J Phys IV France **12** (2002)
© EDP Sciences, Les Ulis
DOI: 10 1051/jp42002005

Markov random fields in image processing application to remote sensing and astrophysics

J. Zerubia, A. Jalobeanu and Z. Kato[1]

Ariana, CNRS/INRIA/UNSA, Joint Research Group INRIA, BP. 93, 06902 Sophia Antipolis cedex, France
[1] *Computer Science Department, School of Computing, NUS, 10 Kent Ridge Crescent, Singapore 119260, Singapore*

1 INTRODUCTION

Early vision directly deals with raw pixel data involving image compression, restoration [12, 25, 66], edge detection [27, 62, 66], segmentation [10, 23, 35, 40, 59], texture analysis [17–19, 46], motion detection, optical flow, etc ... Most of these problems can be formulated within a general framework, called *image labeling*, where we associate a label to each pixel from a finite set. The meaning of this label depends on the problem that we are trying to solve. For image restoration, it means a grey-level; for edge detection, it means the presence or the direction of an edge; for image segmentation, it means a class (or region); etc ... The problem here is how to choose a label for a pixel, which is *optimal* in a certain sense.

Our approach is probabilistic. at each pixel, we want to select the most likely labeling. To achieve this goal, we need to define some probability measure on the set of all possible labelings. In real scenes, neighboring pixels have usually similar intensities; edges are smooth and often straight. In a probabilistic framework, such regularities are well expressed by Markov Random Fields (MRF) [15, 21, 32, 49, 64]. Another reason for dealing with MRF models [1, 26, 55] is of course the Hammersley-Clifford theorem [6, 52] which allows to define MRF through clique-potentials. In the labeling problem, this leads to the following Bayesian formulation [54]: we are looking for the Maximum A Posteriori (MAP) estimate of the label field yielding to the minimization of a usually non-convex energy function.

Unfortunately, finding such an estimate is a heavy computing problem. For example, if we consider a 16×16 image with only two possible labels for each pixel, we obtain a configuration space of 2^{256} elements. It is then impossible to find the optimum by computing the possible values of the cost function. On the other hand, due to the non-convexity, classical gradient descent methods cannot be used since they get stuck in a local minimum. In the early 80's, a Monte-Carlo algorithm [31, 58], called Simulated Annealing [24, 63], has been proposed independently by Černy [11] and Kirkpatrick *et al.* [45] to tackle the optimization problem. However, with the first substantial mathematical results [25, 34], it becomes clear that successful applications of Simulated Annealing (SA) require a very slow temperature cooling schedule and thus large computing time. To avoid this drawback, two solutions have been proposed: One of them deals with the possible parallelization of SA algorithms [2]. The other solution is to use *deterministic* algorithms, which are suboptimal, but converge within a few iterations requiring therefore less computing time [5].

On the other hand, multigrid [8,41,57] models can also significantly improve the convergence rate and the quality of the final result of iterative relaxation techniques. Multigrid methods have a long existence in numerical analysis (partial differential equations, for instance). In image processing, they have also been used in various context from the mid 70's [41]. Hierarchical methods [33] applied to MRF image modeling have been more and more popular [3, 14, 47]. The essence of such an approach is to represent an image model at multiple resolutions from coarse to fine.

We are talking about *multigrid* methods, if the layers in the pyramid are not connected. In this case, the optimization algorithm is usually parallelizable only on the layers, but it is still sequential between layers. An important question in multigrid modeling is how to define the cliques and their potentials at coarse resolutions. There are various ideas [48] including the Renormalization Group approach of Gidas *et al.* [30, 53], a consistent multiscale approach of Heitz *et al.* [37], or Bouman's causal pyramidal model [9].

If there is an inter-level communication, the model is called 3D pyramidal [33]. While the optimization algorithms associated with such models can be parallelized on the whole pyramid, the underlying MRF model becomes more complicated requiring more computation but getting better results [42, 43].

2 FUNDAMENTALS

At the beginning of the previous century, mostly inspired by the Ising model, a new type of stochastic process appeared in the theory of probability called *Markov Random Fields* (MRF). MRF become rapidly a broadly used tool in various problems, including statistical mechanics. Its use in image processing became popular with the two famous papers of [18] and [25] in the 80's. Below, we briefly give an introduction to the theory of MRF [20, 44, 52].

2.1 The Ising Model

Following Ising, we consider a sequence, $0, 1, 2, \ldots, n$ on the line. At each point, there is a spin which is either *up* or *down* at any given time (see Fig. 1). We define a probability measure on the set Ω of all possible configurations $\omega = (\omega_0, \omega_1, \ldots, \omega_n)$. In this context, each spin is a function

$$\delta_i(\omega) = \begin{cases} 1 & \text{if } \omega_i \text{ is } up \\ -1 & \text{if } \omega_i \text{ is } down \end{cases} \tag{1}$$

An *energy* $U(\omega)$ is assigned to each configuration:

$$U(\omega) = -J \sum_{i,j} \delta_i(\omega)\delta_j(\omega) - mH \sum_i \delta_i(\omega) \tag{2}$$

In the first sum, Ising made the simplifying assumption that only interactions of points with one unit apart need to be taken into account. This term represents the energy caused by the spin-interactions. The constant J is a property of the material. If $J > 0$, the interactions tend to keep neighboring spins in the same directions (*attractive case*). If $J < 0$, neighboring spins with opposite orientation are favored (*repulsive case*). The second term represents the influence of an external magnetic field of intensity H and $m > 0$ is a property of the material. The probability measure on Ω is then given by

$$P(\omega) = \frac{\exp\left(-\frac{1}{kT}U(\omega)\right)}{Z}, \tag{3}$$

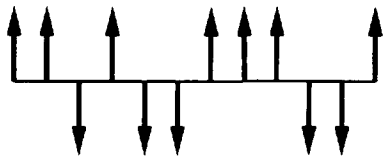

Fig. 1. One dimensional Ising model

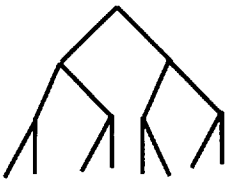

Fig. 2. Cayley tree model.

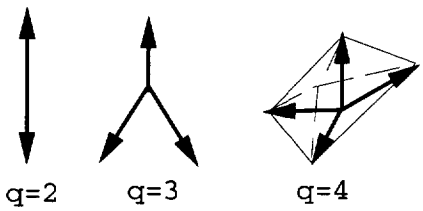

Fig. 3. Potts model.

where T is the temperature and k is a universal constant. The normalizing constant (or *partition function*) Z is defined by

$$Z = \sum_{\omega \in \Omega} \exp\left(-\frac{1}{kT} U(\omega)\right). \tag{4}$$

The probability measure defined in Eqn. (3) is called a *Gibbs distribution.* One could extend the model to two dimensions in a natural way. The spins are arranged on a lattice, they are represented by two coordinates and a point have 4 neighbors unless it is on the boundary. In the 2D case, the limiting measure P is unstable, there is a *phase transition.* As it is pointed out in [44], considering the *attractive* case and an external field h, the measure P_h converges to P^- if h goes to zero through negative values but it converges to $P^+ \neq P^-$ if h goes to zero through positive values. It has been shown that there exists a *critical temperature* T_C and below this temperature phase transition always occurs. The temperature depends on the vertical (J_1) and horizontal (J_2) interaction parameters.

As a special example, we mention the *Cayley tree model* [44], originally proposed by Bethe [4] as an approximation to the Ising model. In this case, the points sit on a tree (see Fig. 2). The root is called the 0^{th} level. From the root, we have q branches ($q = 2$ in Fig. 2). The $q = 1$ case simply gives a 1D Markov chain. A configuration on a tree of n levels is an assignment of a label *up* or *down* to each point. A similar energy function can be defined as for the Ising model.

