aim of this paper is to propose two original PMF parameter estimation methods and test their efficiency in the unsupervised image segmentation context. The first one is based on the general Stochastic Gradient (SG) principle, and the second one is based on the general Iterative Conditional Estimation (ICE) principle.

The organization of the paper is as follows. PMF are shortly recalled in the next section. Section 3 is devoted to the parameter estimation and some unsupervised image segmentation results are presented in Section 4. The last Section 5 contains conclusions and perspectives.

2. PAIRWISE MARKOV FIELDS

The random pairwise field $Z=(X,Y)$ is a PMF when its distribution is a Gibbs distribution, given by

$$p(z)=\gamma\exp[-\sum_{c\in C}\varphi_c(z_c)] \tag{1}$$

The simplest PMF is the HMF with Independent Noise (HMF-IN), in which X is a Markov field, and the conditional distribution $p(y|x)$ verifies :

(H1) $p(y_s|x)=p(y_s|x_s)$ for each $s\in S$;

(H2) $p(y|x)=\prod_{s\in S}p(y_s|x)$.

(H2) means that the random variables $(Y_s)_{s\in S}$ are independent conditionally on X. The distribution of (X,Y) is then :

$$p(x,y)=\gamma\exp\left[-\sum_{c\in C}\varphi_c(x_c)+\sum_{s\in S}Log[p(y_s|x_s)]\right] \tag{2}$$

HMF-IN models are the most currently used ones. A HMF is a model such that both (X,Y) and X are Markov fields. This is a more general model than HMF-IN; in particular, (H2) can be relaxed. Finally, PMF is more general than HMF because in PMF X is not necessarily a Markov field.

Let us consider a Gaussian PMF (X,Y), with $p(x,y)\propto\exp[-U(x,y)]$ and

$$U(x,y)=\frac{1}{2}\left[\sum_{s\in S}a_{x_s}(y_{x_s}-m_{x_s})^2+\sum_{(s,t)\in C}[\alpha(1-2\delta(x_s,x_t))+q_{x_sx_t}(y_{x_s}-m_{x_s})(y_{x_t}-m_{x_t})]\right] \tag{3}$$

(where $\delta(x_s,x_t)=1$ for $x_s=x_t$ and $\delta(x_s,x_t)=0$ for $x_s\neq x_t$). We see that both $p(x|y)$ and $p(y|x)$ are Markov, and the latter is Gaussian.

Let us see how to calculate $p(z_s | z_{V_s})$, which is useful in sampling of the PMF (X,Y). We have $p(z_s | z_{V_s}) = p(x_s, y_s | x_{V_s}, y_{V_s}) =$ $p(x_s | x_{V_s}, y_{V_s}) p(y_s | x_s, x_{V_s}, y_{V_s})$, with

$$p(x_s | x_{V_s}, y_{V_s}) \propto \frac{1}{\sqrt{a_{x_s}}} \exp[-\frac{1}{2}(\sum_{t \in V_s} \alpha(1 - 2\delta(x_s, x_t)) - \frac{1}{4a_{x_s}}(\sum_{t \in V_s} q_{x_s x_t}(y_t - m_{x_t}))^2)] \quad (4)$$

and $p(y_s | x_s, x_{V_s}, y_{V_s})$ Gaussian with mean and variance

$$M_{x_s} = m_{x_s} - \frac{\sum_{t \in V_s} q_{x_s x_t}(y_t - m_{x_t})}{2a_{x_s}}, \ \Sigma^2_{x_s} = \frac{1}{a_{x_s}} \quad (5)$$

Let us notice that $p(x_s | x_{V_s}, y)$ are in HMF-IN proportional to $\exp[-\sum_{c \subset V_s} p(x_s | x_c, y_s)]$, while in PMF they are proportional to $\exp[-\sum_{c \subset V_s} p(x_s | x_c, y_s, y_c)]$, which is much richer.

