



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Computer Vision and Image Understanding 99 (2005) 476–498

Computer Vision
and Image
Understanding

www.elsevier.com/locate/cviu

Unsupervised image segmentation using triplet Markov fields

Dalila Benboudjema, Wojciech Pieczynski *

GET/INT, Département CITI, 9, rue Charles Fourier, 91000 Evry, France

Received 14 April 2004; accepted 22 April 2005

Available online 13 June 2005

Abstract

Hidden Markov fields (HMF) models are widely applied to various problems arising in image processing. In these models, the hidden process of interest X is a Markov field and must be estimated from its observable noisy version Y . The success of HMF is mainly due to the fact that the conditional probability distribution of the hidden process with respect to the observed one remains Markovian, which facilitates different processing strategies such as Bayesian restoration. HMF have been recently generalized to “pairwise” Markov fields (PMF), which offer similar processing advantages and superior modeling capabilities. In PMF one directly assumes the Markovianity of the pair (X, Y) . Afterwards, “triplet” Markov fields (TMF), in which the distribution of the pair (X, Y) is the marginal distribution of a Markov field (X, U, Y) , where U is an auxiliary process, have been proposed and still allow restoration processing. The aim of this paper is to propose a new parameter estimation method adapted to TMF, and to study the corresponding unsupervised image segmentation methods. The latter are validated via experiments and real image processing.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Hidden Markov fields; Pairwise Markov fields; Triplet Markov fields; Bayesian classification; Mixture estimation; Iterative conditional estimation; Stochastic gradient; Unsupervised image segmentation

* Corresponding author. Fax: +33 1 60 76 44 33.

E-mail address: Wojciech.Pieczynski@int-evry.fr (W. Pieczynski).

1. Introduction

Hidden Markov fields (HMF) are widely used in solving 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 $Y = y$. This wide use is due to the fact that standard Bayesian restoration methods can be used in spite of the large size of S : see [3,12,19] for seminal papers and [14,33], among others, for general books. The qualifier “hidden Markov” means that the hidden process X has a Markov law. When the distributions $p(y|x)$ of Y conditional on $X = x$ are simple enough, the pair (X, Y) then retains the Markovian structure, and likewise for the distribution $p(x|y)$ of X conditional on $Y = y$. The Markovianity of $p(x|y)$ is crucial because it 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 in standard HMF to ensure the Markovianity of $p(x|y)$ can pose problems; in particular, such situations occur in textured images segmentation [21]. To remedy this, the use of pairwise Markov fields (PMF), in which one directly assumes the Markovianity of (X, Y) , has been discussed in [26]. Both $p(y|x)$ and $p(x|y)$ are then Markovian, the former ensuring possibilities of modeling textures without approximations, and the latter allowing Bayesian processing, similar to those provided by HMF. PMF have then been generalized to “triplet” Markov fields (TMF), in which the distribution of the pair $Z = (X, Y)$ is the marginal distribution of a Markov field $T = (X, U, Y)$, where $U = (U_s)_{s \in S}$ is an auxiliary random field [27]. Once the space \mathcal{A} of possible values of each U_s is simple enough, TMF still allow one to estimate the unobservable $X = x$ from the observed $Y = y$. Given that in TMF $T = (X, U, Y)$ the distribution of $Z = (X, Y)$ is its marginal distribution, the Markovianity of T does not necessarily imply the Markovianity of Z ; and thus a TMF model is not necessarily a PMF one. Therefore, TMF are more general than PMF and thus are likely to be able to model more complex situations. Conversely, a PMF model can be seen as a particular TMF model in which $X = U$.

There are some studies concerning triplet Markov chains [18,28], where general ideas somewhat similar to those discussed in the present paper, have been investigated. However, as Markov fields based processing is quite different from the Markov chains based one, we will concentrate here on Markov fields with no further reference to Markov chains.

This paper is mainly devoted to study situations in which using TMF can improve the image segmentation results obtained with PMF and HMF. Its principal purpose is to propose an efficient TMF parameter estimation method and show its suitability in unsupervised image segmentation. The new method, briefly introduced in [1], is based on the general Iterative Conditional Estimation (ICE) principle [25]. This principle, whose relationship to the well-known “Expectation-Maximization” (EM) principle is described in [8], already has given satisfactory results in different classical HMF-based processing [4,16,22,23,30]. The new method we propose is applied in different situations and shows, via experiments, its superiority over an adaptation to TMF of the “Stochastic Gradient,” which works well in the context of HMF [35]. It is also shown that the corresponding unsupervised Bayesian MPM image seg-

mentation method can improve the efficiency of the classical HMF and PMF-based analogous methods.

The organization of the paper is as follows. Some basic principles of Bayesian segmentation, as well as classical HMF, PMF, and TMF, are cited in the next section, along with some original models and comments. Section 3 is devoted to the original parameter estimation methods we propose, and some unsupervised image segmentation results are presented in Section 4. In particular, we deal with real image segmentation and show the interest of the new unsupervised methods based on TMF with respect to classical ones based on HMF and PMF. Section 5 contains conclusions and perspectives.

2. Bayesian segmentation with triplet Markov fields

2.1. Bayesian segmentation

Let $X = (X_s)_{s \in S}$ and $Y = (Y_s)_{s \in S}$ be two random fields as above, each X_s taking its values in $\Omega = \{\omega_1, \dots, \omega_k\}$ and each Y_s taking its values in R , and let us consider the problem of estimating the unknown $X = x$ from the observation $Y = y$. In this paper, $Y = y$ is the observed image and $X = x$ is the class image. Therefore, for each $s \in S$ the problem is to propose an “estimated” $\hat{x}_s \in \Omega$, that would be “optimal” in some sense. In other words, each pixel $s \in S$ is associated with a class, which can be, for example, “water,” “forest,” “urban area,” ... and so on. To be more precise, let us consider the Maximum Posterior Mode (MPM) segmentation method classically used in the HMF context. This produces an estimated $\hat{X} = \hat{x} = (\hat{x}_s)_{s \in S}$, with $\hat{x}_s = \arg \max_{\omega \in \Omega} p(x_s = \omega | y)$, which means that the class associated with each pixel $s \in S$ is that which maximizes its probability conditionally on the observation of the whole image $y = (y_t)_{t \in S}$. Thus, MPM is workable once the marginal distributions can be obtained in some way. The main property allowing one to obtain marginal distributions $p(x_s | y)$, is the fact that in HMF the distribution of X conditionally on $Y = y$ is a Markov distribution. In fact, the latter allows one to sample X according to $p(x | y)$ and estimate $p(x_s | y)$ from the obtained samples. Let us also recall that the Bayesian MPM method is optimal in the following sense. If $X = x$ is the true class image and x^* is the “estimated” one, possibly different from x , let us assume that the gravity of the error is measured by the number of pixels on which the classes are different. This number is thus given by the “loss function” $L(x^*, x) = \sum_{s \in S} \chi(x_s^*, x_s)$, with $\chi(x_s^*, x_s) = 0$ for $x_s^* = x_s$ and $\chi(x_s^*, x_s) = 1$ for $x_s^* \neq x_s$. The MPM method \hat{s}_{MPM} is then the best one, with respect to all possible methods (statistical or other), with respect to the mean error: $E[L(\hat{s}_{\text{MPM}}(Y), X)] = \min_{\hat{s}} E[L(\hat{s}(Y), X)]$. According to the large number law, the latter has quite an intuitive interpretation: using MPM ensures minimizing the loss in the long run. In fact, if \hat{s}_{MPM} is applied to n digital images y^1, \dots, y^n and if x^1, \dots, x^n are the true unknown corresponding class images, we have $L(\hat{s}_{\text{MPM}}(y^1), x^1) + \dots + L(\hat{s}_{\text{MPM}}(y^n), x^n) \approx nE[L(\hat{s}_{\text{MPM}}(Y), X)]$ and so, for m large enough, the loss is smaller than the loss obtained by any other method $y \rightarrow \hat{x} = \hat{s}(y)$.

To summarize, the important point is the possibility of sampling X according to $p(x|y)$, which exists in models below.

2.2. Hidden Markov fields

The family of triplet Markov fields offers numerous and rich possibilities for particular models. However, to better emphasize the interest of TMF with respect to classical HMF, let us first specify some general properties of the latter. In fact, although the simplest models of the HMF family are almost exclusively used in practice, this family is itself quite rich. Let us specify three hidden Markov fields (HMF) of growing generality, the third being an original model. In the following, we will say that $Z = (X, Y)$ is an HMF if both X and $Z = (X, Y)$ are Markov fields.