Another extension of the Ising model to more than two states per point is the well known Potts model [59, 65]. The problem is to look at the Ising model as a system of interacting spins that can be either parallel or antiparallel. More generally, we consider a system of spins, each spin pointing one of the q equally spaced directions. These vectors are the linear combinations of q unit vectors pointing in the q symmetric directions of a hypertetrahedron in $q - 1$ dimensions. For $q = 2, 3, 4$, examples are shown in Fig. 3. The energy function of the

Potts model can be written as

$$U(\omega) = \sum_{i,j} J(\Theta_{ij}), \tag{5}$$

where $J(\Theta)$ is 2π periodic and Θ_{ij} is the angle between two neighboring spins in i and j. The $q = 2$ case is equivalent to the Ising model.

2.2 Gibbs Distribution and MRF

The most natural way to define MRF [1, 25] related to image models is to define them on a lattice. However, here we will define MRF more generally on graphs (for an excellent overview, see [55]). Let $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ be a graph where $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$ is a set of vertices (or sites) and $\mathcal{E}$ is the set of edges.

2.2.1 Neighbors

Definition: Two points s_i and s_j are neighbors if there is an edge $e_{ij} \in \mathcal{E}$ connecting them. The set of points which are neighbors of a site s (ie. the neighborhood of s) is denoted by $\mathcal{V}_s$.

2.2.2 Neighborhood system

Definition: $\mathcal{V} = \{\mathcal{V}_s \mid s \in \mathcal{S}\}$ is a neighborhood system for $\mathcal{G}$ if

1. $s \notin \mathcal{V}_s$
2. $s \in \mathcal{V}_r \Leftrightarrow r \in \mathcal{V}_s$

To each site of the graph, we assign a label λ from a finite set of labels Λ. Such an assignment is called a configuration ω having some probability $P(\omega)$. The restriction to a subset $T \subset \mathcal{S}$ is denoted by ω_T and $\omega_s \in \Lambda$ denotes the label given to the site s. In the following, we are interested in the probability measures assigned to the set Ω of all possible configurations. First, let us define the *local characteristics* as the conditional probabilities $P(\omega_s \mid \omega_r, r \neq s)$.

2.2.3 Markov Random Field

Definition: $\mathcal{X}$ is a Markov Random Field (MRF) with respect to $\mathcal{V}$ if

1. for all $\omega \in \Omega$: $P(\mathcal{X} = \omega) > 0$,
2. for every $s \in \mathcal{S}$ and $\omega \in \Omega$:

$$P(X_s = \omega_s \mid X_r = \omega_r, r \neq s) = P(X_s = \omega_s \mid X_r = \omega_r, r \in \mathcal{V}_s).$$

To continue our discussion about probability measures on Ω, the notion of *cliques* will be very useful.

2.2.4 Clique

Definition: A subset $C \subseteq \mathcal{S}$ is a clique if every pair of distinct sites in C are neighbors. $\mathcal{C}$ denotes the set of cliques and $\deg(\mathcal{C}) = \max_{C \in \mathcal{C}} \mid C \mid$.

Using the above definition, we can define a *Gibbs measure* on Ω. Let V be a *potential* which assign a number $V_C(\omega)$ to each subconfiguration ω_C. V defines an *energy* $U(\omega)$ on Ω by

$$U(\omega) = -\sum_T V_T(\omega). \tag{6}$$

2.2.5 Gibbs distribution

Definition: A Gibbs distribution is a probability measure π on Ω with the following representation:

$$\pi(\omega) = \frac{1}{Z} \exp\left(-U(\omega)\right), \tag{7}$$

where Z is the normalizing constant:

$$Z = \sum_\omega \exp\left(-U(\omega)\right),$$

The following theorem establish the equivalence between Gibbs measures and MRF [6,52].

Theorem 1 (Hammersley-Clifford theorem). $\mathcal{X}$ is a MRF with respect to the neighborhood system $\mathcal{V}$ if and only if $\pi(\omega) = P(\mathcal{X} = \omega)$ is a Gibbs distribution, that is

$$\pi(\omega) = \frac{1}{Z} \exp\left(-\sum_{C \in \mathcal{C}} V_C(\omega)\right) \tag{8}$$

The main benefit of this equivalence is that it provides us with a simple way to specify MRF, namely specifying potentials instead of local characteristics, which is usually very difficult.

2.3 Spatial Lattice Schemes

In this section, we deal with a particular subclass of MRF which are the most commonly used schemes in image processing. We consider $\mathcal{S}$ as a lattice $\mathcal{L}$ so that $\forall s \in \mathcal{S} : s = (i, j)$ and define the so-called n^{th} *order homogeneous neighborhood systems* as

$$\mathcal{V}^n = \{\mathcal{V}^n_{(i,j)} : (i,j) \in \mathcal{L}\}, \tag{9}$$

$$\mathcal{V}^n_{(i,j)} = \{(k,l) \in \mathcal{L} : (k-i)^2 + (l-j)^2 \leq n\}. \tag{10}$$

Obviously, sites near the boundary have fewer neighbors than interior ones (free boundary condition). Furthermore, $\mathcal{V}^0 \equiv \mathcal{S}$ and for all $n \geq 0 : \mathcal{V}^n \subset \mathcal{V}^{n+1}$. Figure 4 shows a first-order neighborhood corresponding to $n = 1$. The cliques are $\{(i,j)\}, \{(i,j),(i,j+1)\}, \{(i,j),(i+1,j)\}$. Figure 5 shows a second order neighborhood. In practice, third order or larger systems are rarely used since the energy function would be too complicated and will require a lot of computation.

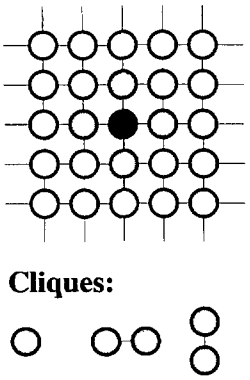

Fig. 4. First order neighborhood system.

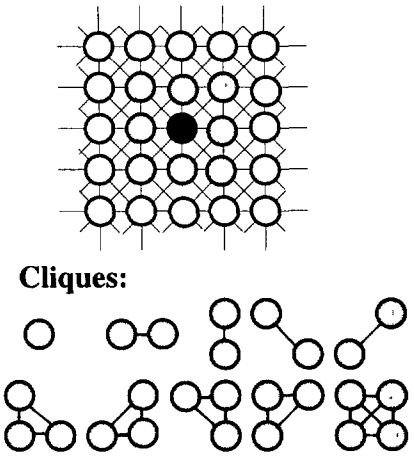

Fig. 5. Second order neighborhood system.

3 A GENERAL MARKOVIAN IMAGE MODEL

MRF models in image processing have become popular with the famous paper of S. Geman and D. Geman on image restoration [25]. The field has grown up in recent years addressing a variety of tasks such as [1]:

Compression: Find a new image as close as possible to the original one but described at a much smaller cost (MRF are used in other compression problems too).

Restoration/deconvolution: Observing a degraded image, which can be blurred and noisy, approximately recover the original one.

Edge Detection: Find smooth boundaries separating image regions.

Segmentation: Partition the image into homogeneous regions where homogeneity is measured in terms of grey-levels or texture characteristics.

Motion Detection: In a sequence of images, try to find a field of velocities linking one image to the next one.

