Condensed Matter Physics 2008, Vol. 11, No 2(54), pp. 293–312

The Gibbs fields approach and related dynamics in image processing*

X.Descombes[1], E.Zhizhina [2]

[1] Ariana, Joint group, CNRS/INRIA/UNSA, INRIA, 2004, route des Lucioles, BP93, 06902, Sophia-Antipolis Cedex, France

[2] Institute for Information Transmission Problems, Bolshoy Karetny per. 19, 127994 GPS-4, Moscow, Russia

Received January 31, 2008

We give in the paper a brief overview of how the Gibbs fields and related dynamics approaches are applied in image processing. We discuss classical pixel-wise models as well as more recent spatial point process models in the framework of the Gibbs fields approach. We present a new multi-object adapted algorithm for object detection based on a spatial birth-and-death process and a discrete time approximation of this process.

Key words: *Gibbs fields, inverse problems in image processing, stochastic dynamics, object detection, marked point processes*

PACS: *05.20.Gg, 02.50.Ey, 05.10.-a, 05.90.tm*

1. Probabilistic approach in image analysis

Statistical physics and probabilistic approaches have been brought in image analysis starting with the famous paper by Besag in 1974 [5]. Ten years later, the papers, based on Gibbs modelling, either in texture analysis [6] or in image restoration [16], mark the beginning of a new field in image processing: the Markov Random Field modelling. Since then, Gibbs fields methods have been intensively developed in the Bayesian framework. These early works generated an explosion of applications of Gibbs fields methods to high-dimensional inverse problems of image processing such as restoration, denoising, deblurring, classification, segmentation, feature extraction, surface reconstruction, stereo matching, etc. Recently, there has been observed a growing interest in this field as a result of numerous applications of marked Markov processes in image analysis.

The basic idea in [16] was to rewrite a restoration procedure in the language of statistical physics using concepts of statistical ensembles, equilibrium and non-equilibrium dynamics. From this point of view, a digital image is considered as a configuration (random variables forming a set of random vectors) $\{X\}$ of a Gibbs field on the lattice with $P(X)$ as joint probability distribution. The implicit assumption behind the probabilistic approach in image analysis is that, for a given problem, there exists a probability distribution such that its ground state represents a regularized solution of the problem. Thus, the first crucial step in the probabilistic approach is the choice of the distribution $P(X)$, or equivalently in the case of Gibbs random field approach, the choice of the energy function $H(X)$. The probability distribution should contain flexible information on relevant image attributes and constraints such as regularity of contours or absence of noise. As there is no real general theory for selecting a model, the choice of a proper distribution $P(X)$ is generally based on the intuition of the desirable properties, see e.g. [15,17,23,35,43].

The basic characteristic of the distributions is their decomposition as a product of factors depending only on a few variables (the so-called, a local interaction property). Moreover, distributions usually involve only a few types of factors. One of them arises from the observable image (the data term) and has the form of an external field term. Other factors are due to generic or prior

*The work was partially supported by ECONET project 10203YK, Elena Zhizhina gratefully acknowledges the financial support of CNRS, RFBR (Grant 05–01–00449) and CRDF (grant RUM1–2693–MO–05).

knowledge on the structure of images. Prior terms in the distribution function are specified by potentials associated with local interactions defined on finite sets of neighboring variables. Thus, each variable directly depends only on its neighborhood, although from a global point of view, all variables are mutually dependent through the combination of successive local interactions.

Once the model specification is achieved, most commonly by using the Bayesian framework, the problem of maximizing the distribution $P(X)$ arises. The very high dimensionality of images (number of pixels), as well as the non-convexity of the energy $H(X)$, usually excludes any direct and deterministic method for the maximization problem. At the same time, the factorization of $P(X)$ permits to use stochastic iterative algorithms involving local changes at each step, i.e. when only one variable (or a few) can be modified at each step, all other ones being fixed. In this scheme, the resulting image is constructed as the limit configuration of a stochastic iterative procedure. At each iteration, the new configuration is obtained according to a transition distribution which depends on the current configuration. Using the local interaction property, computations of the transition probabilities become also local, i.e. they involve only a finite set of neighboring variables. In this connection, a choice of stochastic dynamics which is maximum adapted to a specific problem under consideration is a crucial step in the construction of the algorithm.

Let us stress that the probabilistic formulation offers a variety of necessary tools to analyze the problems. Statistical tools permit parameter learning, generating typical configurations and inferring unknown variables in different ways including minimization problem, capturing and combining all sorts of priors within the Bayesian machinery.

The goal of this paper is to present a brief overview of how the Gibbs field approach is applied in image processing. In the paper we discuss two aspects of modelling: constructions of the energy function capturing the key structural information concerning the images of interest, as well as the choice of a proper stochastic dynamics for the iterative procedure. Starting with a short description of classical pixel-wise (lattice based) models for restoration, segmentation and texture analysis problems, we will address more recent works on spatial point processes in the framework of the Gibbs fields approach used for object detection problems.