Remark 1

Introducing a third, possibly latent, random field $U = (U_s)_{s \in S}$, one can consider a Triplet Markov Field $T = (X, U, Y)$ (TMF [14]), whose distribution is given by (1), with $t = (x, u, y)$ instead of $z = (x, y)$. If each U_s takes its values in a finite set $\Lambda = \{\lambda_1, ..., \lambda_m\}$, we can write $p(x_s = \omega_j | y) = \sum_{\lambda \in \Lambda} p(x_s = \omega_j, u_s = \lambda | y)$. Otherwise, $V = (X, U)$ is Markovian conditionally on $Y = y$, which implies that $p(v_s = (\omega_i, \lambda_j) | y)$ can be classically estimated from sampled values of $V = (X, U)$. Finally, in TMF $p(x_s = \omega_j | y)$ are calculable and thus Bayesian MPM method is workable. How to estimate the parameters of TMF? As TMF $T = (X, U, Y)$ also is a PMF $T = (V, Y)$, the parameter estimation problem in TMF is the same that the parameter estimation problem in PMF we deal with in this paper.

3. PARAMETER ESTIMATION

The parameter estimation in HMF is a difficult one and the implementation of the popular "Expectation-Maximization" (EM) method poses problems [9]. So, some alternative methods have been proposed [1, 4, 5]. Here we propose two new methods valid in PMF : the first one is an adaptation to PMF of the Stochastic Gradient successfully applied in HMF [16]. The second one is of Iterative Conditional Estimation (ICE) kind, which also has already been successfully used in different HMF based problems [2, 7, 11, 13, 15]. ICE resembles EM, and a comparative study can be seen in [3]

We will specify the different methods in a simple particular Gaussian PMF defined by (3), with the neighborhood system reduced to four nearest neighbors. On the one hand, its generalization does not pose problems and, on the other hand, such models, on which are based the experiments presented below, are generally sufficient to practical applications.

So, for k classes $\Omega = \{\omega_1, ..., \omega_k\}$, the vector of parameters θ to be estimated contains α, k components a_i and m_i, $1 \le i \le k$ and k^2 components q_{ij}, $1 \le i, j \le k$.

3.1. Stochastic Gradient (SG)

Adapting SG proposed in [16] to PMF we put :

- Initialize parameter vector $\theta = \theta_0$;
- Calculate θ_{n+1} from θ_n and $Y = y$ by:

$$\theta_{n+1} = \theta_n + \frac{K}{n+1}[\nabla_{\theta_n} U(x_{n+1}, y_{n+1}) - \nabla_{\theta_n} U(x^*_{n+1}, y)] \quad (6)$$

where $\nabla_{\theta_n} U(x_{n+1}, y_{n+1})$ is the value at the point θ_n of the $U(x_{n+1}, y_{n+1})$'s gradient with respect to θ, (x_{n+1}, y_{n+1}) is a realization of (X, Y) simulated by Gibbs sampler using θ_n, x^*_{n+1} is a realization of X simulated by Gibbs sampling according to $p(x|y)$ based on θ_n, and K is a constant.

Adapting (6) to the Gaussian PMF defined by (3), we obtain :

$$\frac{\partial U}{\partial a_i} = 0.5 \sum_{s \in S} (y_s - m_i)^2 1_{[x_s = i]}$$
$$\frac{\partial U}{\partial \alpha} = 0.5 \sum_{(s,t) \in C} [1 - 2\delta(x_s, x_t)]$$
$$\frac{\partial U}{\partial q_{ij}} = \sum_{(s,t) \in C} (y_s - m_i)(y_t - m_j) 1_{[x_s = i]} 1_{[x_t = j]} \quad (7)$$
$$\frac{\partial U}{\partial m_i} = \sum_{s \in S} a_i (m_i - y_i) 1_{[x_s = i]} + \sum_{(s,t) \in C} q_{ij}(m_j - y_t) 1_{[x_s = i]} 1_{[x_t = j \neq i]} + \sum_{(s,t) \in C} q_{ii}(2m_i - y_s - y_t) 1_{[x_s = i]} 1_{[x_t = i]}$$

Let us notice that when a learning sample is known, it is to say when $X = x$ is observed, Stochastic Gradient can still be used replacing in (6) the simulated x^*_{n+1} (which varies with iterations) by the fixed observed $X = x$.