We now turn to the mathematical formulation of a MRF image model. Let $\mathcal{R} = \{r_1, r_2, \ldots, r_M\}$ be a set of sites and $\mathcal{F} = \{F_r : r \in \mathcal{R}\}$ a set of image data (or observations) on these sites. The set of all possible observations $f = (f_{r_1}, f_{r_2}, \ldots, f_{r_M})$ is denoted by Φ. Furthermore, we are given another set of sites $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$, each of these sites may take a label from $\Lambda = \{0, 1, \ldots, L-1\}$. The configuration space Ω is the set of all global discrete labeling $\omega = (\omega_{s_1}, \ldots, \omega_{s_N}), \omega_s \in \Lambda$. The two set of sites $\mathcal{R}$ and $\mathcal{S}$ are not necessarily disjunct, they may have common parts (see for example Geman's image restoration model involving a line process [25]) or refer to a common set of sites. Our goal is to model the labels and observations with a joint random field $(\mathcal{X}, \mathcal{F}) \in \Omega \times \Phi$. The field $\mathcal{X} = \{X_s\}_{s \in \mathcal{S}}$ is called the *label field* and $\mathcal{F} = \{F_r\}_{r \in \mathcal{R}}$ is called the *observation field.*

3.1 Bayesian estimation

First, we construct a Bayesian estimator for the *label field* [16]. We can define both the joint and conditional probabilities in terms of the a priori and a posteriori distributions:

$$P_{\mathcal{X},\mathcal{F}}(\omega, f) = P_{\mathcal{F}|\mathcal{X}}(f \mid \omega) P_{\mathcal{X}}(\omega) \tag{11}$$

$$P_{\mathcal{X}|\mathcal{F}}(\omega \mid f) = \frac{P_{\mathcal{X},\mathcal{F}}(\omega, f)}{P_{\mathcal{F}}(f)} = \frac{P_{\mathcal{F}|\mathcal{X}}(f \mid \omega) P_{\mathcal{X}}(\omega)}{P_{\mathcal{F}}(f)} \tag{12}$$

Since the realization of the observation field is known, $P(f)$ is constant and we can write:

$$P_{\mathcal{X}|\mathcal{F}}(\omega \mid f) \propto P_{\mathcal{F}|\mathcal{X}}(f \mid \omega) P_{\mathcal{X}}(\omega) \tag{13}$$

The estimator is the following decision function δ:

$$\delta : \Phi \longrightarrow \Omega \tag{14}$$

$$f \mapsto \delta(f) = \hat{\omega} \tag{15}$$

and the corresponding *Bayes risk* is given by

$$r(P_X, \delta) = E\left[R(\omega, \delta(f))\right] \tag{16}$$

where $R(\omega, \delta(f))$ is a cost function. Our estimator must have the minimum Bayes risk:

$$\hat{\omega} = \arg\min_{\omega' \in \Omega} \int_{\omega \in \Omega} R(\omega, \omega') P_{\mathcal{X}|\mathcal{F}}(\omega \mid f) d\omega \tag{17}$$

We mention hereafter the three best known Bayesian estimators [50].

3.1.1 Maximum A Posteriori (MAP)

The MAP estimator is the most frequently used estimator in image processing. Its cost function is defined by

$$R(\omega, \omega') = 1 - \Delta_{\omega'}(\omega), \tag{18}$$

where $\Delta_{\omega'}(\omega)$ is the Dirac mass in ω'. Clearly, this function has the same cost for all configurations different from ω'. From Eqn. (17) and Eqn. (18), the MAP estimator of the label field is given by

$$\hat{\omega}^{MAP} = \arg\max_{\omega \in \Omega} P_{\mathcal{X}|\mathcal{F}}(\omega | f). \tag{19}$$

This estimator gives for a given observation f, the modes of the posterior distribution, that is the most likely labelings given the observation f. Eqn. (19) is a combinatorial optimization problem which requires special algorithms such as Simulated Annealing [2, 63].

3.1.2 Marginal A Posteriori Modes (MPM)

We define the cost function of the MPM estimator as

$$R(\omega, \omega') = \sum_{s \in \mathcal{S}} (1 - \Delta_{\omega'_s}(\omega_s)). \tag{20}$$

Remark that the above function is related to the number of sites $s \in \mathcal{S}$ such that $\omega_s \neq \omega'_s$. The solution of Eqn. (17) is given by

$$\forall s \in \mathcal{S} : \hat{\omega}_s^{MPM} = \arg\max_{\omega_s \in \Lambda} P_{X_s|\mathcal{F}}(\omega_s \mid f), \tag{21}$$

which gives the configuration which maximizes at each site the a posteriori marginal $P_{X_s|\mathcal{F}}(. \mid f)$.

3.1.3 Mean Field (MF)

Here, we have the following cost function:

$$R(\omega, \omega') = \sum_{s \in \mathcal{S}} (\omega_s - \omega'_s)^2. \tag{22}$$

From Eqn. (17) and Eqn. (22), we have

$$\forall s \in \mathcal{S} : \hat{\omega}_s^{MF} = \int_{\omega \in \Omega} \omega_s P_{\mathcal{X}|\mathcal{F}}(\omega \mid f) d\omega, \tag{23}$$

which is nothing else but the conditional expected value of $\mathcal{X}$ given $\mathcal{F} = f$ that is the *mean field* of $\mathcal{X}$.

3.2 Defining a Priori and a Posteriori Distributions

Within the Bayesian framework, our knowledge about the "world" is represented by a priori probabilities. However, in practice, it is extremely difficult to define such probabilities globally, even if we focus on a specific area of image processing. But there are some well defined properties if we are considering images *locally*. Usually, neighboring pixels have similar intensities, edges are smooth and often straight and textures have also well defined local properties. It is then a better idea to represent our knowledge in terms of some local random variables. This kind of knowledge is well described by means of MRF.

3.2.1 Prior Distribution

Let us suppose that $\mathcal{X}$ is a MRF with some neighborhood system $\mathcal{V}' = \{\mathcal{V}'_s : s \in \mathcal{S}\}$ and distribution

$$P(\mathcal{X} = \omega) = \frac{1}{Z} \exp\left(-U'(\omega)\right), \tag{24}$$

$$U'(\omega) = \sum_{C \in \mathcal{C}'} V'_C(\omega) \tag{25}$$

where $U'(\omega)$ is the energy function. The above equations give another good reason using MRF priors, namely their Gibbs representation through *clique-potentials*, which are more convenient than working directly with probabilities.

3.2.2 Degraded Image Model and Posterior Distribution

The observations are related to the label process through a *degradation model* which models the relation between the *label field* $\mathcal{X}$ and the *observation process* $\mathcal{F}$. In image restoration/deconvolution for example, what we observe is a blurred noisy image and we want to restore the original one. So, the label process represents grey-levels in this case. We are now considering a similar model but in a more general manner. Most of the problems result in the following function:

$$\mathcal{F} = \Psi(H(\mathcal{X}), N), \tag{26}$$

or at the pixel level:

$$\forall r \in \mathcal{R} : F_r = \Psi(H_r(X_{\psi(r)}), N_r) \tag{27}$$

where $\Psi(a, b)$ is an invertible function in a. H_r is a local function defined on a small part $\psi(r)$ of $\mathcal{S}$ such that $\psi(r) \in \mathcal{S}, \mid \psi(r) \mid \ll \mid \mathcal{S} \mid$ and $\psi^{-1}(s) = \{r \in \mathcal{R} \mid s \in \psi(r)\}$. N is a random component (usually a Gaussian white noise but in tomography N_r are Poisson variables whose means are related to $\mathcal{X}$). In [25], for instance, H is a *blurring matrix* and N is an additive Gaussian noise. If we assume that the distribution of N is given by

$$P_N(.) = \prod_{r \in \mathcal{R}} P_{N_r}(.) \tag{28}$$

then we obtain

$$P_{\mathcal{F}|\mathcal{X}}(f \mid \omega) = \prod_{r \in \mathcal{R}} P_{N_r}(\Psi^{-1}(H_r(\omega_{\psi(r)}), f_r)). \tag{29}$$