2. Lattice based models: the Bayesian paradigm

In this section we give a general overview of lattice based models which represent the classical approach when using Gibbs fields in image analysis. We consider a lattice spin system (an image lattice) in a finite volume $S \subset Z^2, |S| = m$, with a spin space Λ. Then any random variable $X_i,\ i \in S$ at any site of S takes the values in the spin space Λ, $\Omega = \Lambda^m$ is called the configuration space, and any configuration

$$X \in \Omega = \Lambda^m, \quad X = \{X_i,\ i \in S\}, \quad X_i \in \Lambda,$$

corresponds to a given image. Various types of spin spaces are used in practice. Most common examples are $\Lambda = \{0, \ldots, 255\}$ (the grey level space) for image restoration or texture analysis, or $\Lambda = \{1, \ldots, M\}$ for semantic labelling or image segmentation involving M classes.

In practice, an observable image (an observation) is obtained as a configuration $Y \in \Omega$ which usually has a form different from the original image $X \in \Omega$. We will call the configuration X the true or the ideal configuration which is a representation of the underlying scene X, seen through a sensor. X is the unknown which can be interpreted as a version of Y, cleaned from artifacts (noise, blurring, etc.) due to the acquisition process. For some applications, X is a first interpretation of the data Y (segmented image or description of the objects composing the scene).

The goal is to search the solution X among a set of images compatible with the observed data Y. A proper estimate should fit the data and it should fulfill some criteria reflecting the prior knowledge we have on the solution, such as regularity, smoothness, etc. In other words, prior expectation and fidelity to data should be properly balanced. The Bayesian approach consists in modelling X knowing data Y. Therefore, we construct the posterior distribution $P(X|Y)$. As this distribution is hardly accessible, as well as the joint distribution $P(X, Y)$, we use the Bayes rule to set the problem in terms of a likelihood $P(Y|X)$ modelling the sensor and a prior $P(X)$ reflecting the knowledge we have on the solution. Therefore, the expected solution maximizes the posterior distribution on the configuration space, i.e.

$$\begin{aligned} \hat{X} &= \arg \max_{X \in \Omega} P(X|Y), \\ &= \arg \max_{X \in \Omega} P(Y|X)P(X). \end{aligned} \tag{1}$$

where $Y = \{y_i, i \in S\}$ denotes data (the observation). This constitutes the Maximum A Posteriori estimator (MAP). Note that if we have no information on the solution, the prior is taken as a uniform distribution. Then we obtain the Maximum Likelihood estimator (ML).

To obtain the solution of (1), we have to address an optimization problem. Due to the interaction involved in the prior definition, this problem is usually non-convex. Therefore, deterministic algorithms, such as gradient descent, do not work for optimization as they only produce a local minima. However, because of the performances in terms of computation time, they can be used, if a "good enough" initial configuration is available or can be computed. Algorithms such as dynamic programming are also rejected because of the size of the configuration space and the lack of natural ordering. In some specific cases, for example, the Ising model, the optimization problem can be rewritten as a minimal graph cut problem, for which efficient methods based on graph flow maximization can be used [7]. However, these approaches are lacking generality. In such a situation, the only alternative which provides optimal and general algorithms is to consider MCMC simulations embedded in a simulated annealing scheme. The Metropolis-Hasting dynamics or the Gibbs sampler are the most popular MCMC schemes in this context.

Although some alternatives to the MAP exist, such as the Maximum Posterior Mode estimator (MPM), or the Mean Field (MF) approximation, reducing the computation complexity, the MAP remains very popular, particularly in the case of Gibbs distributions when the posterior distribution is simply connected with the associated energy function:

$$P(X|Y) = \frac{1}{Z} \mathrm{e}^{-H(X|Y)}. \tag{2}$$

Here $H(X|Y)$ is the energy function, Z is the normalizing factor. Thus, under the Gibbs fields approach with the posterior distribution given by equation (2) in the Gibbs form, the solution of problem (1) is a configuration (or configurations) minimizing the total energy of the system:

$$E_{\min} = \arg\min H(X|Y), \quad \hat{X} \in E_{\min}\,. \tag{3}$$

The energy function $H(X|Y)$ in (2) is defined as the sum of an interaction term, derived from the prior and associated finite-range potentials $\{U_C\}$, and a data driven term:

$$H(X|Y) \;=\; \Phi_1(X) \;+\; \Phi_2(Y|X). \tag{4}$$

Embedded into a simulated scheme, the solution of problem (1) can be obtained by iteratively simulating the following distributions:

$$P_T(X|Y) = \frac{1}{Z_T} \exp\left\{ -\frac{H(X|Y)}{T} \right\}. \tag{5}$$

If the temperature parameter T decreases slowly enough during iterations, it has been proven that simulated configurations converge to a ground state (3) of the posterior, see e.g. [16,24,25,43].

For a given problem, we have to select a model (likelihood and prior), to select an optimization algorithm (i.e. to choose a dynamics or equivalently to choose proposition kernels in the MCMC scheme) and to select parameters involved in the model (i.e. to define an estimation procedure). Below there are concrete examples to illustrate these issues.

3. Stochastic annealing as a method for global optimization

The common way to simulate a Gibbs distribution as defined in equation (5) is to consider a proper stochastic dynamics which is reversible with respect to the targeted distribution (5). For example, the Metropolis-Hastings (MH) or other Glauber type dynamics are appropriate candidates for this purpose in case of discrete spin spaces.