(i) HMF with Independent Noise (HMF-IN)

The simplest HMF is the widely used classical HMF with Independent Noise (HMF-IN), in which the conditional distribution $p(y|x)$ verifies the two following hypotheses:

$$(H1) \quad p(y|x) = \prod_{s \in S} p(y_s|x_s);$$

$$(H2) \quad p(y_s|x) = \bar{p}(y_s|x_s) \text{ for each } s \in S.$$

We say that the noise is “independent” because (H1) means that the random variables $(Y_s)_{s \in S}$ are independent conditionally on X . The distribution of the Markov field X is a Gibbs distribution with respect to a neighboring system $(A_s)_{s \in S}$; for example, A_s can be the set of four nearest neighbors of $s \in S$. Denoting by C the set of cliques (a clique being either a singleton or a set of mutually neighboring pixels), the distribution of X is classically written $p(x) = \gamma \exp[-\sum_{c \in C} \varphi_c(x_c)]$. Thus, the distribution of (X, Y) is:

$$p(x, y) = \gamma \exp \left[- \sum_{c \in C} \varphi_c(x_c) + \sum_{s \in S} \log [p(y_s|x_s)] \right]. \quad (2.1)$$

Although the hypotheses (H1) and (H2) are rarely satisfied in practical situations, the HMF-IN model is widely used and gives satisfactory results in numerous contexts. However, in some others, as in the presence of textured classes, these hypotheses can turn out to be too strong [21].

(ii) HMF with Correlated Noise (HMF-CN)

In the HMF-CN hypothesis (H1) is relaxed and (H2) is maintained. A simple way of constructing such models follows [14]. Let $X = (X_s)_{s \in S}$ be a Markov field on $\Omega^{\text{Card}(S)}$ as above, and $W = (W_s)_{s \in S}$ be a Markov field on $R^{\text{Card}(S)}$, both being Markovian with respect to the same neighborhood, which produces the set of cliques C . For each $s \in S$ and $x_s \in \Omega$, let f_{x_s} be a bijective and derivable function from R to R , and let us assume that for each fixed x , $Y = f(x, W)$, with $f(x, w) = (f_{x_s}(w_s))_{s \in S}$. Denoting by

$\sum_{c \in C} \varphi_c(x_c)$ the energy of the Markov field X and $\sum_{c \in C} \phi_c(w_c)$ the energy of the Markov field W , the joint distribution $p(x, y)$ is then given by $p(x, y) = p(x)p(y|x)$. Thus, we have

$$p(x, y) = \gamma \exp \left[- \left(\sum_{c \in C} \varphi_c(x_c) + \sum_{c \in C} \phi_c(f_{x_c}^{-1}(y_c)) - \sum_{s \in S} \log \left| \frac{\partial(f_{x_s}^{-1}(y_s))}{\partial y_s} \right| \right) \right]. \quad (2.2)$$

We see that since $Z = (X, Y)$ and X are Markov fields, Z is a HMF. Moreover, we notice that (H2) is verified while (H1) is not, consequently Z is not a HMF-IN.

(iii) HMF, which are neither HMF-IN nor HMF-CN

Let us consider the following HMF, which is strictly more general than HMF-IN and HMF-CN. In other words, both (H1) and (H2) are relaxed but X and (X, Y) remain Markov fields. Let X be a Markov field on $\Omega^{\text{Card}(S)}$ and $W = (W_s)_{s \in S}$ be a Markov field on $R^{\text{Card}(S)}$, as in (ii) above. Let C_1, \dots, C_q be a set of cliques of cardinal $a \geq 2$ forming a partition of S . So, $S = C_1 \cup \dots \cup C_q$ and $C_i \cap C_j = \emptyset$ for $i \neq j$. Of course, there are many other cliques which are not in $\{C_1, \dots, C_q\}$. For each C_i and $x_{C_i} \in \Omega^a$, let $f_{x_{C_i}}$ be a bijective and derivable function from R^a to R^a , and let us assume that for each fixed x , $Y = f(x, W)$, with $f(x, w) = (f_{x_{C_i}}(w_{C_i}))_{1 \leq i \leq q}$. Denoting by $[\partial(f_{x_{C_i}}^{-1}(y_{C_i}))/\partial y_{C_i}]$ the Jacobian of $f_{x_{C_i}}^{-1}$ (which is a matrix of size $a \times a$), we have

$$p(x, y) = \gamma \exp \left[- \left(\sum_{c \in C} \varphi_c(x_c) + \sum_{c \in C} \phi_c(f_{x_c}^{-1}(y_c)) - \sum_{1 \leq i \leq q} \log \left| \det \left[\frac{\partial(f_{x_{C_i}}^{-1}(y_{C_i}))}{\partial y_{C_i}} \right] \right| \right) \right]. \quad (2.3)$$

We see how (2.3) generalizes (2.2), the latter being obtained for particular $f_{x_{C_i}}$ verifying $f_{x_{C_i}}(y_{C_i}) = f_{(x_1, \dots, x_a)}(y_1, \dots, y_a) = (f_{x_1}(y_1), \dots, f_{x_a}(y_a))$. Further, X is a Markov field by the very construction of $p(x, y)$ (which is defined by a Markov $p(x)$ and some $p(y|x)$), and $p(x, y)$ is a Markov field because the last sum in (2.3) is taken over cliques.

Remark 2.1. When sampling X according to $p(x|y)$, one uses the conditional local marginal distributions $p(x_s|x_{V_s}, y)$, where V_s is the set of neighbors of s . Furthermore, completing the global distributions (2.1)–(2.3), these are useful to figure out the degree of generality of the corresponding model. So, let us see how these distributions vary in the three models above. In HMF-IN, we have according to (2.1): $p(x_s|x_{V_s}, y) \propto \exp[-\sum_{s \in c} \varphi_c(x_c) + \log[p(y_s|x_s)]]$ (s is fixed and the sum is taken over the cliques c such that $s \in c$). In an analogous way, these distributions are proportional to $\exp \left[- \left(\sum_{s \in c} \varphi_c(x_c) + \sum_{s \in c} \phi_c(f_{x_c}^{-1}(y_c)) - \log \left| \frac{\partial(f_{x_s}^{-1}(y_s))}{\partial y_s} \right| \right) \right]$ in HMF-CN, and in HMF given with (2.3) they are proportional to $\exp \left[- \left(\sum_{s \in c} \varphi_c(x_c) + \sum_{s \in c} \phi_c(f_{x_c}^{-1}(y_c)) - \log \left| \det \left[\frac{\partial(f_{x_{C_i}}^{-1}(y_{C_i}))}{\partial y_{C_i}} \right] \right| \right) \right]$, where C_i is such

that $s \in C_i$. We see that these conditional distributions depend on y_s in the case of HMF-IN, and depend on y_{V_s} , with increasing complexity, in the HMF-CN and HMF cases.

2.3. Pairwise Markov fields

In PMF one assumes the Markovianity of $Z = (X, Y)$. As marginals do not necessarily retain Markovianity, neither X nor Y are Markov fields, and thus PMF are more general than HMF in (i)–(iii) above. Therefore, the distribution of $Z = (X, Y)$ can be written as $p(x, y) = \gamma \exp[-\sum_{c \in C} \psi_c(x_c, y_c)]$, and the local conditional distributions mentioned in Remark 2.1 are $p(x_s | x_{V_s}, y) = \exp[-\sum_{s \in C} \psi_c(x_c, y_c)] / \sum_{\omega \in \Omega} \exp[-\sum_{s \in C} \psi_c(x_{c-\{s\}}, x_s = \omega, y_c)]$, where the functions ψ_c are of arbitrary form. We can see how these distributions generalize those of the previous subsection.

Let us consider simple ‘‘Gaussian’’ PMF, meaning that $p(y|x)$ is Gaussian, which will be used in experiments below. Therefore, let (X, Y) be a random field with

$$\begin{aligned}
 p(x, y) &= \gamma \exp[-G(x, y)] \\
 &= \gamma \exp \left[-\frac{1}{2} \left[\sum_{s \in S} a_{x_s} (y_{x_s} - m_{x_s})^2 + \sum_{(s,t) \in C} [\alpha(1 - 2\delta(x_s, x_t)) \right. \right. \\
 &\quad \left. \left. + q_{x_s x_t} (y_{x_s} - m_{x_s})(y_{x_t} - m_{x_t})] \right] \right], \tag{2.4}
 \end{aligned}$$

(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 Markovian, and that the latter is also Gaussian.