The conditional distribution of the observation field $\mathcal{F}$ given $\mathcal{X}$ can be written as

$$P_{\mathcal{F}|\mathcal{X}}(f \mid \omega) = \exp\left(\sum_{r \in \mathcal{R}} -\ln(P_{N_r}(\Psi^{-1}(H_r(\omega_{\psi(r)}), f_r)))\right), \tag{30}$$

assuming that $P_{N_r}(.) > 0$ at each site r in $\mathcal{R}$. Combining the above equation with Eqn (13) and Eqn. (24), the posterior distribution is of the following form:

$$P_{\mathcal{X}|\mathcal{F}}(\omega \mid f) \propto \frac{1}{Z} \exp\left(\sum_{r \in \mathcal{R}} -\ln(P_{N_r}(\Psi^{-1}(H_r(\omega_{\psi(r)}), f_r))) + \sum_{C \in \mathcal{C}'} V'_C(\omega)\right) \tag{31}$$

Notice that the posterior distribution is also a Gibbs distribution with the smallest neighborhood system $\mathcal{V}$ containing all the cliques in $\mathcal{C}'$ and the sets $\{\psi(r), r \in \mathcal{R}\}$:

$$\forall s \in \mathcal{S} : \mathcal{V}_s = \left(\bigcup_{r \in \psi^{-1}(s)} \psi(r) \setminus \{s\}\right) \cup \mathcal{V}'_s \tag{32}$$

Let us denote the corresponding energy function by $U(\omega, f)$:

$$\begin{aligned} U(\omega, f) &= \sum_{r \in \mathcal{R}} -\ln(P_{N_r}(\Psi^{-1}(H_r(\omega_{\psi(r)}), f_r))) + \sum_{C \in \mathcal{C}'} V'_C(\omega) \\ &= \sum_{r \in \mathcal{R}} V_r(\omega_{\psi(r)}), f_r) + \sum_{C \in \mathcal{C}'} V'_C(\omega) \end{aligned} \tag{33}$$

In the following, we will be more specific about $V_r(\omega_{\psi(r)}), f_r)$ and suppose that it is of the form:

$$V_r(\omega_{\psi(r)}), f_r) = V_r(\omega_{\psi(r)}) + \sum_{s \in \psi(r)} V_{s,r}(\omega_s, f_r). \tag{34}$$

This restriction is less severe than it might be expected. As we will see, most of the nowadays used models have this kind of energy function. The above equation can be rewritten as

$$\sum_{r \in \mathcal{R}} V_r(\omega_{\psi(r)}), f_r) = \sum_{r \in \mathcal{R}} V_r(\omega_{\psi(r)}) + \sum_{r \in \mathcal{R}} \sum_{s \in \psi(r)} V_{s,r}(\omega_s, f_r) \tag{35}$$

$$= \sum_{r \in \mathcal{R}} V_r(\omega_{\psi(r)}) + \sum_{s \in \mathcal{S}} \underbrace{\sum_{r \in \psi^{-1}(r)} V_{s,r}(\omega_s, f_r)}_{V_s(\omega_s, f_{\psi^{-1}(s)})} \tag{36}$$

$$\tag{37}$$

Finally, we have the following energy function associated with the posterior distribution of the label field $\mathcal{X}$:

$$U(\omega, f) = \sum_{s \in \mathcal{S}} V_s(\omega_s, f_{\psi^{-1}(s)}) + \sum_{C \in \mathcal{C}} V_C(\omega) \tag{38}$$

$$= U_1(\omega_s, f_{\psi^{-1}(s)}) + U_2(\omega). \tag{39}$$

where the clique-potentials $V_C(\omega)$ are defined as

$$V_C(\omega) = \begin{cases} V'_C(\omega) & \text{if } C \in \mathcal{C}' \text{ and } C \notin \{\psi(r), r \in \mathcal{R}\} \\ V_r(\omega_{\psi(r)}) & \text{if } C = \psi(r) \text{ and } \psi(r) \notin \mathcal{C}' \\ V'_C(\omega) + V_r(\omega_{\psi(r)}) & \text{if } C = \psi(r) \text{ and } \psi(r) \in \mathcal{C}' \end{cases} \tag{40}$$

If we assume that the observed image $\mathcal{F}$ is affected at site s only by the pixel s itself then Eqn. (38) can be further simplified: $\psi(r)$ reduces to s and the neighborhood system of the posterior distribution is equivalent to the one of the prior distribution.

3.3 Some Examples of Markov Models

Herein, we present some models applied to a variety of image processing tasks. Most of them uses the general model discussed in the previous section. Let us begin this discussion with the restoration/deconvolution model proposed by D. Geman and S. Geman in [25].

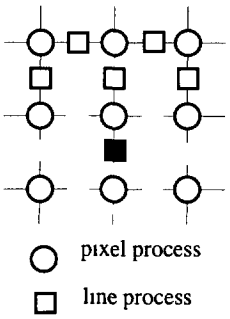

Fig. 6. Geman's image restoration model

3.3.1 Image Restoration/Deconvolution

We observe a blurred noisy image $\mathcal{F}$ and we want to restore the original one. The components of the degraded model in Eqn. (26) have the following meanings: N is supposed to be a white Gaussian noise with mean μ and variance σ^2. H is a shift-invariant blurring matrix. First of all, we define the lattices on which the label process and the observation process are defined. The *observation process* simply consists of the grey-level at each pixel of the given image. Thus $\mathcal{R}$ is a lattice, each site corresponds to a pixel. The *label process* is more sophisticated involving both *pixel sites* and *line sites*. $\mathcal{X}$ is then a "mixed" process having two subprocesses: a *pixel process* and a *line process*. The lattice $\mathcal{S}$ contains $\mathcal{R}$ (*pixel sites*) and another lattice with sites siting between each vertical and horizontal pair of pixels representing a possible location of edge elements (see Fig. 6).

We now turn to the posterior distribution and its energy function. Let us denote the *line process* by $\mathcal{X}^l$ and the *pixel process* by $\mathcal{X}^p$. We assume that $\mathcal{X}^p$ is a MRF over an homogeneous neighborhood system (see Section 2.3) $\mathcal{V}$ on $\mathcal{R}$ and $\mathcal{X}^l$ is also a MRF over a neighborhood system shown on Fig. 6. (the neighbors of the black site are the grey sites). $\mathcal{X}$ has a prior distribution of

$$P(\mathcal{X}^p = \omega^p, \mathcal{X}^l = \omega^l)$$

$$= \frac{1}{Z}\exp(-U'(\omega^p, \omega^l)) = \frac{1}{Z}\exp\left(-\sum_{C\in\mathcal{C}} V_C(\omega)\right), \tag{41}$$

where $\omega = (\omega^p, \omega^l)$. ω^p takes values among the available grey-levels and ω^l among the line states. If we choose $\mathcal{V}$ such that it is large enough to encompass the dependencies caused by the blurring H then the posterior distribution also defines a MRF with energy function

$$U(\omega^p, \omega^l) = U'(\omega^p, \omega^l) + \frac{\|\vec{\mu} - \Psi^{-1}(H(\omega^p), f)\|^2}{2\sigma^2} \tag{42}$$

The optimal labeling $\hat{\omega}$ is found by the MAP estimate minimizing the above energy function. The restored image is then given by the *pixel process* $\hat{\omega}^p$.