Such a method will be called SG from complete data (SGCD), while the new SG above will be called SG from incomplete data (SGID).

3.2. Iterative Conditional Estimation (ICE)

ICE is based on the following assumptions : (i) there exists an estimator of θ from the complete data: $\hat{\theta} = \hat{\theta}(X, Y)$; (ii) for each $\theta \in \Theta$, either the conditional expectation $E_\theta[\hat{\theta}(X,Y)|Y = y]$ is computable, or simulations of X according to its distribution conditional to $Y = y$ are feasible.

It is an iterative method which runs as follows :

1. Initialize $\theta = \theta_0$;
2. for $q \in N$,

- put $\theta_{q+1} = E_{\theta_q}[\hat{\theta}(X,Y)|Y = y]$ if the conditional expectation is computable;
- if not, simulate l realizations $x^1,...,x^l$ of X (each x^i is a class image) according to its distribution conditional to $Y = y$ and based on θ_q and put $\theta_{q+1} = \frac{\hat{\theta}(x^1, y) + ... + \hat{\theta}(x^l, y)}{l}$.

Let us notice that it can occur that the conditional expectation is computed for some components of θ, and is approximated for the other, where the exact computation is not feasible. Such cases occur in Hidden Markov Chains [6] or still in spatially independent data [13].

As in PMF the distribution of X conditional on $Y = y$ is a Markov distribution, its simulations are feasible, and thus the condition (ii) is always verified. So, we only need an estimator $\hat{\theta} = \hat{\theta}(z)$. The new method we propose is mainly based on least squares method proposed by Derin and Elliott [4] to estimate the parameters α, and on the use of conditional distributions to estimate the remaining model parameters, relative to Gaussian distribution of Y conditionally on X. The method below can be either directly used when a learning sample of (X,Y) is available, or inside of ICE, when not.

Although the proposed method can be extended to higher order neighborhood systems, we present it in the case of simple four-nearest neighbors, as shown in Figure 1.

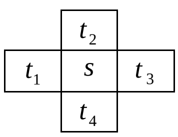

Figure 1. *Set W_s containing pixel s and its' 4 neighbors.*

Let us consider a sequence of sets of pixels W_1, …, W_n, where each W_i is of the form presented in Figure 1. Parameters $\alpha = [\beta_1, ..., \beta_k, \alpha_1, \alpha_2]$, $q = [q_{ij}]_{1 \le i, j \le k}$, $a = [a_i]_{1 \le i \le k}$ and $m = [m_i]_{1 \le i \le k}$ are then estimated from W_1, …, W_n, in the following way.

Estimation of α

The procedure, which is strictly the same that the Derin *et al.* method, consists of the following steps:

(i) Find the relationship between the joint probability $p(x_s, x_{V_s})$ and the parameter α.

(ii) Use histogram techniques to estimate all such probabilities

(iii) Construct an over-determined system of equations in terms of the probabilities and parameters. This system of linear equation is given by:

$$\left[\Phi(\omega_i, x_{V_s}) - \Phi(\omega_j, x_{V_s})\right]^T \alpha = Log\left(\frac{p(\omega_j, x_{V_s})}{p(\omega_i, x_{V_s})}\right) \qquad (8)$$

where

$\Phi(\omega, x_{V_s}) = [J_1(\omega), \cdots, J_k(\omega), I(\omega, x_{t_1}) + I(\omega, x_{t_3}),$
$I(\omega, x_{t_2}) + I(\omega, x_{t_4})]^T$

with

$$I(\omega, x_{t_1}, \cdots, x_{t_n}) = \begin{cases} -1 & if \;\; \omega = x_{t_1} = \cdots = x_{t_n} \\ +1 & \text{otherwise} \end{cases} \qquad (9)$$

and

$$J_k(\omega) = \begin{cases} +1 & if \;\; \omega = \omega_k \\ 0 & \text{otherwise} \end{cases} \qquad (10)$$

(iv) Solve the over-determined system using the Least Squares (LS) method.