Let us see how to calculate $p(z_s | z_{V_s})$, which is used to sample the PMF (X, Y) . We have $p(x_s, y_s | x_{V_s}, y_{V_s}) = p(x_s | x_{V_s}, y_{V_s}) p(y_s | x_s, x_{V_s}, y_{V_s})$, where:

$$\begin{aligned}
 p(x_s | x_{V_s}, y_{V_s}) &\propto \frac{1}{\sqrt{a_{x_s}}} \\
 &\times \exp \left[-\frac{1}{2} \left(\sum_{t \in V_s} \alpha(1 - 2\delta(x_s, x_t)) - \frac{1}{4a_{x_s}} \left(\sum_{t \in V_s} q_{x_s x_t} (y_t - m_{x_t}) \right)^2 \right) \right] \tag{2.5}
 \end{aligned}$$

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

$$M_{x_s} = m_{x_s} - \frac{\sum_{t \in V_s} q_{x_s x_t} (y_t - m_{x_t})}{2a_{x_s}}, \tag{2.6}$$

$$\Sigma_{x_s}^2 = \frac{1}{a_{x_s}}. \tag{2.7}$$

An important feature of such Gaussian PMF models is that they take into account noise correlation, which potentially contains some useful information. For example, let us consider the case of two classes, where the two Gaussian distributions of Y_s

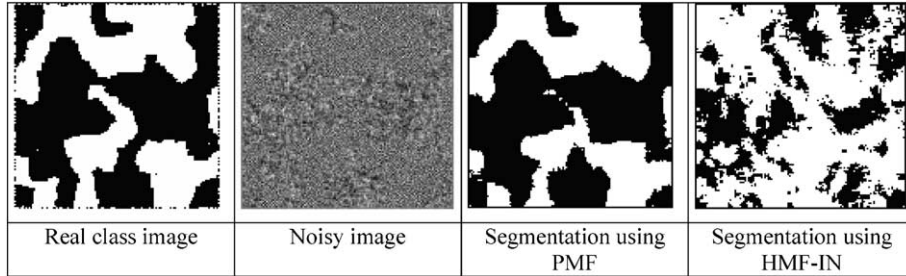


Fig. 1. Simulated PMF, with both distributions of Y_s conditionally on $X_s = \omega_1, \omega_2$ being the same ($p(y_s|x_s = \omega_1) = p(y_s|x_s = \omega_2)$) and comparison between PMF and HMF-IN unsupervised segmentation.

conditionally on $X_s = \omega_1$ and $X_s = \omega_2$ are the same. So, the only possible information is contained in the spatial correlations of the noise. Can this be enough to perform a suitable segmentation? We give an example in Fig. 1, where we see that PMF give an acceptable result while HMF-IN lead to a quite poor one (in both cases the parameters have been estimated by a new method specified in the next section).

2.4. Triplet Markov fields

Let $X = (X_s)_{s \in S}$ and $Y = (Y_s)_{s \in S}$ be two random fields as above. A “triplet” Markov field is defined as $T = (X, U, Y)$ by introducing a third random field $U = (U_s)_{s \in S}$, with each U_s taking its values in a finite set of values, and by assuming that T is a Markov field. So, X and Y have a real interpretation but, which is the main idea behind “triplet” fields, the random field U can be latent, possibly without any physical interpretation. As T is Markovian, its distribution is a Gibbs distribution with respect to a neighboring system $(A_s)_{s \in S}$. As for X in Section 2.2, the distribution of T is then written:

$$p(t) = \gamma \exp \left[- \sum_{c \in C} \varphi_c(t_c) \right], \tag{2.8}$$

where C is the set of cliques associated with $(A_s)_{s \in S}$. Classically, the distribution (2.8) of $T = (X, U, Y)$ results in a distribution of $V = (X, U)$ conditionally on $Y = y$, which is also a Markov distribution. In fact, the distribution $p(x, u|y)$ is classically obtained from $p(x, u, y)$ by considering y as a constant in (2.8). We obtain

$$p(x, u|y) = \gamma^*(y) \exp \left[- \sum_{c \in C} \varphi_c(x_c, u_c, y_c) \right]. \tag{2.9}$$

As mentioned at the end of Section 2.1, it is important to be able to simulate X from the distribution $p(x|y)$. This is possible because realizations of $V = (X, U)$ according to $p(x, u|y)$ can be simulated, and the latter realizations give realizations of X . However, let us insist on the fact that TMF are more general than PMF in that $p(x|y)$ is the marginal distribution of the Markov distribution $p(x, u|y)$, and thus is not necessarily a Markov distribution itself. This greater generality permits better modelling

of more complex situations; however, its practical interest in an unsupervised context is subject to the reliability of the parameter estimation methods used. As mentioned above, proposing such a reliable method is the main aim of the paper.

2.5. Three interpretations of TMF

As specified above, PMF can be seen as a particular case of TMF (where $X = U$), and thus the latter should be able, in theory, to better model more complex situations. This is true for general latent processes U , without any physical interpretation. However, U can also admit an interpretation in some situations, and thus some TMF can also be justified by a physical reality. So, let us specify three possible interpretations of the latent process U .

1. Let $Z = (X, Y)$ be a PMF model with two classes such that the distributions of Y_s conditionally on $X_s = x_s$ have complex densities f_1, f_2 impossible to express in closed form. One can then aim to approximate f_1, f_2 by mixtures: $f_1 \approx \sum_{i=1}^m \alpha_1^i f_1^i$, $f_2 \approx \sum_{i=1}^m \alpha_2^i f_2^i$, where $f_1^1, \dots, f_1^m, f_2^1, \dots, f_2^m$ are given densities. Such a situation can be seen as a particular TMF; in particular, if $Z = (X, Y)$ is a HMF-IN where X is a Markov field and $p(y|x) = \prod_{s \in S} p(y_s|x_s)$, we can introduce $U = (U_s)_{s \in S}$, each U_s taking its values in a set $A = \{\lambda_1, \dots, \lambda_m\}$, such that X has the same structure and $p(u, y|x) = \prod_{s \in S} p(u_s, y_s|x_s) = \prod_{s \in S} p(u_s|x_s) p(y_s|x_s, u_s)$, where $p(u_s = \lambda_i|x_s = \omega_j) = \alpha_j^i$ and $p(y_s|x_s = \omega_j, u_s = \lambda_i) = f_j^i(y_s)$.

2. The second interpretation is like the first one, except for the fact that we know $f_1 = \sum_{i=1}^{m_1} \alpha_1^i f_1^i$ and $f_2 = \sum_{i=1}^{m_2} \alpha_2^i f_2^i$. Such a situation occurs when each of two classes is an union of finite subclasses, the latter admitting known densities. An example of real images verifying such a situation is treated in the next section.

3. The last interpretation is more complex and calls on the “theory of evidence” [32]. Considering a HMF-IN $Z = (X, Y)$, let us assume that the distribution of X is not well known and cannot be estimated because, for example, of a strong non-stationarity. In such cases, the distribution of X can be replaced by a “Markovian mass function” and the posterior distribution of X is obtained by the so-called “Dempster–Shafer combination rule” [32] which fuses this mass function with a probability obtained from the observation $Y = y$. Such a fusion destroys Markovianity; however, its result remains workable because it is formally a TMF model. These cases are only mentioned here and will be not treated further in this paper (see [29] for more details).

Remark 2.2. So TMF generalize PMF, which generalize HMF, the latter generalizing HMF-IN. As HMF-IN are often complicated, one could imagine that TMF would be very complicated and thus their practical interest would be quite limited. This is not necessarily true because there are very simple TMF which are neither HMF, nor even PMF. In fact, let U be a Markov field, and let us assume random variables (Y_s) and random variables (X_s) independent conditionally on U , with $p(y_s|u) = p(y_s|u_s)$ and $p(x_s|u) = p(x_s|u_s)$. The distribution of T is written $p(t) = \gamma \exp[-W(t)]$, with $W(t) = \sum_{c \in C} \varphi(u_c) - \sum_{s \in S} [\log(p(x_s|u_s)) + \log(p(y_s|u_s))]$. For-

mally, T is then a classical HMF-IN, where the hidden Markov field is $V = (X, U)$. So, the complexity of such a TMF model is equivalent to the complexity of a HMF-IN model $T = (V, Y)$, which means that the computational complexity is comparable in both models. For example, if $\Omega = \{\omega_1, \omega_2\}$ and $\Lambda = \{\lambda_1, \lambda_2\}$, the complexity of the TMF model T is equivalent to the complexity of a HMF-IN model with four classes (each component V_s of the hidden field V takes its values in $\Omega \times \Lambda$). Otherwise, the distribution of X is the marginal distribution of the distribution of (X, U) given by $p(x, u) = \gamma \exp[-\sum_{c \in C} \varphi(u_c) + \sum_{s \in S} \log(p(x_s | u_s))]$, and thus is not a Markov distribution. The latter means that (X, Y) is not a HMF.