3 3.2 Texture Segmentation

The observations consist of a set of various texture features describing spatial statistics of the image. These features are computed on local windows around each pixel including mean, variance, correlation, entropy, contrast, homogeneity, etc [17–19, 27, 32] ... Here, both the *observation process* and the *label process* are defined on the same lattice $\mathcal{S}$ with sites corresponding to image pixels. The terms U_1 and U_2 from Eqn. (39) are defined in the following way: The prior energy U_2 usually favors spatially homogeneous regions assigning lower potentials to homogeneous cliques. The term U_1 does not have such a "standard" definition. It has various forms in the literature. In [36], it measures the distance, at a given point s, between the distribution of the texture features in a small block B_s centered at s and the one in the whole (candidate) region R_s to which we want to assign s. This technique, as claimed in [36], permits to automatically determine the number of regions. The energy function is defined as

$$U_1(\omega, \vec{f}) = \sum_{s \in \mathcal{S}} V_s(B_s, R_s) \tag{43}$$

$$V_s(B_s, R_s) = \sum_{i=1}^{m} (2\Delta(d(\vec{f}^{\,i}_{R_s}, \vec{f}^{\,i}_{B_s}) > c^i) - 1) \tag{44}$$

where m is the number of considered features. $\vec{f}_{B_s}$ and $\vec{f}_{R_s}$ denote the set of feature vectors on block B_s and on the region R_s respectively. $d(a, b)$ stands for the *Kolmogorov-Smirnov distance* and c^i is a threshold given by statistical tables associated to the Kolmogorov limit distribution. The function Δ returns 1 if its argument is true, 0 otherwise.

3.3.3 Edge Detection

MRF models for edge detection are often *compound Gauss-Markov random fields* (CGMRF) [39, 66]. The local characteristics of a CGMRF is given by

$$P(f_s \mid f_r, r \in \mathcal{S}) = \frac{1}{\sqrt{2\pi}\varsigma} \exp\left(\frac{1}{2\varsigma^2}\left(f_s - \mu_m - \sum_{r \in \mathcal{V}_I} \theta_r (f_r - \mu_m)\right)^2\right), \tag{45}$$

where ς is the deviation, μ_m is the mean and θ_r is the model parameter. The supporting graph is similar to the one reported in [25] (cf. Fig. 6). The observations $\mathcal{F}$ are considered to be corrupted by an additive Gaussian noise with zero mean and variance σ^2. The label field is again a "mixed" process containing both a *pixel process* $\mathcal{X}^p$ and a *line process*. Assuming a first order neighborhood system (cf. Fig. 4) and denoting the horizontal and vertical line process by $\mathcal{X}^h$ and $\mathcal{X}^v$ respectively, a possible form of the energy function is given by [66]:

$$\begin{aligned} U(\omega, f) = & \frac{1}{2\sigma^2} \sum_{s=(i,j) \in \mathcal{S}} ((f_{i,j} - \omega^p_{i,j})^2 + \beta^2 (1 - 2(\theta_h + \theta_v)) f^2_{i,j} \\ & + \theta_h (\beta^2 (f_{i,j} - f_{i-1,j})^2 (1 - \omega^h_{i,j}) + \alpha \omega^h_{i,j}) \\ & + \theta_v (\beta^2 (f_{i,j} - f_{i,j+1})^2 (1 - \omega^v_{i,j}) + \alpha \omega^v_{i,j})) \end{aligned} \tag{46}$$

with $1 - 2(\theta_h + \theta_v) > 0$. θ_v and θ_h are the model parameters for the vertical and horizontal cliques. β^2 corresponds to a regularization term reflecting the confidence in the data. In [66], $\beta^2 = \sigma^2/\varsigma^2$ expressing that when $\mathcal{F}$ is very noisy, we have no confidence in the data (β^2 is high). This model is related to the weak membrane model presented in [7]. The estimation of the *line process* is done by a Mean Field approach.

3.3.4 Motion Analysis

In [56] a MRF model for motion detection is presented. The *observation process* is defined both on the image-lattice $\mathcal{S}$ and on a time axis t. The detection of moving objects relies on the analysis of the variation of the intensity distribution in time. At each pixel, we have a two-element observation vector:

$$\vec{f}_s^1(t) = \mid y_s(t) - y_s(t - dt) \mid, \tag{47}$$

where $y_s(t)$ stands for the intensity value at pixel s at time t. $\vec{f}^2$ is a logical map of temporal changes between time t and $t - dt$. It equals to 1 if a temporal change of the intensity is valid at site s and 0 otherwise. The *label process* is binary valued ($X_s(t) = 1$ if s is on a mask of a mobile object at time t). The energy function U consists of three terms. Two of them are related to the observations and labels simultaneously (taking the role of U_1 in Eqn. (39)). One of them is used to reconstruct the mask of a mobile object at a given time:

$$U_1^2(\omega, \vec{f}_s^2) = \sum_{s \in \mathcal{S}} V_1(\omega_s(t), \vec{f}_s^2(t), \vec{f}_s^2(t + dt)), \tag{48}$$

the other one expresses consistency between the current labeling and the intensity variation:

$$U_1^1(\omega, \vec{f}_s^1) = \sum_{s \in \mathcal{S}} \left(\frac{1}{2\sigma^2} (\vec{f}_s^1(t) - \mu\omega_s(t))^2 + (\vec{f}_s^1(t + dt) - \mu\omega_s)^2 \right) \tag{49}$$

where μ and σ are model parameters. The third term of the energy function U corresponds to U_2 with potentials favoring homogeneous masks.

4 AN APPLICATION TO REMOTE SENSING AND ASTROPHYSICS: IMAGE DECONVOLUTION

Remote sensing and planetary images can be modeled by MRF, which enables us to use the stochastic framework defined herein to construct an automatic image deconvolution method. First, we describe the prior model, then we show how to compute the MAP to get the image estimate, and finally we give the outline of an unsupervised method to estimate the parameters of the model and the image at the same time.

Let us recall the expression of the prior probability associated to the MRF model:

$$P(\mathcal{X} = \omega) = \frac{1}{Z} \exp(-U_2(\omega)) = \frac{1}{Z} \exp\left(- \sum_{C \in \mathcal{C}} V_C(\omega) \right) \tag{50}$$

We choose a MRF with a first order neighborhood to model the image which has to be deconvolved, but this model can be easily generalized to handle longer interactions. We define the clique potentials by:

$$V_C(\omega_s, \omega_t) = \lambda^2 \, \varphi\left(\frac{\omega_s - \omega_t}{\delta} \right) \tag{51}$$

where the function φ is positive, increasing on $\mathbb{R}^+$ and symmetric, as introduced in [13] and [28]. Properties of the φ-function have been studied in order to preserve the edges, avoiding noise amplification. If we use a convex function, ensuring the uniqueness of the solution, then restoration can be made by a deterministic minimization algorithm. Thus, it is possible to

preserve the edges, since the potential function is not quadratic. It means that the high pixel differences (corresponding to the edges) are not penalized as much as if the potential was quadratic, while low pixel differences, corresponding to noise, are penalized following a quadratic function. Therefore the noise is efficiently filtered on homogeneous areas, while preserving the edges.

Then the energy of the prior probability distribution can be expressed as:

$$U_2(\omega) = \lambda^2 \sum_{s=(i,j)\in\mathcal{S}} \left[\varphi\left(\frac{\omega_{i+1,j} - \omega_{i,j}}{\delta}\right) + \varphi\left(\frac{\omega_{i,j+1} - \omega_{i,j}}{\delta}\right)\right] \tag{52}$$

The observation model is a convolution by the kernel h corrupted by a white Gaussian noise of variance σ^2. It corresponds to the following energy:

$$U_1(\omega,\ f) = \sum_{s\in\mathcal{S}} \frac{(f - h\star\omega)_s^2}{2\sigma^2} \tag{53}$$

The energy associated to the posterior density is $U = U_1 + U_2$. Computing the MAP to deconvolve the image is equivalent to minimize the energy U. Since U is convex in this case, we use a deterministic method to perform the minimization, which is faster than stochastic methods.