Let us remind that the MH algorithm is associated with the following two step single spin dynamics. If we denote by p the proposal distribution on the spin space $\Lambda = \{x_1, \dots, x_k\}$ defined by the symmetric transition matrix $p_{x,y}$, then one can randomly pick up a new configuration value $\tilde{x}_i \in \Lambda$ at the site $i \in S$ following p. We denote by $\tilde{X}$ the new configuration

$$\tilde{X} \; = \; \{x_1, \dots, \tilde{x}_i, \dots x_m\},$$

which differs from the configuration X only at one site $i \in S$. Then the new configuration $\tilde{X}$ is accepted with probability

$$Q_{X,\tilde{X}} = \exp\left\{-\frac{1}{T}[H(\tilde{X}|Y) - H(X|Y)]^+\right\},$$

where $[u]^+ = u$ as $u \geqslant 0$ and $[u]^+ = 0$ as $u < 0$, and correspondingly, rejected with probability $1 - Q_{X,\tilde{X}}$. Values of the new configuration $X(n+1) = \{X_i(n+1),\ i \in S\}$ are chosen consequently over all sites of the volume S.

As $T \to \infty$, the Gibbs distribution becomes uniform over all possible realizations. In this case, a typical realization looks random. On the other hand, as $T \to 0$, the only probable realizations are the ones that minimize the energy function. Consequently, when parameter T tends to zero, the Gibbs distribution generated by (5) will be more and more concentrated in the vicinity of the ground states $E_{\min}$, i.e. configurations where $H(X|Y)$ reaches global minima. In the case of a discrete single spin space Λ, the limit distribution will be the uniform distribution on $E_{\min}$.

If we are inside the iterative scheme depending on parameter T, then the problem is how to define the right decreasing speed of parameter $T = T(n)$ during iterations in order to escape from local minima of $H(X|Y)$:

$$\lim_{n\to\infty} P(X(n) = X|\; X(0) = X_0) \; = \; \nu(X), \quad \forall X_0 \in \Omega, \tag{6}$$

where $\nu(X)$ is a probability measure concentrated on $E_{\min}$.

In [16,25], it was shown that (6) holds for the MH scheme, if

$$\lim_{n\to\infty} T(n) \cdot \log n > R,$$

where the large enough constant R depends on the energy function $H(X|Y)$. This controlled decrease of parameter T is called the annealing schedule.

The idea behind the simulated annealing comes from physics, see [26]. If cooled down slowly enough, large physical systems tend to the states of minimal energy, called ground states. These ground states are usually highly ordered crystals. The emphasis is on “slow cooling”, and in fact, annealing means controlled cooling. If, for example, melted silicate is cooled down too quickly one gets the meta-stable material known as glass. Glass is a fluid, and not a ground state but one of the states of local minima. The analogy to physics also explains why the parameter T is called temperature. It corresponds to the factor kT with absolute temperature T and the Boltzmann constant k.

4. Some examples of inverse problems in image processing

4.1. Image restoration

Image restoration problems were first real and important applications of the Gibbs fields approach in image analysis, see e.g. [15,16,43]. Consider that we have some data (an image) Y, corrupted by noise and/or a linear distortion:

$$Y = \mathcal{K}(X) \oplus \eta, \tag{7}$$

where η represents the noise and $\mathcal{K}$ is a linear operator for distortion.

The restoration problem consists in recovering X from Y. Embeding the problem into a Bayesian framework, we maximize the following posterior:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \propto P(Y|X)P(X). \tag{8}$$

Assuming, for example, that we have an additive independent Gaussian noise, in the case of image denoising ($\mathcal{K}$ is the identity), we then have:

$$P(Y|X) = \exp\left\{-\sum_{s\in S} \frac{(y_s - x_s)^2}{2\sigma^2} - \log\sqrt{2\pi\sigma^2}\right\}, \tag{9}$$

where σ^2 is the noise variance.

Figure 1. Image denoising using a Φ-model.

The prior $P(X)$ aims at regularizing the solution, i.e. smoothing the resulting image and is usually modelled by a pairwise interaction Gibbs field:

$$P(X) = \frac{1}{Z}\exp\left\{-\sum_{c=\langle s,s'\rangle} V_c(x_s, x_{s'})\right\}. \tag{10}$$

A Gaussian Gibbs field with $V_c(x_s, x_{s'}) = (x_s - x_{s'})^2$ could achieve this goal. However, a Gaussian prior field leads to blur edges. To avoid blurring, more sophisticated priors are considered to preserve edges [17,33], such as the Φ-model:

$$V_c(x_s, x_{s'}) = \frac{-\beta}{1 + \frac{(x_s - x_{s'})^2}{\delta^2}}, \tag{11}$$

where β is a parameter representing the strength of the prior and δ is the minimum grey level gap to define an edge. To summarize, the Hamiltonian to be minimized, in the case of image denoising, is written as follows:

$$H_{\mathrm{res}}(X|Y) = \sum_{c=\langle s,s'\rangle} \frac{-\beta}{1 + \frac{(x_s - x_{s'})^2}{\delta^2}} + \sum_{s\in S} \frac{(y_s - x_s)^2}{2\sigma^2}, \tag{12}$$