Remark 2.3. As we place ourselves in the context of Bayesian segmentation, the distribution of the hidden field X can be interpreted as modelling our prior knowledge about possible forms of class images. As this distribution is not known in a closed form in TMF, modelling prior knowledge does not appear in a straightforward way. However, we have seen that, on the one hand $p(x, u | y)$ was Markovian and, on the other hand, it generalized the classical Markovian $p(x | y)$ corresponding to HMF-IN. Furthermore, both $p(x_s, u_s | x_{V_s}, u_{V_s}, y)$ in TMF, and $p(x_s | x_{V_s}, y)$ HMF-IN are calculable. Comparison of these conditional distributions, which are easier to interpret than the whole distribution of X , could then be used to elicit the form of energy in TMF, which would generalize the simpler energy of HMF-IN. In other words, we can start from a classical HMF-IN with a given energy, and then search for a TMF energy such that the corresponding $p(x_s, u_s | x_{V_s}, u_{V_s}, y)$ would generalize $p(x_s | x_{V_s}, y)$ obtained from HMF-IN in a useful and intuitively satisfying way. To summarize, we have two random fields (X, Y) and the problem is to recover X from Y . Classical HMF-IN provide an interesting and widely used model. Recent TMF provide a more general model, which can be used in more complex situations. Furthermore, TMF include models which are not HMF and yet present a comparable complexity.

3. Parameter estimation

The problem of parameter estimation from the observed data $Y = y$ is a “mixture estimation” problem, which comes from the fact that the distribution of Y is a “mixture” of distributions $p(y) = \sum_x p(x, y)$. The most widely used general method is the so-called “Expectation-Maximization” (EM) method [20]; however, its implementation in the hidden Markov fields context is difficult [6,20,24] and some alternative methods have been proposed [3,9,15,17,24,34,35]. In particular, one may consider Stochastic Gradient (SG [35]), whose aim is to approach the maximum of the likelihood $p(y)$ in a stochastic manner to remedy the difficulties encountered by EM. Another possibility is to use the general “Iterative Conditional Estimation” (ICE [25]) method, which has given good results in different classical HMF situations [4,7,10,16,22,23,30], and in more complex Markov models with a Dempster–Shafer fusion [2] or fuzzy hidden fields [31]. Moreover, first applications of ICE in a simple TMF context also gave promising results [27]. ICE resembles EM, and some

relationships are specified in [8]; however, it is more flexible and seems better suited to the Markov field context. In fact, ICE does not necessarily use the likelihood, and the latter is difficult to handle in different Markov field models.

Therefore, we choose SG and ICE for general principles and propose two original parameter estimation methods based on them and valid in the general TMF context. The first one is an adaptation to TMF, in a rather straightforward manner, of the SG proposed in [35]. The second method, which is the main novelty of this paper, is new.

We notice that the parameter estimation problem in TMF is exactly the same as that in PMF. In fact, TMF (X, U, Y) can be seen as PMF (V, Y) , with $V = (X, U)$. So, we will deal in this section with the parameter estimation problem of PMF (X, Y) and, when dealing with TMF, X will simply be replaced with $V = (X, U)$.

To simplify things, we will specify the different methods in a simple particular Gaussian PMF defined by (2.4). On the one hand, its generalization does not pose problems and, on the other, such models, on which the experiments presented in this paper are based, are generally sufficient for practical applications.

Hence, for k classes $\Omega = \{\omega_1, \dots, \omega_k\}$, the vector of parameters θ to be estimated contains the following real numbers: $\alpha, a_1, \dots, a_k, m_1, \dots, m_k$, and k^2 components q_{ij} , $1 \leq i, j \leq k$.

The SG-based method of estimating θ from $Y = y$ is described in Section 3.1, and that based on ICE is described in Sections 3.2 and 3.3.

3.1. Stochastic gradient algorithm

The aim of SG is to search for the maximum of the likelihood: for observed $Y = y$, the problem is to find θ maximizing $p_\theta(y)$. We do not propose here any theoretical properties, whose study is rather tedious, but the extension of results presented in [35] in the case of HMF-IN could be considered. In fact, the problem is difficult because $p_\theta(x, y) = \gamma(\theta) \exp[-G_\theta(x, y)]$ and thus θ appears simultaneously in the energy G and in the constant γ . The proposed method runs as follows:

- Initialize parameter vector θ_0 .
- Calculate θ_{n+1} from θ_n and $Y = y$ by:

$$\theta_{n+1} = \theta_n + \frac{K}{n+1} [\nabla_{\theta_n} G_\theta(x_{n+1}, y_{n+1}) - \nabla_{\theta_n} G_\theta(x_{n+1}^*, y)], \quad (3.1)$$

where $\nabla_{\theta_n} G_\theta(x_{n+1}, y_{n+1})$ is the gradient of $G_\theta(x_{n+1}, y_{n+1})$ with respect to θ taken at θ_n , (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 sampler according to its posterior distribution $p(x|y)$ based on θ_n , and K is a constant.

SG can be initialized by different classical methods like k -means, or, in the case of image segmentation, by some histogram based method. To stop SG, different criteria can be used subject to each particular case considered.

Such a method will be called Stochastic Gradient from Incomplete Data (SGID, see also Remark 3.1).

Applying the general equation (3.1) to the Gaussian PMF defined by (2.4), we obtain:

$$\begin{aligned}
 \frac{\partial G}{\partial a_i} &= 0.5 \sum_{s \in \mathcal{S}} (y_s - m_i)^2 \mathbf{1}_{[x_s=i]}, \\
 \frac{\partial G}{\partial \alpha} &= 0.5 \sum_{(s,t) \in \mathcal{C}} [1 - 2\delta(x_s, x_t)], \\
 \frac{\partial G}{\partial q_{ij}} &= \sum_{(s,t) \in \mathcal{C}} (y_s - m_i)(y_t - m_j) \mathbf{1}_{[x_s=i]} \mathbf{1}_{[x_t=j]}, \\
 \frac{\partial G}{\partial m_i} &= \sum_{s \in \mathcal{S}} a_i (m_i - y_i) \mathbf{1}_{[x_s=i]} + \sum_{(s,t) \in \mathcal{C}} q_{ij} (m_j - y_t) \mathbf{1}_{[x_s=i]} \mathbf{1}_{[x_t=j \neq i]} \\
 &\quad + \sum_{(s,t) \in \mathcal{C}} q_{ii} (2m_i - y_s - y_t) \mathbf{1}_{[x_s=i]} \mathbf{1}_{[x_t=i]}.
 \end{aligned} \tag{3.2}$$

Remark 3.1. When a learning sample is available, that is to say when $X = x$ is observed, SG can still be used replacing in (3.1) the simulated x_{n+1}^* (which varies with iterations) by the fixed observed $X = x$. So the method, which will be called Stochastic Gradient from Complete Data (SGCD), runs as follows:

- Initialize parameter vector θ_0 ;
- Calculate θ_{n+1} from θ_n and Y by:

$$\theta_{n+1} = \theta_n + \frac{K}{n+1} [\nabla_{\theta_n} G(x_{n+1}, y_{n+1}) - \nabla_{\theta_n} G(x, y)]. \tag{3.3}$$

3.2. Iterative conditional estimation

ICE is based on the following assumptions:

- (i) there exists an estimator $\hat{\theta} = \hat{\theta}(X, Y)$ of θ from the complete data (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 on $Y = y$ are feasible.

ICE is an iterative method which runs as follows:

1. Initialize $\theta = \theta_0$;
2. for $n \in N$,
 - (a) put $\theta_{n+1} = E_{\theta_n}[\hat{\theta}(X, Y)|Y = y]$ if the conditional expectation is computable;

- (b) if not, simulate l realizations x^1, \dots, x^l of X (each x^i is a class image) according to its distribution conditional on $Y = y$ and based on θ_n and put
$$\theta_{n+1} = \frac{\hat{\theta}(x^1, y) + \dots + \hat{\theta}(x^l, y)}{l}.$$

The problems of initialization and stopping ICE are analogous to those of initialization and stopping SG, and so the same methods will be used for both below.

As in PMF (X, Y) the distribution of X conditional on $Y = y$ is a Markov field distribution, its simulations are feasible using the Gibbs sampler or Metropolis algorithm, and thus the condition (ii) always holds. Therefore, all we have to do in PMF is to search for an estimator from complete data $\hat{\theta} = \hat{\theta}(x, y)$.