The idea is to *diagonalize* the convolution operator, which is equivalent to replace the convolution by a point to point multiplication, in the *frequency space*. The difficulty is to diagonalize at the same time the non quadratic regularization term. This is achieved by taking a half-quadratic expansion of φ, in order to linearize the problem [29]:

$$\varphi(u) = \inf_{b\in\mathbb{R}} \left[(u-b)^2 + \psi(b)\right] \tag{54}$$

Two auxiliary MRF b^x and b^y are defined, so that:

$$\begin{aligned} U_2^*(\omega) \quad = \lambda^2 \sum_{s=(i,j)\in\mathcal{S}} \quad & \left[\left(b^x_{i,j} - \frac{\omega_{i+1,j} - \omega_{i,j}}{\delta}\right)^2 + \psi(b^x_{i,j})\right. \\ & \left.\left(b^y_{i,j} - \frac{\omega_{i,j+1} - \omega_{i,j}}{\delta}\right)^2 + \psi(b^y_{i,j})\right] \end{aligned} \tag{55}$$

Instead of minimizing w.r.t. ω only, we minimize alternately w.r.t. ω, with b^x and b^y fixed, and w.r.t. b^x and b^y with ω fixed. The first step is easy to achieve, since each auxiliary variable is updated independently according to $b^k_s = u - \varphi'(u)/2$ where k is either x or y, and u is the respective pixel difference divided by δ. The second step is done in the Fourier domain, since the Fourier transform diagonalizes the quadratic form U_2^* when b^x and b^y are fixed. We use in fact a Cosine transform instead of a Fourier transform, which enables us to fulfill the symmetric boundary conditions, which avoid artefacts on the borders of the image.

The half-quadratic expansion enables us to define an implicit "mixed" MRF, with a line process and a pixel process related to b^x and b^y and ω, which are linked through the φ function.

The quality of the restored image is highly sensitive to the value of the parameters λ and δ. Therefore, they must be accurately determined. The first parameter controls the smoothness of the solution to be restored, while the second one is a threshold, which controls the noise cancellation and the edge preservation (in fact it enables us to choose gradually between a quadratic model and a total variation model).

To estimate these parameters, we have developed a method called MCMCML, for Markov Chain Monte Carlo Maximum Likelihood [38]. The optimal parameter values are determined by maximizing the incomplete data likelihood

$$(\hat{\lambda}, \hat{\delta}) = \arg\max_{\lambda,\delta} P_{\mathcal{F}}(f \mid \lambda, \delta) \tag{56}$$

Using Bayes rule, the likelihood can be expressed as:

$$P_{\mathcal{F}}(f \mid \lambda, \delta) = \frac{Z_{\mathcal{X}|\mathcal{F}}}{Z_{\mathcal{F}|\mathcal{X}}.Z_{\mathcal{X}}} \tag{57}$$

where $Z_{\mathcal{X}|\mathcal{F}}$, $Z_{\mathcal{F}|\mathcal{X}}$ and $Z_{\mathcal{X}}$ are respectively the normalizing constants corresponding to the posterior, the likelihood of ω and the prior distribution. The main difficulty of parameter estimation comes from these functions, which depend on (λ, δ) and f but are impossible to evaluate in practice. So we optimize this criterion without explicitly computing them, as we only compute the derivatives, by using a gradient descent algorithm.

To optimize the likelihood we minimize the -log-likelihood, and we need its derivatives. To estimate them we use the following property: let Z denote the normalizing constant related to a distribution $P_\theta(\omega) = Z^{-1}e^{-\Phi(X,\theta)}$, where θ is a vector parameter, then we have :

$$\frac{\partial \log Z}{\partial \theta^i} = -E_{\omega \sim P_\theta(\omega)}\left[\frac{\partial \Phi(X,\theta)}{\partial \theta^i}\right] \tag{58}$$

where $E[\,]$ is the expectation of ω w.r.t. $P_\theta(\omega)$. This expectation can be estimated using a Monte Carlo method [31, 58], by sampling from the distribution $P_\theta(\omega)$.

We need to sample from both prior and posterior densities. Sampling from the posterior density is intractable by means of classical algorithms such as Gibbs sampler [25] or Metropolis dynamics [51] due to the large support of the PSF, inducing a large neighborhood for the conditional probability. We use the idea introduced by Geman & Yang [29] to derive a simulated annealing algorithm for MAP estimation of X. The idea is to use the same half-quadratic expansion of φ as before, to introduce two auxiliary MRF b^x and b^y defined the same way. This defines an augmented stochastic process. Instead of sampling only ω from a distribution $P(\omega)$, we sample (ω, b^x, b^y) from a joint density $P(\omega, b^x, b^y)$ and keep only the ω samples. The sampling is done alternately: ω is fixed and the auxiliary fields are sampled, which is not difficult because the components of these fields are independent, then the auxiliary fields are fixed and ω is sampled. This step is made possible by using the same trick as before, i.e. by diagonalizing. There is a multivariate Gaussian distribution to be sampled, whose covariance matrix is diagonalized by a Fourier transform [29] (we use a Cosine transform instead [38]). Then the sampling is performed in the frequency space, by sampling each Fourier coefficient independently.

Finally, the MCMCML algorithm enables us to estimate the parameters of the unknown image to be restored, directly from the blurred and noisy observation. We prefer to fix the threshold δ, because the joint ML estimator defined here is degenerate, since it produces multiple solutions. Then, we estimate the value of the regularizing parameter λ and use it to deconvolve the image with a deterministic algorithm using alternate minimizations. The estimation and the deconvolution are simultaneous, because the posterior sampling algorithm is initialized with the MAP estimate. At each gradient descent step, the parameter value is updated and the MAP is computed, then prior and posterior sampling are performed, to estimate the gradient and go to the next step. This defines a convergent method, which needs maximum 10 iterations to produce a deblurred image with optimal parameter values [38].

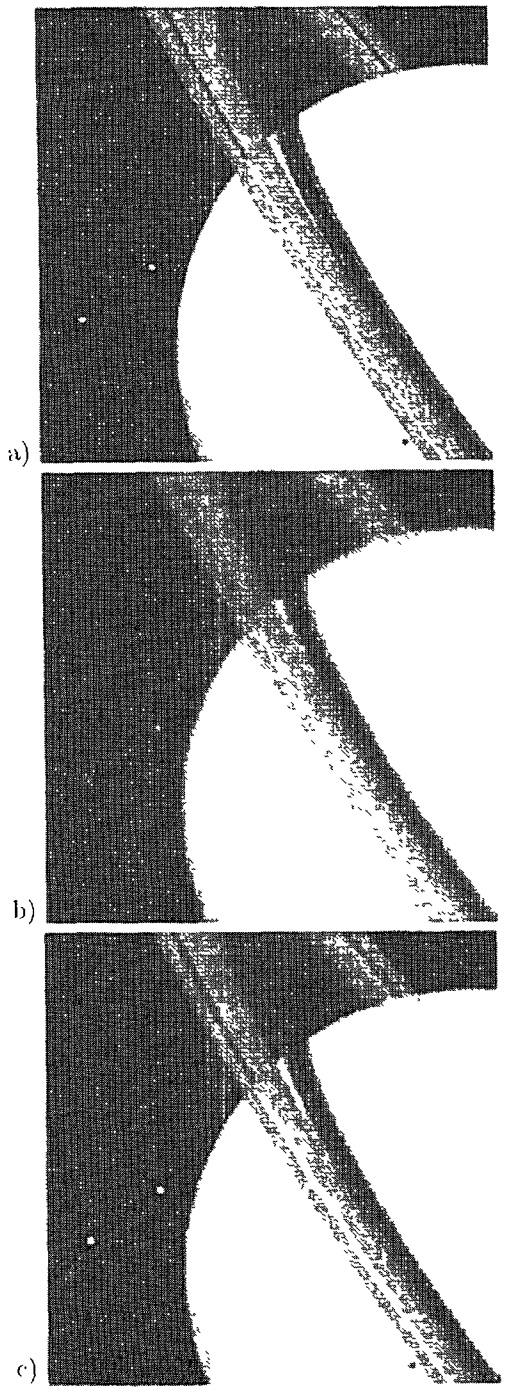

Fig. 7. a) original image of Saturn, b) blurred and noisy image (Gaussian blur with standard deviation=2) and noise variance σ^2=2, c) deconvolved image with δ=5 and estimated λ=0.63, $\varphi(t) = \log(1 + t^2)$