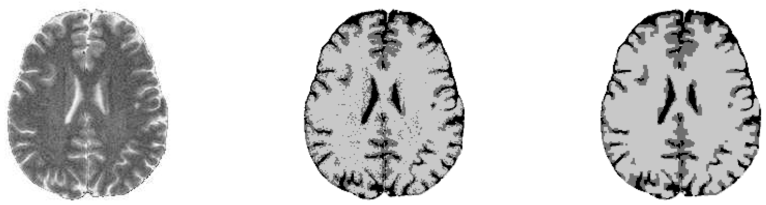
where c are sets of two neighboring pixels (sites). For each pixel we usually consider four or eight closest pixels. Another classical example is image deconvolution, when the image is blurred due to the movements of the camera or defocusing. In this case the operator $\mathcal{K}$ is a convolution by a kernel K. A denoising result obtained by the Hamiltonian defined in equation (12) and using the Langevin dynamics, see [13], is shown in figure 1 for different levels of noise.

4.2. Segmentation problem

The segmentation consists in partitioning the image in such a way that each region may represent a given object or feature of the image, see e.g. [3,9,10,29,32,41,44]. Let $S \subset \mathbb{Z}^2$ be the image lattice. A partition of S is a set of regions $\{R_i, i \in \Lambda = \{1, \ldots, I\}\}$, such that $\cup_{i\in\Lambda} R_i = S$ and $R_i \cap R_j = \emptyset$ when $i \neq j$. Consider a random field $Y = (y_s)_{s\in S}$, where $y_s \in \Lambda$. The likelihood term $P(Y|X)$ model the grey level distribution of the pixels belonging to a given class or region. For example, we may consider that each class representing a given feature (sea, sand, crops in remote sensing data or grey, white matter and CSF for brain images, as in the example given in figure 2) exhibits a Gaussian distribution and is therefore characterized by its means and variance. We then write the likelihood as follows:

$$P(Y|X) = \exp\left\{-\sum_{s\in S}\sum_{i\in\Lambda}\left(\frac{(y_s-\mu_i)^2}{2\sigma_i^2} - \log\sqrt{2\pi\sigma_i^2}\right)\delta(x_s = i)\right\}, \tag{13}$$

where $\delta(a)$ is equal to 1, a is true and 0 otherwise.



Initial image — Maximum Likelihood — Potts model

Figure 2. Magnetic Resonnance Image segmentation using a Potts model.

By maximizing the likelihood function, we obtain a first segmentation which is not spatially homogeneous (see figure 2). To regularize the solution, i.e. to obtain smooth regions without holes, we consider a Gibbs field $P(X)$ as prior. The most widely used prior in image segmentation is the Potts model [5,16], which is written as follows:

$$P(X) = \frac{1}{Z}\exp\left\{-\sum_{c=\{s,s'\}}\beta\delta(x_s \neq x_{s'})\right\}, \tag{14}$$

where $\beta > 0$ represents the strength of the prior. The Hamiltonian to be minimized is then written as follows:

$$H_{\text{seg}}(X|Y) = \sum_{c=\{s,s'\}}\beta\delta(x_s \neq x_{s'}) + \sum_{s\in S}\sum_{i\in\Lambda}\left(\frac{(y_s-\mu_i)^2}{2\sigma_i^2} - \log\sqrt{2\pi\sigma_i^2}\right)\delta(x_s = i). \tag{15}$$

Figure 2 shows the segmentation results with and without the Potts model. We can see that the prior removes the local errors due to data.

Although the Potts model succeeds in regularizing the solution, it is not always adapted for image segmentation [32]. Indeed, the obtained solution is at least an approximation of the Hamiltonian ground state. Therefore, in the presence of the phase transition phenomenon the smallest regions may disappear in the segmentation process. Indeed, it has been shown in [31] that, in case of a non-homogeneous external field, phase transition may occur, and as a result the geometry of objects is affected by the Potts model, and especially the elongated areas may disappear. To overcome this problem, several models have been proposed [41,44]. The main idea of these models is to distinguish local noise on configurations from edges and lines. To define the edges, higher range interactions are needed. In [10], a binary model is proposed (the chien-model) taking into account the links between neighboring cliques (supports of the potential). This model has been generalized to the m-ary case in [9]. This model, although regularizing, preserves fine structures and linear shapes in images. In this model, the set of cliques is composed of 3×3 squares. Three parameters (n, l and e) are associated to these patterns.

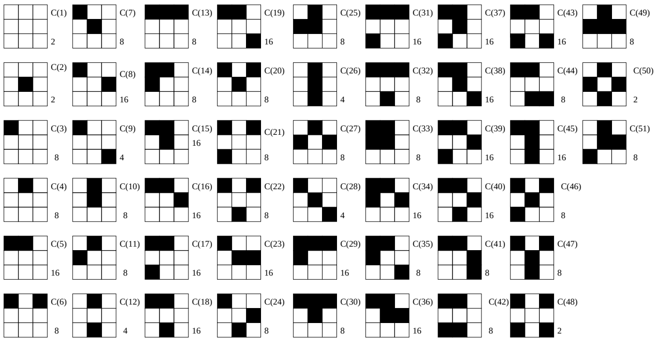


Figure 3. The different classes induced by a binary 3×3 model and their number of elements.