3.3. Estimation from complete data (X, Y)

In this subsection, we present an original approach for estimating PMF parameters from complete data. The proposed procedure, which will be called “New Method from Complete Data” (NMCD) in the following, is mainly based on the least squares method proposed by Derin and Elliott [9] to estimate the parameters α , and on the use of conditional distributions to estimate the remaining model parameters, relative to the Gaussian distribution of Y conditional on X .

Of course, the method below can be either directly used when a learning sample of (X, Y) is available, or inside ICE, when not. When it is used inside ICE, that is to say when only the incomplete data Y are available, it will be called “New Method from Incomplete Data” (NMID).

Although the proposed method can be extended to higher order neighborhood systems, we present it in the case of simple neighborhood containing the four nearest neighbors, as shown in Fig. 2.

Let us consider a sequence of sets of pixels W_1, \dots, W_n , where each W_i is of the form presented in Fig. 2. Next, we choose a slightly more general energy, replacing in (2.4) $\sum_{(s,t) \in C} \alpha(1 - 2\delta(x_s, x_t))$ by $\sum_{(s,t) \in C^H} \alpha_1(1 - 2\delta(x_s, x_t)) + \sum_{(s,t) \in C^V} \alpha_2(1 - 2\delta(x_s, x_t)) + \sum_{s \in S} \beta(x_s)$ (α_1 corresponds to “horizontal” cliques C^H , and α_2 corresponds to “vertical” cliques C^V). Putting $\beta(\omega_1) = \beta_1, \dots, \beta(\omega_k) = \beta_k$, the real α in (2.4) is thus replaced by $\alpha = [\beta_1, \dots, \beta_k, \alpha_1, \alpha_2]$. The parameters $\alpha = [\beta_1, \dots, \beta_k, \alpha_1, \alpha_2]$, $q = [q_{ij}]_{1 \leq i, j \leq k}$, $a = [a_i]_{1 \leq i \leq k}$ and $m = [m_i]_{1 \leq i \leq k}$ are then estimated from W_1, \dots, W_n , in the following way.

- *Estimation of α .* The procedure, which is strictly the same as the Derin et al. method, consists of the following steps:

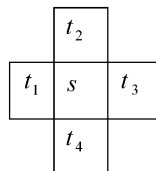


Fig. 2. Set W_s containing pixel s and its four neighbors.

- Step 1: Use the histogram techniques to estimate $p(x_s, x_{V_s})$;
 Step 2: Consider the following over-determined system of equations in terms of the probabilities and parameters:

$$[\Phi(\omega_i, x_{V_s}) - \Phi(\omega_j, x_{V_s})]^T \alpha = \ln \left(\frac{p(\omega_j, x_{V_s})}{p(\omega_i, x_{V_s})} \right); \tag{3.4}$$

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

$$I(\omega, x_{t_1}, \dots, x_{t_n}) = \begin{cases} -1 & \text{if } \omega = x_{t_1} = \dots = x_{t_n}, \\ +1 & \text{otherwise.} \end{cases} \tag{3.5}$$

and

$$J_k(\omega) = \begin{cases} +1 & \text{if } \omega = \omega_k, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

The location of t_i 's with respect to s is shown in Fig. 2.

Step 3: Solve the over-determined system using the least squares method.

- Estimation of $q = [q_{ij}]_{1 \leq i, j \leq k}$, $a = [a_i]_{1 \leq i \leq k}$, and $m = [m_i]_{1 \leq i \leq k}$
 As above, we consider a sequence of sets of pixels W_1, \dots, W_n as in Fig. 2, centered on pixels $1, \dots, n$. Let us denote by Y_{W_i} and X_{W_i} , respectively, the restriction of Y and X to W_i . So, we have $X_{W_i} = \begin{bmatrix} X_i \\ X_{V_i} \end{bmatrix}$ and $Y_{W_i} = \begin{bmatrix} Y_i \\ Y_{V_i} \end{bmatrix}$, 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}, \dots, 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, \dots, W_n a subsequence W'_1, \dots, W'_t such that $x_{W'_1} = \dots = x_{W'_t} = x_W$ and use $y_{W'_1}, \dots, y_{W'_t}$ to estimate m_{x_W} and Γ_{x_W} in a standard fashion by

$$\hat{m}_{x_W} = \frac{1}{t} \sum_{j=1}^t y_{W'_j}, \tag{3.7}$$

$$\hat{\Gamma}_{x_W} = \frac{1}{t} \sum_{j=1}^t (y_{W'_j} - \hat{m}_{x_W})^T (y_{W'_j} - \hat{m}_{x_W}). \tag{3.8}$$

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 given, respectively, by

$$M_{x_s} = m_{x_s} + AB^{-1}A^T(y_{V_i} - m_{V_i}), \tag{3.9}$$

$$\Sigma_{x_s}^2 = \sigma^2 - AB^{-1}A^T. \tag{3.10}$$

Comparing the estimates (3.7)–(3.10) with (2.6) and (2.7), we calculate m_i , a_i , and q_{ij} (with $i = x_s$). More precisely, m_{x_s} is given by (3.7), a_{x_s} is given by $a_{x_s} = \frac{1}{\Sigma_{x_s}^2}$ (see

(2.7)) with $\Sigma_{x_s}^2$ given by (3.10). Further, putting $q_{x_s} = (q_{x_s x_{i_1}}, q_{x_s x_{i_2}}, q_{x_s x_{i_3}}, q_{x_s x_{i_4}})$ and comparing (3.9) to (2.6) gives

$$q_{x_s} = -2a_{x_s}AB^{-1}. \tag{3.11}$$

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

$$\hat{m}_i = \frac{1}{n} \sum_{t=1}^r d_t m_i^t, \quad \hat{a}_i = \frac{1}{n} \sum_{t=1}^r d_t a_i^t, \quad \text{and} \quad \hat{q}_{ij} = \frac{1}{n} \sum_{t=1}^r d_t q_{ij}^t. \tag{3.12}$$

Finally, the approach we propose can be summarized in the following steps:

- Step 1: Estimation of α using Derin et al.'s method described above;
- Step 2: For each configuration x_W , estimation of means m_{x_W} and variance–covariance matrices Γ_{x_W} with (3.7) and (3.8);
- Step 3: For each configuration x_W , computation of m_i , a_{x_s} , and the $q_{x_s} = (q_{x_s x_{i_1}}, q_{x_s x_{i_2}}, q_{x_s x_{i_3}}, q_{x_s x_{i_4}})$ with (3.9) and (3.11);
- Step 4: Calculation of final q_{ij} , a_i , and m_i with (3.12) applied to the estimates obtained from r configurations of x_W used.

Let us remark that it is not necessary—and it would even be impossible when the number of classes becomes large—to take all the configurations x_W into account. For example, one could decide to consider only the configurations with no more than two classes.

Otherwise, let us notice that the NMCD above can be used in some practical situations. For example, imagine that the ground truth is known on some part of a country, and it has to be estimated on some other part. Thus, all parameters needed can be estimated from the known part with the NMCD, and used to segment the other part by some Bayesian method.

Finally, using NMCD inside ICE produces the new NMID we propose, which is summarized in Fig. 3.

Let us also mention the important problem of estimating the number of classes, which is not addressed here. It can be viewed as a particular model selection problem, and different techniques have been proposed in the classical HMF case (see, among others, the recent paper [11] and the references therein). Thus, extending these techniques to TMF will undoubtedly be among the interesting perspectives of further work.

4. Unsupervised image segmentation

This section is devoted to different experiments aiming to show the interest of the new ICE-based parameter estimation method and unsupervised image segmentation

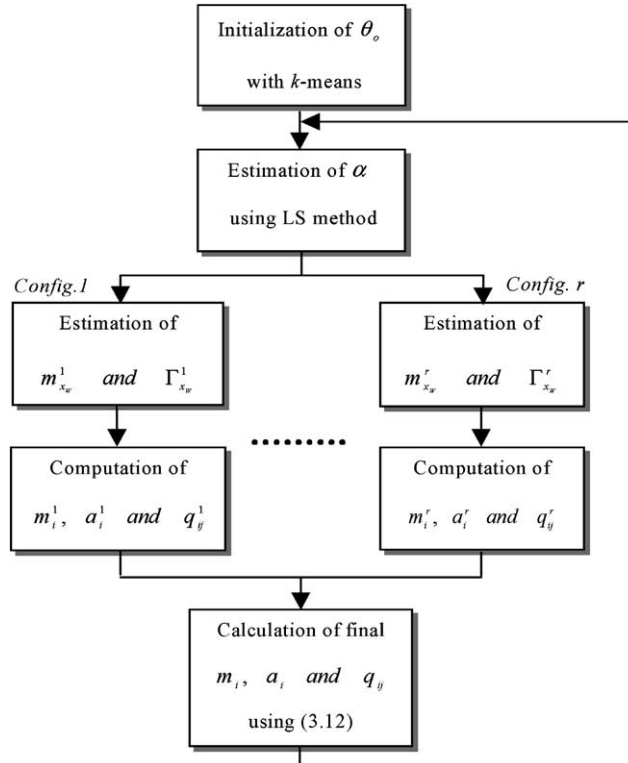


Fig. 3. Parameter estimation with NMID.

based on it, in both PMF and TMF cases. Subsection 4.1 contains comparisons of the new parameter estimation method with a classical one in the case of simulated PMF and synthetic images, while Subsection 4.2 aims to show that the use of TMF instead of PMF or HMF-IN can be of interest. Finally, an example of unsupervised segmentation of a real image, where TMF perform better than HMF-IN, is presented in Subsection 4.3.