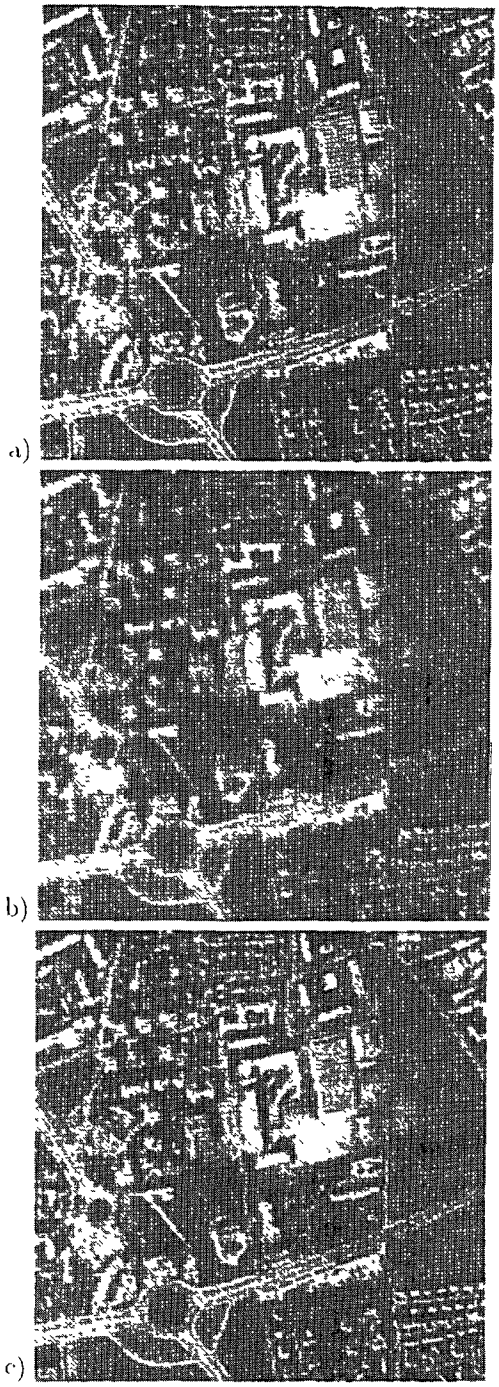

Fig. 8. a) original image of Nîmes (SPOT 5) © CNES, b) blurred and noisy image (blurring kernel provided by the CNES) and variance σ^2=2, c) deconvolved image with δ=10 and estimated λ=0.5, $\varphi(t) = 2\sqrt{1+t^2} - 2$

The results are shown on Fig. 7 and Fig. 8 for astrophysics and remote sensing data.

They exhibit both sharp edges and clean homogeneous areas. For an image made of large constant areas, such as Saturn, the model enables us to obtain good results. They can still be improved by choosing a non-convex potential function, but the optimization becomes more difficult: a stochastic technique is needed to avoid falling in a local maximum of the posterior density. It is possible to use the mean field estimate to compute the unknown image, the mean is estimated using the posterior sampler.

For aerial and satellite images, the textures are not always preserved, because they are not taken into account in the prior model, but the results are globally correct, since the small features are preserved.

The framework developed herein can be used to define other models, with higher order interactions (to take into account the textures, for instance), and using potential functions allowing half-quadratic expansions lead to the same type of MCMC algorithm for simultaneous parameter and image estimation.

5 CONCLUSION

As we have seen above, MRF models are widely used in image processing, mainly due to the fact that they enable to express global constraints or hypothesis in a local way. This is true, in particular, for remote sensing applications such as road network detection, urban area extraction, image classification for precision farming, super-resolution, digital elevation map etc. using satellite images and for astrophysics applications such as planetary imaging.

Nevertheless, for images taken by very high resolution satellites such as Ikonos, IRS, Helios or the future Pléïades for instance, classical MRF models are not well adapted to incorporate geometrical constraints on the shape of objects. New models based on marked point processes, which can exhibit the Markov property, are currently under investigation in Ariana research group [22, 60, 61].

References

[1] R Azencott Image Analysis and Markov Fields. In *Proc ICIAM'87*, pages 53–61, Paris, France, June 29-July 3 1987.

[2] R Azencott. Parallel Simulated Annealing: An Overview of Basic Techniques In R. Azencott, editor, *Simulated Annealing· Parallelization Techniques*, pages 37–46. John Wiley & Sons, Inc , 1992

[3] R Azencott and C Graffigne. Non supervised segmentation using multi-level Markov Random Fields In *Int Conf. Pattern Rec.*, III, pages 201–204, The Hague, The Netherlands, Sep. 1992.

[4] R J Baxter. *Exactly Solved Models in Statistical Mechanics*. Academic Press, 1990.

[5] J Besag. On the statistical analysis of dirty pictures. *Jl. Roy. Statis. Soc. B.*, 1986.

[6] J. Besag Spatial interaction and the statistical analysis of lattice systems (with discussion). *Jl. Roy. Statis Soc. B. 36* , pages 192–236, 1974

[7] A. Blake and A Zisserman. Visual reconstruction. *MIT Press, Cambridge - MA*, 1987

[8] C Bouman and B. Liu Multiple Resolution segmentation of Texture Images. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 13·99–113, 1991.

[9] C Bouman and M Shapiro. Multispectral Image Segmentation using a Multiscale Model In *Proc. ICASSP'92*, San Francisco, USA, March 1992.

[10] B. Braathen, W. Pieczynski, and P. Masson Global and Local Methods of Unsupervised Bayesian Segmentation of Images. *Machine Graphics and Vision*, 2(1).39–52, 1993.

[11] V. Černy Thermodynamical Approach to the Traveling Salesman Problem An Efficient Simulation Algorithm. *J. Opt. Theory Appl* , 45(1):41–51, January 1985

[12] B Chalmond Image restoration using an estimated markov model. *Signal Processing*, 15·115–129, 1988.

[13] P Charbonnier, L Blanc-Féraud, G Aubert et M Barlaud Deterministic Edge-Preserving Regularization in Computed Imaging. *IEEE Trans on Image Processing*, 6(2):298–311, Feb. 1997.

[14] A Chardin and P. Pérez. Unsupervised Image Classification with a Hierarchical EM Algorithm. In *Proc. ICCV*, 2, pages 969–974, Kerkyra, Greece, Sep. 1999.

[15] R. Chellappa and A Jain Ed. Markov Random Fields · Theory and Application. *Academic Press*, 1993.

[16] P Chou and C. Brown. The theory and practice of bayesian image labeling *International Journal of Computer Vision*, 4 185–210, 1990.

[17] F. S. Cohen and D. B Cooper. Simple Parallel Hierarchical and Relaxation Algorithms for Segmenting Noncausal Markov Random Fields *IEEE Trans on Pattern Analysis and Machine Intelligence*, 9(2)·195–219, March 1987.

[18] G. R Cross and A. K Jain Markov Random Field Texture Models *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 5(1):25–39, January 1983.

[19] H. Derin and H. Elliott. Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 9(1)·39–55, January 1987.

[20] P L. Dobruschin The Description of a Random Field by Means of Conditional Probabilities and Constructions of its Regularity *Theory of Probability and its Applications*, XIII(2)·197–224, 1968.