Estimation of $q = [q_{ij}]_{1 \le i, j \le k}$, $a = [a_i]_{1 \le i \le k}$ and $m = [m_i]_{1 \le i \le k}$

As above, we consider a sequence of sets of pixels W_1, .., W_n as in Figure 1, centered on pixels 1, .., n. Let us denote by Y_{W_i} and X_{W_i} the restriction of Y and X to W_i. So, we have $Y_{W_i} = \begin{bmatrix} Y_i \\ Y_{V_i} \end{bmatrix}$ and $X_{W_i} = \begin{bmatrix} X_i \\ X_{V_i} \end{bmatrix}$, where V_i contains the four neighbors of the pixel i. The idea of the estimator is the following. For each given configuration x_W, which is a possible realization of X_{W_1}, .., X_{W_n}, let m_{x_W} and Γ_{x_W} be the mean vector and the variance-covariance matrix of the distribution of Y_W conditional on $X_W = x_W$. So, for a given x_W, we can consider in the sequence W_1, .., W_n a sub-sequence W_1', .., W_m' such that $x_{W_1'} = ... x_{W_m'} = x_W$ and use $y_{W_1'}$, .., $y_{W_m'}$ to classically estimate m_{x_W} and Γ_{x_W} by:

$$\hat{m}_{x_W} = \frac{1}{m}\sum_{j=1}^{m} y_{W_{j'}} , \quad (11)$$

$$\hat{\Gamma}_{x_W} = \frac{1}{m}\sum_{j=1}^{m} (y_{W_{j'}} - \hat{m}_{x_W})^T (y_{W_{j'}} - \hat{m}_{x_W}) \quad (12)$$

Otherwise, fixing x_W let us omit it and let us put $\Gamma_{x_W} = \begin{bmatrix} \sigma^2 & A \\ A^T & B \end{bmatrix}$, where σ^2 is the variance of Y_i conditional on x_i. The distribution of Y_i conditionally on Y_{V_i} (recall that $Y_{W_i} = \begin{bmatrix} Y_i \\ Y_{V_i} \end{bmatrix}$) is then a Gaussian distribution of mean and variance

$$M_{x_s} = m_{x_s} + AB^{-1}A^T(y_{V_i} - m_{V_i}) \quad (13)$$

$$\Sigma_{x_s}^2 = \sigma^2 - AB^{-1}A^T \quad (14)$$

Comparing the estimates (11)-(14) with (5), we calculate m_i, a_i, and q_{ij} (with $i = x_s$). More precisely, m_{x_s} is given by (11), a_{x_s} is given by $a_{x_s} = \frac{1}{\Sigma_{x_s}^2}$ with $\Sigma_{x_s}^2$ given by (14). Further, putting $q_{x_s} = (q_{x_s x_{t_1}}, q_{x_s x_{t_2}}, q_{x_s x_{t_3}}, q_{x_s x_{t_4}})$ and comparing (13) to (5) gives :

$$q_{x_s} = -2a_{x_s} AB^{-1} \quad (15)$$

So, we have the estimates of m_i, a_i, and q_{ij} for each configuration x_W. So, when the configurations vary, say, from 1 to r, we obtain $m_i^1, \ldots, m_i^r$ possibly different estimates for m_i, and the same for a_i, and q_{ij}. Let us put d_1 the cardinal of configurations of type $1, \ldots, d_r$ the cardinal of configurations of type r (we have $d_1 + \ldots + d_r = n$). Then we take as final estimates the means of the estimates associated with particular configurations x_W :

$$\hat{m}_i = \frac{1}{n}\sum_{t=1}^{r} d_i m_i^t , \quad (16)$$

$$\hat{a}_i = \frac{1}{n}\sum_{t=1}^{r} d_i a_i^t , \quad (17)$$

$$\hat{q}_{ij} = \frac{1}{n}\sum_{t=1}^{r} d_i q_{ij}^t \quad (18)$$

Finally, the whole approach can be summarized in the following steps:

(a) Estimation of α_i using Derin and Elliott method;

(b) For each configuration x_W, estimation of means m_{x_W} and variance-covariance matrices Γ_{x_W} with (11) and (12);

(c) For each configuration x_W, computation of m_i, a_{x_s} and $q_{x_s} = (q_{x_s x_{t_1}}, q_{x_s x_{t_2}}, q_{x_s x_{t_3}}, q_{x_s x_{t_4}})$ with (13) and (15);

(d) Calculation of final q_{ij}, a_i and m_i with (16)-(18) applied to the estimates obtained from r configurations of x_W used.

Let us notice that it is not necessary to take all the configurations x_W into account. For example, one could decide to consider the only configurations with no more than two classes.

The estimator $\hat{\theta} = \hat{\theta}(z)$ above will be called "new method from complete data" (NMCD), and its use inside ICE will be called "new method from incomplete data" (NMID).

4. UNSUPERVISED IMAGE SEGMENTATION

We present below the results of two series of experiments. The first series concerns simulated PMF with the energy given by (3). Considering two classes, the parameter is: $\theta = (\alpha_1, \alpha_2, a_1, a_2, m_1, m_2, q_{11}, q_{22}, q_{12})$. The aim here is to make different comparisons in the unsupervised image segmentation context. The first point of interest is to compare the estimates obtained with the four methods SGCD, SGID, NMCD and NMID considered. The second one is to look how these different estimates act upon the unsupervised segmentation methods based on them.

We performed two series of experiments, each of which having contained numerous and various situations.

The first series concerns simulated PMF. One result is presented in Table 1 and the corresponding images are presented in Figure 2. Roughly speaking, we can put forth the following conclusions :

(i) the estimates SGCD and NMCD are of comparable efficiency. The latter is rather a good one, even in the case of very strong noise. In particular, there is little difference when segmenting the noisy image with real parameters or with the estimated ones;

(ii) the estimate NMID works better than SGID in every kind of situations. When the noise is not too strong, their efficiency can be comparable; but, when it is strong, NMID can be significantly better than SGID.

In the case of SG estimation, 15 iterations have been used from the complete data and 30 iterations from the incomplete data. In the proposed parameter estimation method, 30 iterations have been used in ICE.

The second series concerns two classes images corrupted with a correlated noise. Let us consider a *"ring"* image corrupted with a synthetic Gaussian correlated noise. So, we have a partial knowledge about the model, or still the data correspond partially to PMF above (we only know that the noise is correlated and Gaussian). Such models are interesting because studying them allows one to understand how the methods work when the used model goes away from theoretical one. In other words, such studies provide some knowledge about the robustness of the methods considered.

So, we consider the two-classes image *"Ring"*, which is noisy with a correlated noise. The observed Y_s is then obtained from an independent noise $W = (W_s)_{s \in S}$ by averaging on the four nearest neighbors :

$$Y_s = \frac{1}{5}(W_s + \sum_{t \in V_s} W_t) \quad (19)$$

So, first W_t (for each $t \in S$) is drawn with the Gaussian distribution $N(m_1, \sigma_1^2)$ if x_t is white, and with the Gaussian distribution $N(m_2, \sigma_2^2)$ if x_t is black, and then Y_s is calculated, for each $s \in S$, using (19). Of course, such a correlated Gaussian field is not a Markov one and thus the model considered is not of PMF kind , in which $p(y|x)$ is a Markov distribution, considered in the previous subsection.