4.1. Simulated PMF and synthetic images

This subsection is devoted to two series of experiments. The first concerns simulated PMF whose energy is slightly more general than the energy given by (2.4): we will consider $U(x, y) = \frac{1}{2} \left[\sum_{s \in S} a_{x_s} (y_{x_s} - m_{x_s})^2 + \sum_{(s,t) \in C^H} \alpha_1 (1 - 2\delta(x_s, x_t)) + \sum_{(s,t) \in C^V} \alpha_2 (1 - 2\delta(x_s, x_t)) + \sum_{(s,t) \in C} q_{x_s x_t} (y_{x_s} - m_{x_s})(y_{x_t} - m_{x_t}) \right]$, where C^H is the set of horizontal cliques and C^V is the set of vertical ones. For the case of two classes considered, the vector of parameters to be estimated is thus $\theta = (\alpha_1, \alpha_2, a_1, a_2, m_1, m_2, q_{11}, q_{22}, q_{12})$. The aim is to compare the estimates obtained with the four methods SGCD, SGID, NMCD, and NMID (SG for stochastic gradient, NM for new method, CD for

complete data, and ID for incomplete data), and to look at how these different estimates act upon the unsupervised segmentation methods based on them.

We have carried out several experiments and the results of one of them, representative of various results we obtained, are presented in Table 1 (see Fig. 4). In general terms, we reached the following conclusions:

- (i) the estimates SGCD and NMCD are of comparable efficiency, which is rather good, even in the case of very strong noise, in the sense that there is little dif-

Table 1
Results corresponding to Fig. 4

	TP	New method		SG	
		NMCD	NMID	SGCD	SGID
α_1	2	2.1	1.87	2.36	2.24
α_2	2	2.01	1.68	2.37	2.24
a_1	1	1	1.06	0.91	0.58
a_2	1	1.03	1.09	1.00	1.03
m_1	0	0.00	-0.01	0.10	0.26
m_2	2	2.00	2.07	1.86	0.78
q_{11}	-0.2	-0.19	-0.19	-0.2	-0.06
q_{22}	-0.2	-0.2	-0.19	-0.21	-0.40
q_{12}	0.00	0.00	0.00	0.02	0.00
ER	1.44%	1.61%	1.63%	2.03%	8.59%

True parameters (TP), estimates from complete data with NMCD and SGCD, and estimates from observed data with NMID and SGID. ER, error ratio of MPM segmentation.

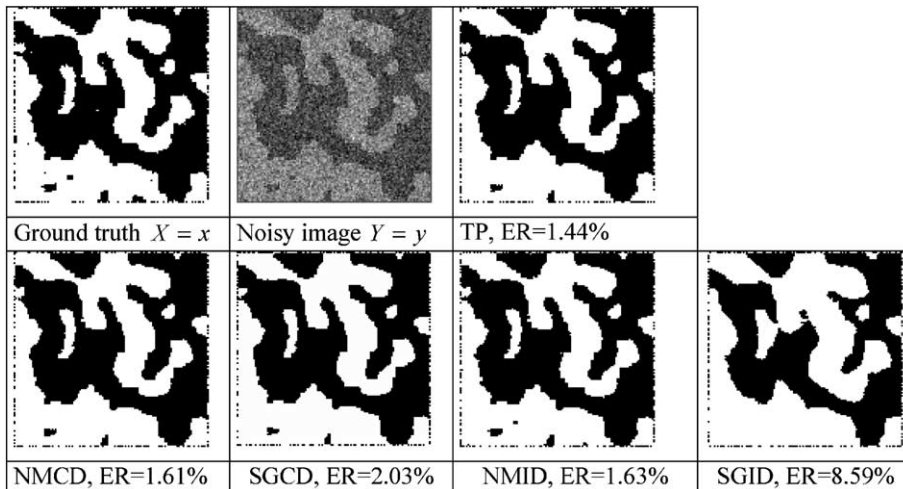


Fig. 4. Simulated PMF (size 128×128) and comparison between PMF segmentation using true parameters (TP) and estimates by different parameters estimation methods.

ference when segmenting the noisy image with real parameters or estimated ones;

- (ii) the estimate NMID works better than SGID in all situations.

In the case of SG estimation, 15 iterations have been used from the complete data and 30 iterations from the incomplete data. In NMID we have used 30 iterations. These numbers of iterations have been found experimentally, from the numerous experiments we performed.

Note that the error rates in all experiments in the paper are obtained from one experiment.

The second series of experiments concerns a ring-image corrupted with synthetic Gaussian correlated noise. We have partial knowledge of the model; in other words, the model corresponds partially to PMF above. Such models are interesting because studying them allows one to understand how the methods work when the model used diverges from the theoretical one. In other words, such studies provide some knowledge about the robustness of the methods considered.

We consider the two-class image “Ring” and its noisy observed version $Y = y$. Each Y_s is then obtained from an independent noise $W = (W_s)_{s \in S}$ by averaging on the four nearest neighbors: $Y_s = \frac{1}{5}(W_s + \sum_{t \in \mathcal{V}_s} W_t)$. 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 black, and with the Gaussian distribution $N(m_2, \sigma_2^2)$ if x_t is white, and then $Y_s = y_s$ is calculated, for each $s \in S$, using the equality above. Of course, such a correlated Gaussian field is not Markovian and thus the model considered is not of the PMF type considered in the previous subsection.

As above, we performed numerous experiments and one case, representative of the different results obtained, is specified in Table 2. On the whole, analogous conclusions to those drawn in the previous subsection hold. In general, SGCD and NMCD are quite efficient, even in very noisy cases. When considering the incomplete data situation, NMID works better than SGID (see Fig. 5).

Table 2
Image “ring”

	New method		SG	
	NMCD	NMID	SGCD	SGID
α_1	2.01	1.17	2.38	2.30
α_2	2.01	1.19	2.38	2.30
a_1	7.34	8.19	4.88	2.04
a_2	7.74	8.23	3.43	2.05
m_1	0.01	-0.05	0.01	0.02
m_2	0.99	1.04	0.98	0.60
q_{11}	-0.50	-0.50	-0.35	0.00
q_{22}	-0.48	-0.45	-0.23	-0.50
q_{12}	0.00	0.00	-0.01	0.00
ER	3.1%	5.5%	3.65%	15.3%

Results corresponding to image “ring” in Fig. 5. Estimates from complete data with NMCD and SGCD, and estimates from observed data with NMID and SGID. ER, error rate of MPM unsupervised segmentation. Real parameters are $m_1 = 0$, $m_2 = 0.5$, and $\sigma_1^2 = \sigma_2^2 = 1$.

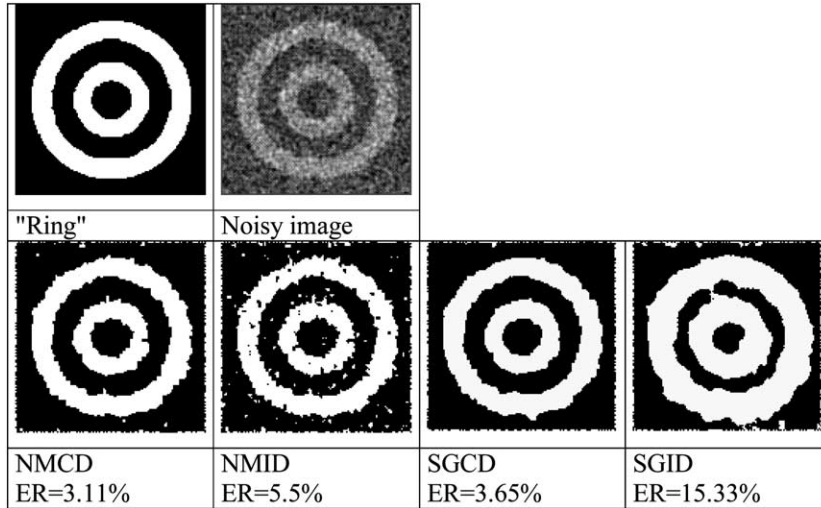


Fig. 5. Image “ring” (size 128×128), its noisy version, and different unsupervised segmentation results. ER, error rate MPM unsupervised segmentation. Computation time is about 80 s for both NMID and SGID, the parameter estimation taking about 2 s in NMID and 6 s in SGID.