[21] M. Figueiredo, J Leitao, and A.K. Jain Elements of Bayesian Theory and Markov Fields for Image Analysis *Springer-Verlag*, 2002.

[22] L Garcin, X Descombes, H. Le Men and J. Zerubia Building detection by Markov object processes. *Invited paper, Proc. ICIP, Thessaloniki, Greece*, Oct. 2001.

[23] D. Geiger and F Girosi. Parallel and deterministic algorithms for MRFs surface reconstruction *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1315, pages 401–412, 1991

[24] S. B Gelfand and S. K Mitter. On Sampling Methods and Annealing Algorithms In R. Chellappa, editor, *Markov Random Fields*, pages 499–515. Academic Press, Inc., 1993

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

[26] D Geman Random fields and inverses problems in imaging. *Ecole d'été de probabilités de St Flour XVIII, Lecture Notes in Mathematics, Springer*, 1427, pages 117–193, 1988.

[27] S. Geman, D. Geman, C. Graffigne, and P. Dong Boundary detection by constrained optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12 609–628, 1990

[28] D Geman et G Reynolds Constrained restoration and the recovery of discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(3):367–383, March 1992.

[29] D Geman et C. Yang. Nonlinear Image Recovery with Half-Quadratic Regularization *IEEE Trans. on Image Processing*, 4(7) 932–946, July 1995

[30] B. Gidas A Renormalization Group Approach to Image Processing Problems. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 11(2).164–180, February 1989

[31] W R. Gilks, S Richardson and D.J Spiegelhalter. Markov Chain Monte Carlo in practice *Chapman & Hall, London*, 1996

[32] G. Gimel'Farb. Image textures and Gibbs random fields *Kluwer*, 1999

[33] C. Graffigne, F Heitz, P Pérez, F Prêteux, M Sigelle and J Zerubia. Hierarchical Markov random field models applied to image analysis. a review. In *Proc. SPIE Conf., 2568 on Neural, Morphological and Stochastic Methods in Image and Signal Processing*, San Diego, USA, July 10-11, 1995.

[34] B Hajek A Tutorial Survey of Theory and Applications of Simulated Annealing. In *Proc. 24. Conf. on Decision and Control*, pages 755–760, Lauderdale, FL, December 1985.

[35] F. R. Hansen and H Elliott Image Segmentation Using Simple Markov Field Models. *CVGIP*, 20.101–132, 1982.

[36] F. Heitz and C. Kervrann. A statistical model-based approach to unsupervised texture segmentation. In *Proc. SCIA, Norway*, 1993.

[37] F Heitz, P. Pérez, and P. Bouthemy. Multiscale Minimization of Global Energy Functions in Some Visual Recovery Problems *CVGIP·IU*, 59(1)·125–134, 1994

[38] A. Jalobeanu, L. Blanc-Féraud and J Zerubia Hyperparameter estimation for satellite image restoration by a MCMCML method. *EMMCVPR Conf.*, York, England, Jul 26-29, LNCS 1654, Springer Verlag, 1999

[39] F. C. Jeng and J. M. Woods. Compound Gauss - Markov Random Fields for Image Estimation. *IEEE Trans. Acoust , Speech and Signal Proc.*, ASSP-39:638–697, 1991.

[40] B. Jeon and D. A. Landgrebe. Classification with Spatio-Temporal Interpixel Class Dependency Contexts. *IEEE Trans. on Geoscience and Remote Sensing*, 30(4).663–672, July 1992

[41] J. M. Jolion and A. Rosenfeld. *A Pyramid Framework for Early Vision* Series in Engineering and Computer Science. Kluwer Academic Publishers, 1994.

[42] Z Kato, M Berthod and J. Zerubia. A Hierarchical Markov Random Field Model and Multitemperature Annealing for Parallel Image Classification. *Graphical Models and Image Processing*, 58(1), pages 18–37, Jan 1996.

[43] Z Kato, J Zerubia, and M. Berthod Unsupervised parallel image classification using Markovian models. *Pattern Recognition*, 32, pages 591–604, 1999.

[44] R Kindermann and J. L. Snell Markov Random fields and their applications *Amer. Math. Soc.*, 1:1–142, 1980.

[45] S. Kirkpatrick, C Gellatt, and M. Vecchi. Optimization by simulated annealing *Science 220, pp 671-680*, 1983

[46] S. Krishnamachari and R. Chellappa. Multiresolution Gauss-Markov Random Field Models for Texture Segmentation *IEEE Trans on Image Processing*, 6(2), Feb 1997.

[47] J M. Laferté, P Pérez and F Heitz Discrete Markov image modeling and inference on the quad-tree *IEEE Trans on Image Processing*, 9(3):390–404, Mar 2000

[48] S Lakshmanan and H. Derin. Gaussian Markov Random Fields at Multiple Resolution. In *Markov Random Fields*, pages 131–157 Academic Press, Inc , 1993

[49] S.Z Li Markov Random field modeling in computer vision *Springer Verlag*, 1995

[50] J L. Marroquin *Probabilistic solution of inverse problems* PhD thesis, MIT-Artificial Intelligence Lab , 1985

[51] N. Metropolis, A Rosenbluth, M. Rosenbluth, A. Teller, and E Teller. Equation of state calculations by fast computing machines *J of Chem. Physics, Vol. 21, pp 1087-1092*, 1953

[52] J Moussouris. Gibbs and Markov Random System with Constraints. *Journal of Statistical Physics*, 10(1):11–33, Jan. 1974

[53] G K Nicholls and M. Petrou A Generalisation of Renormalisation Group Methods for Multi Resolution Image Analysis In *Proc. ICPR'92*, pages 567–570, 1992

[54] J O'Ruanaidh and W Fitzgerald. Numerical Bayesian methods applied to signal processing *Springer Verlag*, 1996.

[55] P. Pérez Markov fields and image analysis. *CWI quaterly Journal*, 11(4), pages 413–437, 1998.

[56] P. Pérez. Markov random fields and multiresolution image analysis: application to visual motion. *PhD Thesis, University of Rennes 1, France*, 1993.

[57] P. Pérez and F. Heitz Multiscale Markov Random Fields and Constrained Relaxation in Low Level Image Analysis In *Proc ICASSP*, San-Francisco, Mar 1992

[58] C Robert and G. Casella. Monte-Carlo statistical methods. *Springer Verlag*, 1999.

[59] M Sigelle, C Bardinet, and R Ronfard Relaxation of Classification Images by a Markov Field Technique - Application to the Geographical Classification of Bretagne Region. In *Proc European Association of Remote Sensing Lab Conf*, Eger, Hungary, Sept 1992.

[60] R. Stoica, X. Descombes and J. Zerubia. Road extraction in remotely sensed images using a stochastic geometry framework. *Proc MaxEnt*, Gif sur Yvette, France, Jul. 8-13, 2000.

[61] R. Stoica, X Descombes, M N M Van Lieshout and J Zerubia An application of marked point processes to the extraction of linear networks from images *Chapter in Spatial Statistics Case Studies*, WIT Press, 2001

[62] H L Tan, S B Gelfand, and E J. Delp. A Cost Minimization Approach to Edge Detection Using Simulated Annealing *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(1):3–18, January 1991

[63] P Van Laarhoven and E Aarts Simulated annealing . Theory and applications *Reidel Pub., Dordrecht, Holland*, 1987

[64] G Winkler Image analysis, random fields and dynamic Monte-Carlo methods *Springer Verlag*, 1995

[65] F. Y. Wu. The Potts model *Reviews of Modern Physics*, 54(1)·235–268, January 1982.

[66] J. Zerubia and R. Chellappa Mean field annealing using Compound Gauss-Markov Random fields for edge detection and image estimation *IEEE Trans on Neural Networks*, 8(4):703–709, July 1993