As for the first series, we performed numerous experiments and some results, representatives of the different results obtained, are specified in Table 2. They correspond to $m_1 = 0.$, $m_2 = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1$, which is a very strongly noisy case. In the whole, analogous conclusions to those specified in the previous sub-section hold. In general, SGCD and NMCD are quite efficient, even in the very noisy cases. When considering the incomplete data situation, NMID works better than SGID. We present in Figure 3 the two images obtained by Bayesian MPM based on SGID and NMID estimates. We have chosen an extreme case; in many others the difference between the two methods is less striking.

	RV	NMCD	NMID	SGCD	SGID
α_1	2	2.11	1.11	2.36	1.75
α_2	2	2.02	1.03	2.35	1.75
a_1	1.	0.96	1.11	0.91	0.91
a_2	1.	0.98	1.13	0.86	0.83
m_1	0.	0.01	0.	0.01	0.2
m_2	1.	0.97	1.14	1.08	0.4
q_{11}	-0.4	-0.37	-0.41	-0.32	-0.41
q_{22}	-0.4	-0.38	-0.41	-0.32	-0.3
q_{12}	0.	0.	0.	-0.01	0.
ER	11.53%	11.65%	12.23%	12.40%	48.67%

Table 1. *Results corresponding to Figure 2. Real values of parameters (RV), estimates from complete data with NM (NMCD), estimates from observed data with NM and ICE.* ***.*** *(NMID), estimates from complete data with (SGCD), estimates from observed data (SGID), and error ration (ER) of MPM segmentation.*

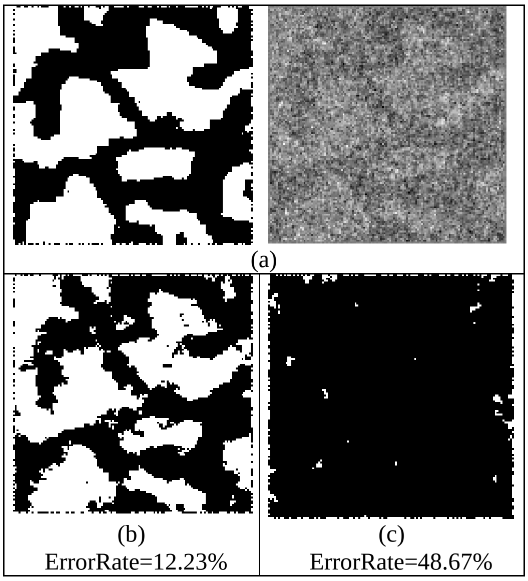

Figure 2. *(a) : simulated PMF. (b) and (c) : new method and SG based unsupervised segmentations (columns NMID and SGID in Table 1). Real parameters based Error Rate =11.53%*

	NM		SG	
	NMCD	NMID	SGCD	SGID
α_1	2.01	1.12	2.42	2.31
α_2	2.01	1.07	2.41	2.3
a_1	7.3	8.47	5.19	3.8
a_2	7.61	8.55	5.11	3.36
m_1	0.	0.	0.	0.
m_2	0.47	0.55	0.47	0.15
q_{11}	-0.5	-0.5	-0.32	0.
q_{22}	-0.49	-0.5	-0.17	-0.5
q_{12}	0.	0.	0.	0.
ER	7.15%	14.59%	7.72%	62.53%

Table 2. *Results corresponding to Image "ring". NMCD : estimated values using complete data, NMID : estimated values with NM and ICE using incomplete data, SGCD: estimated values with SG using complete data, SGID: estimated values with SG using incomplete. data. ER: error rate using MPM based on estimated parameters.*