Both methods NMIC and SGID have been initialized with k -means procedure with 20 iterations. Further, we used 20 iterations to estimate the parameters, and 100 samples to estimate the posterior marginal distributions $p(x_s|y)$ used in MPM. The optimal number of different iterations has been found experimentally. The computation time is about 80 s for both NMID and SGID, the parameter estimation taking about 2 s in NMID and 6 s in SGID. All experiments in this paper have been performed with IBM PC, Pentium IV, 2 GHz.

4.2. PMF and TMF based unsupervised segmentation

This subsection is devoted to an example showing one kind of situation in which TMF perform better than PMF. Let us consider a two-classes hand-drawn image $X = x$ depicted as “synthetic data” in Fig. 6, and its observed noisy version $Y = y$. The latter is the class image corrupted with Gaussian white noise; so $p(y_s|x_s = \omega_1) = N(0, 1)$ and $p(y_s|x_s = \omega_2) = N(2, 1)$. The observed image $Y = y$ is then segmented using HMF-IN, PMF, and TMF models, all segmentations being performed in an unsupervised manner with parameters estimated by the new method proposed above. All HMF-IN, PMF, and TMF used are Markovian with respect to the four nearest neighbors, and in TMF the latent random variables U_s take their values in $\Lambda = \{\lambda_1, \lambda_2\}$. The energies of the models considered are $G_{\text{HMF}}(x, y) = \sum_{(s,t) \in C} \alpha(1 - 2\delta(x_s, x_t)) - \sum_{s \in S} \log[p(y_s|x_s)]$, $G_{\text{PMF}}(x, y) = \sum_{(s,t) \in C} \alpha(1 - 2\delta(x_s, x_t)) + \sum_{s \in S} a_{x_s}(y_{x_s} - m_{x_s})^2 + \sum_{(s,t) \in C} q_{x_s x_t}(y_{x_s} - m_{x_s})(y_{x_t} - m_{x_t})$, and $G_{\text{TMF}}(x, u, y) = \sum_{(s,t) \in C} \alpha(1 - 2\delta(u_s, u_t)) - (\alpha' \delta(u_s = u_t = \lambda_1) + \alpha \delta(u_s = u_t = \lambda_2))(1 - \delta(x_s, x_t)) - \sum_{s \in S} \log[p(y_s|x_s)]$, respectively. We have chosen the HMF-IN mod-

Synthetic data	Noisy data	Optimal « pixel by pixel » method : ER=13.85%
HMF-IN ID : ER=8.22%	PMF ID : ER=7.55%	TMF ID : ER=4.94%

Fig. 6. Image “Synthetic data” (size 128×128), HMF-IN, TMF, and PMF-based unsupervised segmentation methods results, and “pixel by pixel” true parameters based one. ER, ratio of wrongly classified pixels.

el, known as “Pott’s model,” as the simplest one. The PMF model chosen is the simplest PMF generalizing the HMF-IN considered; in fact, when the noise correlations disappear, the latter PMF model gives the former HMF-IN one. Similarly, the TMF (X, U, Y) considered is a simple one again giving the HMF (X, Y) for $U = X$ (let us recall that TMF generalize PMF because the latter are particular TMF for which $U = X$). In fact, when $u_s = x_s$ and $u_t = x_t$ in $G_{\text{TMF}}(x, u, y)$, we have $(\alpha' \delta(u_s = u_t = \lambda_1) + \alpha' \delta(u_s = u_t = \lambda_2))(1 - \delta(x_s, x_t)) = 0$, which is due to the fact that $\alpha' \delta(u_s = u_t = \lambda_1) + \alpha' \delta(u_s = u_t = \lambda_2) = 0$ for $u_s \neq u_t$ and $1 - \delta(x_s, x_t) = 0$ for $u_s = u_t$. So, $U = X$ implies $G_{\text{TMF}} = G_{\text{HMF}}$.

According to the experiments presented in Fig. 6 we see that the TMF-based method works significantly better than the PMF-based one, and the latter works slightly better than the classical HMF-based method. Thus, we have here a simple example of a situation in which PMF contribute little to HMF, but the use of TMF is of interest and significantly improves both HMF- and PMF-based methods results.

We have also added in Fig. 6 the optimal “pixel by pixel” segmentation, in which each \hat{x}_s is obtained by maximizing $p(x_s|y_s)$. The latter also means that \hat{x}_s is obtained by maximizing $p(x_s = \omega_1)p(y_s|x_s = \omega_1)$ and $p(x_s = \omega_2)p(y_s|x_s = \omega_2)$. As the proportions $p(x_s = \omega_1) = 0.26$ and $p(x_s = \omega_2) = 0.74$ are efficiently estimated from synthetic data and the two Gaussian densities $p(y_s|x_s = \omega_1)$ and $p(y_s|x_s = \omega_2)$ are known, the “pixel by pixel” segmentation is optimal and thus this complementary result incident-

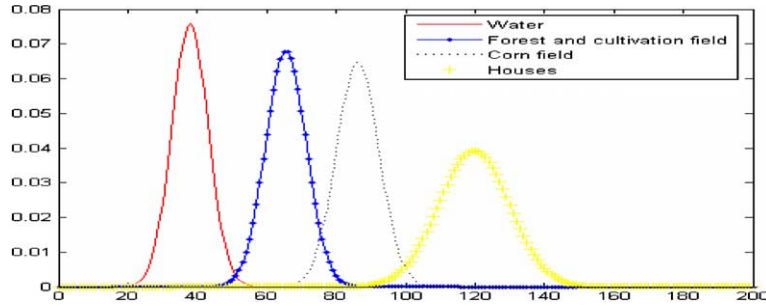


Fig. 7. Four Gaussian distributions (“Water,” “Forest and cultivation field,” “Corn field,” and “Houses,” respectively) estimated with NMID from Istres image in Fig. 8. Estimated $\alpha = (\alpha_H, \alpha_V)$ is (0.93; 0.89) and estimated $q_{x,y}$ are $-0.12, -0.09, -0.07,$ and -0.02 .

tally shows the usefulness of Markov models, in which the use of $p(x_s|y_s)$ is replaced by the use of $p(x_s|y)$ (see Section 2.1).

The computation time is comparable to that of the previous subsection.

4.3. Real image segmentation

This section is devoted to the segmentation of a real radar image. The considered image, shown in Fig. 8, is an image of Istres city, which can be divided into four classes: (1) Water, (2) Forest and cultivation, (3) Corn field, and (4) Houses, with the noise distributions given in Fig. 7. We want to obtain two classes: “Water” and “Land,” the latter regrouping “Forest and cultivation,” “Corn fields,” and “Houses.” Thus, we use two different procedures: the first considers classical PMF, given by (2.4), with two classes, estimating all parameters with NMID as described above. The second considers the same PMF (2.4) with four classes, and then considers that “Land” is union of three of them (which gives a TMF: see the second interpretation in Subsection 2.4). All parameters $\theta = (\alpha, a_i, m_i, q_{ii})_{1 \leq i \leq 4}$ are still estimated with NMID (we have fixed $q_{ij} = 0$ for $i \neq j$). We have used 30 iterations in ICE and the posterior marginal distributions, used in the MPM segmentation, are estimated from 100 samples, each of which was sampled with 20 Gibbs sampler iterations.

5. Conclusions and perspectives

In this paper, we have dealt with the recent triplet Markov fields (TMF [27]) model, which is more general than the classical hidden Markov fields (HMF) model and still allows one to recover, via some Bayesian methods, the hidden signal. We proposed two kinds of results:

- (i) the first concerns the TMF models themselves. We specified how several variants can be simply obtained from classical HMF with correlated noise. None of these new variants is much more complicated to use than the classical HMF, and thus new possibilities are offered to deal with each particular problem.