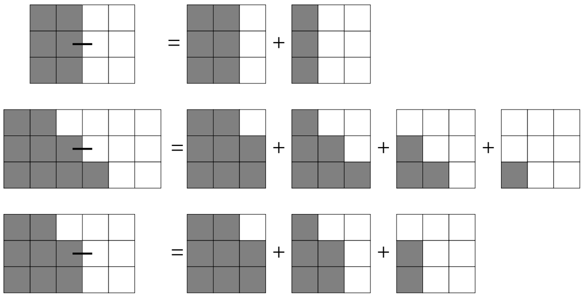


Figure 4. Equations associated with edge constraints.

Before constructing the model, different configurations induced by a 3×3 square are classified using the symmetries (symmetry black-white, rotations, etc.). This classification and the number of elements in each class are described in figure 3. A parameter $C(i)$ is associated to each class and it refers to the value of the potential function for the considered configuration. So, under the hypothesis of isotropy of the model, which induces some symmetries, plus the black/white symmetry, we have fifty one degrees of freedom for such a topology (cliques of 3×3). The construction of the model consists in imposing constraints by relations between its parameters. Two energy functions which differ only by a constant are equivalent, so we suppose that the minimum of the energy is equal to zero. We suppose that constant realizations are ground states for the prior model, so we have the first equation for the parameters given by $C(1) = 0$. We then define the different constraints with respect to those two constant realizations. The first class of constraints concerns the energy of the edges which is noted e per unit of length. Due to symmetries and rotations we just have to define three orientations of the edges corresponding to the eight ones induced by the size of cliques. These constraints and the derived equations are represented in figure 4. Similar constraints are considered to define the energy associated with lines.

To extend the binary chien-model in an m-ary model, we define the energy of a given configuration as the sum of several energies given by the binary model. Consider a configuration and a

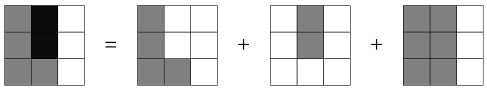


Figure 5. M-ary extension of the chien-model.

given label σ_0. We put all pixels of the configuration that are in state σ_0 to 0 and the others to 1. We then have a binary configuration. The energy of the m-ary model is the sum of the energies obtained by all these deduced binary configurations for the m labels (see figure 5). The potential associated with each configuration is then a linear combination of the three parameters e, l and n:

$$\forall i = 0, \ldots, 51 \; C(i) = \epsilon(i)e + \lambda(i)l + \eta(i)n, \tag{16}$$

and coefficients $\epsilon(i), \lambda(i), \eta(i)$ are defined through the relations between potentials $C(i)$. Then the resulting distribution is written:

$$P_{e,l,n}(X) = \frac{1}{Z(e,l,n)} \exp\left[-eN_0(X) - lN_1(X) - nN_2(X)\right], \tag{17}$$

where:

$$N_0(X) \;=\; \sum_{i=1,\ldots,51} \epsilon(i)\#_i(X), \quad N_1(X) = \sum_{i=1,\ldots,51} \lambda(i)\#_i(X), \quad N_2(X) = \sum_{i=1,\ldots,51} \eta(i)\#_i(X),$$

$\#_i(X)$ being the number of configurations of type i in the realization X.

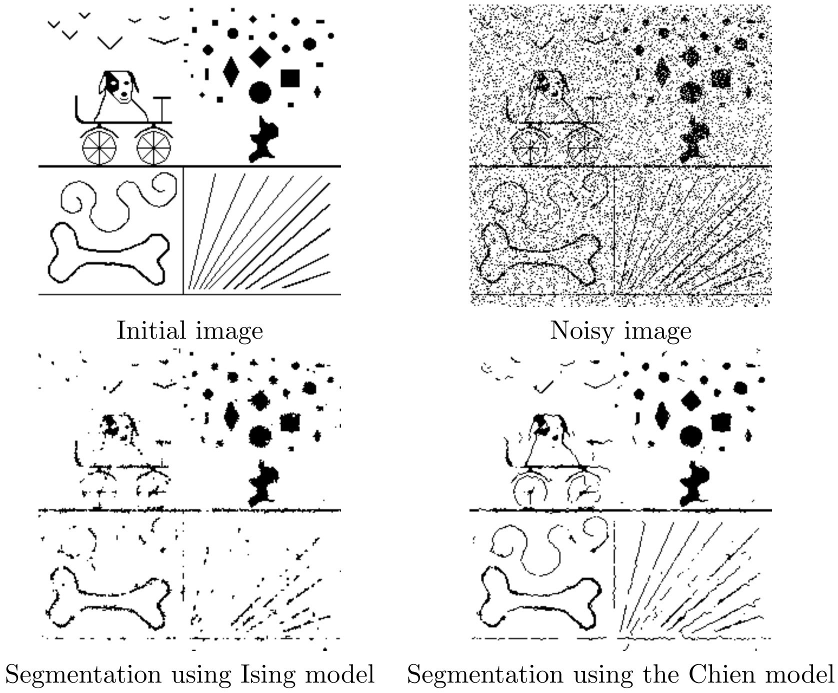


Figure 6. Ising model vs Chien model.