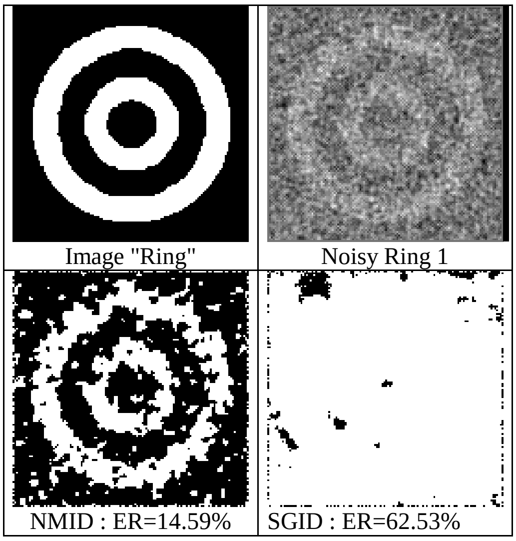

Figure 3. *Image "ring", its noisy version, and two Bayesian MPM segmentation results based on parameters estimation by NMID and SGID, respectively*

5. CONCLUSIONS

The Hidden Markov Fields model find numerous applications in various problems occurring in image processing. In such models, the hidden field X is a Markov one, and, when the noise in not too complex, its posterior distribution remains a Markov distribution. The latter property is vital, because it allows one to simulate X and thus apply different Bayesian processing methods. However, when willing model texture by a Markov noise, such a noise is too complex and thus the posterior distribution is no longer a Markov one. To remedy this, in Pairwise Markov Fields (PMF) one directly assumes that the couple (hidden field, observed field) is a Markov field. Although X is possibly no longer a Markov field in such models, they still allow one to recover the hidden process from the observed one.

In this paper we tackled the problem of parameter estimation. We proposed two original methods based on Stochastic Gradient (SG) and Iterative Conditional Estimation (ICE), the latter being associated with an original generalization of Derin's *et al.* method. The general conclusion is that the ICE based method is faster and more efficient than the SG based one. In particular, the ICE based estimation is close to the estimation obtained from complete data, which attests that the use of ICE is interesting in the context considered.

As perspective, we may view different use of the new parameter estimation method in more complex TMF, and their applications in real image unsupervised segmentation.

6. 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] J.-M. Boucher and P. Lena, Unsupervised Bayesian classification, application to the forest of Paimpont (Brittany), Photo Interprétation, Vol. 32, No. 1994/4, 1995/1-2, 1995, pp. 79-81.

[3] J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, IEEE Trans. on Signal Processing, Vol. 45, No. 10, pp. 2613-2615, 1997.

[4] H. Derin and H. Elliot, Modeling and segmentation of noisy and textured images using Gibbs random fields, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 9, pp. 39-55, 1987.

[5] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 6, pp. 721-741, 1984.

[6] 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.

[7] Z. Kato, J. Zeroubia, and M. Berthod, Unsupervised parallel image classification using Markovian models, Pattern Recognition, Vol. 32, pp. 591-604, 1999.

[8] 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.

[9] G. J. McLachlan and T. Krishnan, EM Algorithm and Extensions, Wiley, Series in Probability and Statistics, 1997.

[10] D. Melas and S. P. Wilson, Double Markov Random Fields and Bayesian Image Segmentation, IEEE Trans. on SP, Vol. 50, No. 2, pp. 357-365, 2002.

[11] M. Mignotte, C. Collet, P. Pérez, and P. Bouthémy, Sonar image segmentation using an unsupervised hierarchical MRF model, IEEE Trans. on Image Processing, Vol. 9, No. 7, pp. 1216-1231, 2000.

[12] W. Pieczynski and A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, Machine Graphics and Vision, Vol. 9, No. 3, pp. 705-718, 2000.

[13] 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.

[14] W. Pieczynski, D. Benboudjema, and P. Lanchantin, Statistical image segmentation using Triplet Markov fields, SPIE's International Symposium on Remote Sensing, September 22-27, Crete, Greece, 2002.

[15] S. Reed, Y. Petillot, and J. Bell, An automatic approach to the detection and extraction of mine features in sidescan sonar, IEEE Journal of Oceanic Engineering, Vol. 28, No. 1, pp. 90-105, 2003.

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