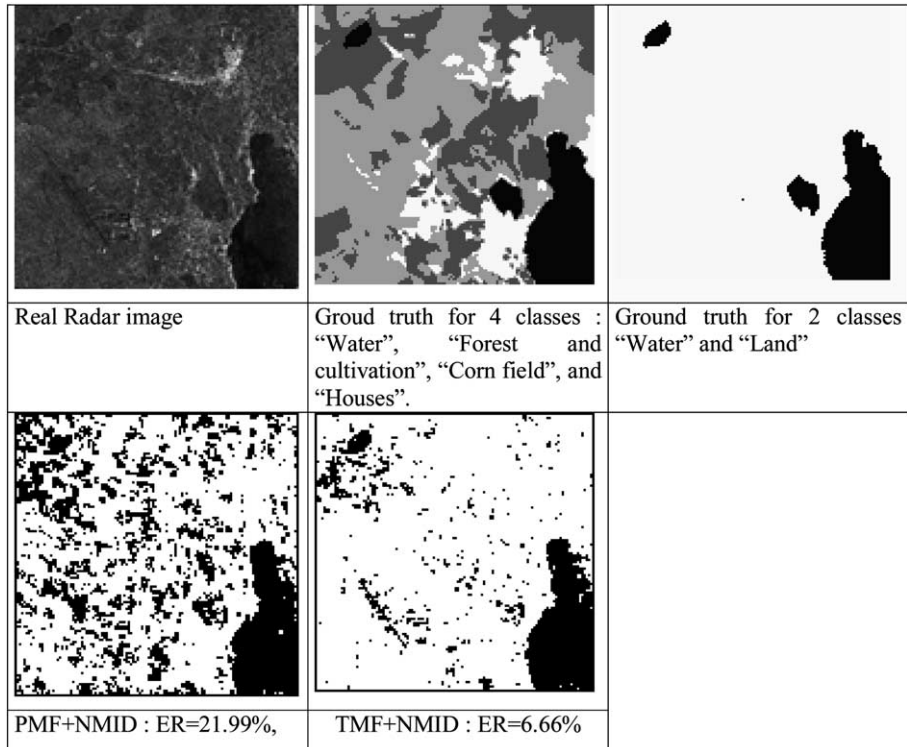


Fig. 8. Image “Istres” (size 256×256), ground truth, and two unsupervised segmentation results obtained with PMF and TMF models. ER: error rate with respect to the two classes ground truth. Computational time: 316 s for PMF + NMID, and 647 s for TMF + NMID.

- (ii) The second main aim was to introduce, and assess by experiment, some original parameter estimation methods. We proposed two of them, respectively, based on Stochastic Gradient (SG [35]) and Iterative Conditional Estimation (ICE [25]) with a generalization of least squares (LS [9]). It turned out that ICE with LS is quite efficient and provides, in particular, interesting unsupervised Bayesian image segmentation methods.

As a perspective for further studies, let us mention the extension to more complex noises, possibly non-Gaussian, using copulas [5]. An integration of different sensors, some of which being possibly “evidential” [2], is another perspective for further investigation of TMF. Finally, extension of the rectangular lattice considered here to more general “Bayesian networks” [13] could also be viewed.

References

- [1] D. Benboudjema, W. Pieczynski, Parameter estimation in pairwise Markov fields, in: *Advanced Concepts for Intelligent Vision Systems (ACVIS 04)*, Aug. 31–Sept. 3, Brussels, Belgium, 2004.

- [2] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, W. Pieczynski, Multisensor images segmentation using Dempster–Shafer fusion in Markov fields context, *IEEE Trans. Geosci. Remote Sensing* 39 (8) (2001) 1789–1798.
- [3] J. Besag, On the statistical analysis of dirty pictures, *J. R. Stat. Soc. Ser. B* 48 (1986) 259–302.
- [4] J.-M. Boucher, P. Lena, Unsupervised Bayesian classification, application to the forest of Paimpont (Brittany), *Photo Interpretation* 32 (1994/4, 1995/1–2) (1995) 79–81.
- [5] N. Brunel, W. Pieczynski, Unsupervised signal restoration using Copulas and Pairwise Markov chains, in: *IEEE Workshop on Statistical Signal Processing (SSP 2003)*, Saint Louis, Missouri, September 28–October 1, 2003.
- [6] B. Chalmond, An iterative Gibbsian technique for reconstruction of m-ary images, *Pattern Recogn.* 22 (1989) 747–761.
- [7] Y. Delignon, A. Marzouki, W. Pieczynski, Estimation of generalized mixture and its application in image segmentation, *IEEE Trans. Image Process.* 6 (10) (1997) 1364–1375.
- [8] J.-P. Delmas, An equivalence of the EM and ICE algorithm for exponential family, *IEEE Trans. Signal Process.* 45 (10) (1997) 2613–2615.
- [9] H. Derin, H. Elliot, Modeling and segmentation of noisy and textured images using Gibbs random fields, *IEEE Trans. Pattern Anal. Mach. Intell.* 9 (1987) 39–55.
- [10] F. Destremes, M. Mignotte, A statistical model for contours in images, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (5) (2004) 626–638.
- [11] F. Forbes, N. Peyrard, Hidden Markov random field model selection criteria based on mean field-like approximations, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (9) (2003) 1089–1101.
- [12] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* 6 (1984) 721–741.
- [13] Z. Ghahramani, An introduction to hidden Markov models and Bayesian networks, *Int. J. Pattern Recogn. Artif. Intell.* 15 (1) (2001) 9–42.
- [14] X. Guyon, *Random Fields on a Network, Probability and its Applications*, Springer-Verlag, Berlin, 1995.
- [15] M.V. Ibáñez, A. Simó, Parameter estimation in Markov random field image modeling with imperfect observations. A comparative study, *Pattern Recogn. Lett.* 24 (14) (2003) 2377–2389.
- [16] Z. Kato, J. Zeroubia, M. Berthod, Unsupervised parallel image classification using Markovian models, *Pattern Recogn.* 32 (1999) 591–604.
- [17] S. Lakshmanan, H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (8) (1989) 799–813.
- [18] P. Lanchantin, W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chains using evidential priors, *IEEE Trans. Signal Process.* 53 (8) (2005) to appear.
- [19] J. Marroquin, S. Mitter, T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *J. Am. Stat. Assoc.* 82 (1987) 76–89.
- [20] G.J. McLachlan, T. Krishnan, *EM Algorithm and Extensions*, in: *Probability and Statistics*, Wiley, New York, 1997.
- [21] D. Melas, S.P. Wilson, Double Markov random fields and Bayesian image segmentation, *IEEE Trans. Signal Process.* 50 (2) (2002) 357–365.
- [22] M. Mignotte, C. Collet, P. Pérez, P. Bouthémy, Sonar image segmentation using an unsupervised hierarchical MRF model, *IEEE Trans. Image Process.* 9 (7) (2000) 1216–1231.
- [23] M. Mignotte, J. Meunier, J.-P. Soucy, C. Janicki, Comparison of deconvolution techniques using a distribution mixture parameter estimation: application in single photon emission computed tomography imagery, *J. Electron. Imaging* 11 (1) (2002) 11–25.
- [24] P. Pérez, Markov random fields and images, *CWI Q.* 11 (4) (1998) 413–437.
- [25] W. Pieczynski, Statistical image segmentation, *Mach. Graph. Vision* 1 (1/2) (1992) 261–268.
- [26] W. Pieczynski, A.-N. Tebbache, Pairwise Markov random fields and segmentation of textured images, *Mach. Graph. Vision* 9 (3) (2000) 705–718.
- [27] W. Pieczynski, D. Benboudjema, P. Lanchantin, Statistical image segmentation using triplet Markov fields, in: *SPIE’s Internat. Symp. on Remote Sensing*, September 22–27, Crete, Greece, 2002.

- [28] W. Pieczynski, C. Hulard, T. Veit, Triplet Markov Chains in hidden signal restoration, in: SPIE's Internat. Symp. on Remote Sensing, September 22–27, Crete, Greece, 2002.
- [29] W. Pieczynski, D. Benboudjema, Triplet Markov fields and theory of evidence, *Image Vision Comput. J.* (2003) submitted.
- [30] S. Reed, Y. Petillot, J. Bell, An automatic approach to the detection and extraction of mine features in sidescan sonar, *IEEE J. Oceanic Eng.* 28 (1) (2003) 90–105.
- [31] F. Salzenstein, W. Pieczynski, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, *Graph. Models Image Process.* 59 (4) (1997) 205–220.
- [32] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.
- [33] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*, Springer, Berlin, 2003.
- [34] Y. Yu, Q. Cheng, MRF parameter estimation by an accelerated method, *Pattern Recogn. Lett.* 24 (9–10) (2003) 1251–1259.
- [35] L. Younes, Parametric inference for imperfectly observed Gibbsian fields, *Probab. Theory Relat. Fields* 82 (1989) 625–645.