A comparison of Potts and Chien models for fine structures segmentation is shown in figure 6. We have reversed 15% of the pixels in this binary image. The Chien model appears to be much more adapted to image modelling than the Potts model.

4.3. Texture modelling

Another important domain of image processing where Gibbs fields play a leading role is texture modelling, see e.g. [6,12,20,30]. To characterize the objects or specific land cover in an image, the pixel grey level by itself is not always relevant. As shown in figure 7, the radiometry (grey level information) is adapted for the purpose of distinguishing the different fields. Within the urban area, the grey levels are almost uniformly distributed. Therefore, to decide if a pixel belongs to an urban area or not, the grey level information is not sufficient. To distinguish urban areas, we then have to consider the local distribution of grey levels. In the Gibbs field approach, we assume that, locally, the grey levels are distributed according to a Gibbs distribution, and we estimate the parameters associated with this Gibbs distribution. The parameter values are used in order to make a decision, instead of the grey level values. If the goal is to analyze texture, for example to make a decision about urban area vs fields, then simple models leading to fast estimation techniques are preferred. When the goal is to model the textures themselves in order to synthesize them, generic models are addressed. In this context, the relevance of Gibbs modelling is shown in [20], where high range pairwise interactions are considered. Herein, we only derive a very simple model, i.e. the four connected isotropic Gaussian Markov Random Fields, to extract urban areas from satellite images [12].

Initial image (© CNES/SPOTIMAGE) $\quad$ β map $\quad$ Urban area

Figure 7. Urban areas detection using a Gaussian Gibbs Field model.

Let us consider an image $X = (x_s), s \in S$, where $x_s \in \Lambda$. The grey level space (or state space) is typically $\Lambda = \{0, 1, \ldots, 255\}$. We assume that locally the considered image is a realization of a Gibbs field with the following Hamiltonian:

$$H_\theta(X) = \beta \left(\sum_{c=\{s,s'\}} (x_s - x_{s'})^2 + \lambda \sum_{s \in S} (x_s - \mu)^2 \right), \tag{18}$$

where $\theta = (\beta, \lambda, \mu)$ are the model parameters. Different estimator can be used to obtain the parameter value. For instance, the Maximum Likelihood (ML) estimator is given as:

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_\theta P_\theta(X). \tag{19}$$

The algorithm for the ML estimator usually requires a long computational time. Therefore, different criteria, easier to estimate, can be used, such as Maximum of the Pseudo Likelihood (MPL):

$$\hat{\theta}_{\mathrm{MPL}} \;=\; \arg\max_\theta \prod_{s \in S} P(x_s | x_t, t \in S/\{s\}), \tag{20}$$

where conditional probabilities in (20) are found using $H_\theta(X)$. In our example parameter β is estimated. The higher β, the higher probability to be in an urban area. After estimating β on each pixel by considering a local window, the β map is segmented to delineate urban areas, as shown in figure 7, see [30].

5. Spatial point processes for object detection problems

5.1. Marked Point Processes and Reversible Jump MCMC

In this section we discuss stochastic algorithms in the framework of the Gibbs fields approach for feature extraction problems. These problems become critical in remote sensing with the development of high resolution sensors for which the object geometry is well defined. In lattice based models each pixel is considered to be a random variable. In this setting a local definition of constraints is more natural, and it is difficult to include strong non-local geometrical constraints into lattice based models. In addition, the pixelwise approach seems to be non-adequate in cases of geometrical noise arising from "trash" information on the scene such as cars or shadows. Consequently, new problems required new models, and the marked point process framework is found very proper for feature extraction problems from remotely sensed data.

The main idea behind a marked point process is to model random sets of objects within a stochastic framework. Random sets of objects are represented in the model by marked point configurations in a continuous space. The shape and the size of each object is described by a mark, and the location of the object by a point. The probability distribution on the configuration space is defined by a density function with respect to the Poisson measure in a finite volume (Radon Nykodim derivative), see [40]. This density consists of several terms, and generally the posterior distribution including a prior and a data term can be written as a Gibbs reconstruction of the Poisson measure. Following the scheme described above for seeking global minimizers of the energy function, we consider various stochastic dynamics with a given stationary Gibbs measure, such as spatial birth-and-death processes or Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithms [1,2,18,19,21,22,37,38], the latter being an extension of the Metropolis-Hastings method adapted for general state spaces. The main property of these two schemes is their ability to manage a random number of points in the configuration.

These iterative algorithms work by choosing a new configuration according to a transition distribution from the current one by proposing a local change. The birth and death algorithms permit the creation of a new object or the removal of an existing object. In addition, the RJMCMC algorithms permit other operations on objects such as splitting, merging, translations, as well as modifications of the marks of objects using, for example, rotations or dilations. Finally, these algorithms are embedded into a simulated annealing scheme.

Thus, the main advantages of a marked point process in image analysis consist in their geometrical adaptativity and generality. Any geometrical properties can easily be introduced into the model through the object geometry. Different types of objects (trees, roads, buildings, etc.) can be considered within the same model but of course with appropriate interactions. Moreover, interactions between the points permit to model some prior information on the object configuration, and the data are taken into account at the object level, thus improving robustness of the algorithms. Together with these evident positive features there is one large drawback of RJMCMC algorithms, they converge very slowly. The slow convergence is caused by the structure of the algorithms: at each step of the iterated scheme, only one operation with one or two objects from the current configuration can be realized. Moreover, the rejection principle of the Metropolis scheme, although ensuring the convergence, introduces computation without any change in the current configuration.

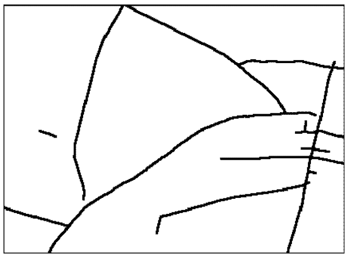
RJMCMC approach has been used in order to detect different features such as road network [28, 39], buildings [34] or trees [36]. We briefly present two models - for road and tree extraction - in the framework of RJMCMC algorithms approach, and then introduce a new model for object detection based on an approximation of a continuous time dynamics.

5.1.1. Quality Candy model: road network extraction

Road network extraction from satellite images is a challenging problem which finds numerous applications in cartography. Different approaches have been proposed but a few provide a fully automatic extraction, and Gibbs fields models among them, see e.g. [28,39,42]. We present one of them, based on a marked point process modelling, referred to as the Quality Candy model. The considered objects are segments which are determined by a point (the position of the center) and marks (the orientation and the length).

The prior knowledge models the high connectivity and the low curvature of a road network. Each segment has two attractive areas surrounding the ends of the segment and a repulsive area around the center of the segment. If a segment intersects the attractive area of another segment, we have an attractive interaction with an increasing intensity under decreasing orientation difference between these two segments. If a segment intersects the repulsive area of another segment, we have a repulsive interaction with a decreasing intensity when the orientation difference between the segments tends to $\pi/2$. In addition, there is some repulsive energy (a penalty term) associated with the segments having one or two unconnected ends. These terms allow us to control the shape of the network and the crossings, and the connectivity of the network. The data term contributes to the energy through a sum of local energy functions on each segment. For a given segment, the local energy depends on a statistical test between the pixel distribution within the segment projection on the lattice and the pixel distribution in two masks covering the left and right sides of this projection. The higher the contrast between the segment and the neighboring masks, the lower the energy. The optimization is performed by an RJMCMC algorithm embedded into a simulated annealing scheme. The RJMCMC perturbation kernel includes birth and death, translation and rotation. A key point in increasing the convergence speed is a new kernel referred to as "birth and death in a neighborhood" which consists in proposing a new segment in the neighborhood of segments from the current configuration. An example of the result is shown in figure 8. On this aerial image, tree shadows partially eclipse the road. The prior permits the detection process to provide a connected network by reconstructing the missing information.

Aerial image (©) IGN

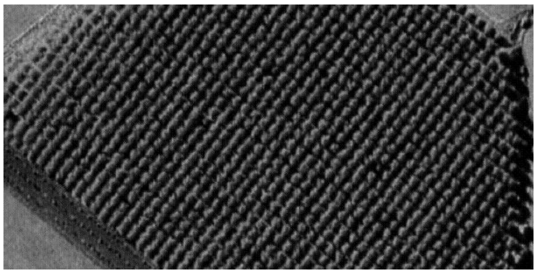


Detected network

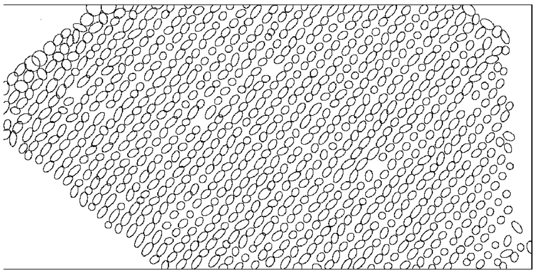
Figure 8. Road network detection using the Quality Candy model.

5.1.2. RJMCMC for tree detection

Tree detection from aerial images provides important information for forestry management and monitoring. Using an ellipse as object, marked point process can perform this task. In [36],



Tree plantation (©) IFN



Detected trees

Figure 9. Tree detection using a marked point process.

the authors proposed a prior penalizing overlapping ellipses and a data term depending on the

Bhattacharya distance between the pixel distribution inside the ellipse projection on the lattice and the pixel distribution in a crown surrounding this ellipse. This model detects a collection of non-overlapping or slightly overlapping ellipses that exhibit a contrast with their surrounding pixels in the considered image. The RJMCMC kernels considered for the optimization contain birth and death, rotation and dilation of ellipses. In case of plantations, the trees are regularly placed along two directions. When this information is available, we also consider a birth and death kernel to favor birth of trees according to this periodicity. The result of this approach is shown in figure 9.

5.2. A new object based algorithm

To improve the convergence properties of RJMCMC algorithms we recently proposed a new multi-object adjusted algorithm for object detection [11], based on a spatial birth and death process, reversible with respect to the given Gibbs distribution, see [4,27], and a discrete time approximation of this process. The reversibility is provided by the so-called detailed balance conditions on birth and death rates. In our scheme, the birth intensity is constant, whereas death intensities depend on the energy function and the current configuration. This choice of rates has been made to optimize the convergence speed. Indeed, the volume of the space for birth is much bigger than the number of discs in the configuration. It is therefore faster to update the death map than the birth map.

Then we embed the defined stationary dynamics into a simulated annealing procedure when the temperature of the system tends to zero in time. Thus we obtain a non-stationary stochastic process, such that all weak limit measures have a support on configurations giving the global minimum of the energy function under a minimal number of discs in the configuration. The final step is discretization of this non-stationary dynamics. This last step leads to a non-homogeneous, in time and in space, Markov chain with transition probabilities depending on temperature, energy function and discretization step. We prove that:

1) discretization process converges to the continuous time process under fixed temperature as the step of discretization tends to zero;

2) if we apply the discretization process to any initial measure with a continuous density w.r.t. the Lebesgue-Poisson measure, then, in the limit, when the discretization step tends to 0, time tends to infinity and the temperature tends to 0, we get a measure concentrated on the global minima of the energy function with a minimal number of discs.

These results confirm that the proposed algorithm based on the discretization scheme together with the cooling procedure solves the problem of searching configurations giving the global minima of the energy function.

We applied this framework to object detection from images. The crucial advantage of our approach is that each step concerns the whole configuration and there is no rejection. Thus we obtained better performances in terms of computational time, which permit to address real images containing several million of pixels. We take an energy function involving prior knowledge on the object configuration such as partial non overlapping between objects and a data term which permits the objects to fit the image under study. Further in this section we briefly present the main constructions and theorems on convergence of the invented algorithm following [11].

5.2.1. The model

We consider finite systems of partially overlapping discs $\{d_{x_1}, \dots, d_{x_k}\}$ of the same radius r with a hard core distance $\epsilon < r$ between any two elements, lying in a bounded domain $V \subset R^2$. Let

$$\gamma = \{x_i\} \in \Gamma_d(V), \quad x_i \in V \subset R^2,$$

be a configuration of the centers of discs and $\Gamma_d(V)$ denotes the configuration space of the discs centers in V. Notice, the set $\Gamma_d(V)$ can be decomposed into strata:

$$\Gamma_d(V) \ = \ \bigcup_{n=0}^{N} \Gamma_d(V, n),$$

where each stratum $\Gamma_d(V,n)$ is the set of configurations containing n discs, $\Gamma_d(V,0) = \{\emptyset\}$, and N is the maximal number of discs in V. As a reference measure in the space $\Gamma_d(V)$ we take the Lebesgue-Poisson measure λ.

On the space $\Gamma_d(V)$ we define a real-valued smooth function $H(\gamma)$ bounded from below which is called the energy function. It can be written as a vector function

$$H = (0,\ H_1(x_1),\ H_2(x_1,x_2),\dots,H_N(x_1,\dots,x_N)) \tag{21}$$

with $H(\emptyset) = 0$. In practice, this energy is a sum of two terms. The first one represents prior knowledge on the discs configuration and is defined by interactions between neighboring discs; the second term is defined by the data for each object, and it can be negative.

The Gibbs distribution μ_β on the space $\Gamma_d(V)$ generated by the energy $H(\gamma)$ is defined by the density $p_V(\gamma) = \frac{\mathrm{d}\mu_\beta}{\mathrm{d}\lambda}(\gamma)$ with respect to the Lebesgue-Poisson measure λ:

$$p_V(\gamma) = \frac{z^{|\gamma|}}{Z_{\beta,V}}\ \exp\{-\beta H(\gamma)\}, \tag{22}$$

with positive parameters $\beta > 0,\ z > 0$ and a normalizing factor $Z_{\beta,V}$:

$$Z_{\beta,V} = \int_{\Gamma_d(V)} z^{|\gamma|}\exp\{-\beta H(\gamma)\}\mathrm{d}\lambda(\gamma) = 1 + \sum_{n=1}^{N}\frac{z^n}{n!}\int_{V_d^n} \mathrm{e}^{-\beta H_n(x_1,\dots,x_n)}\mathrm{d}x_1\dots\mathrm{d}x_n.$$

Now we formulate some assumptions on the energy function $H_V(\gamma)$. Denote by

$$\bar{H} = \min_{\gamma\in\overline{\Gamma_d(V)}} H(\gamma),$$

where $\overline{\Gamma_d(V)}$ is the closure of $\Gamma_d(V)$, and let

$$T_V = \{\bar{\gamma}\in\overline{\Gamma_d(V)}:\ H(\bar{\gamma}) = \bar{H}\}$$

be a set of all global minimizers of the function $H(\gamma)$. The set T_V can be written as

$$T_V = \bigcup_{n=0}^{N} T_{V,n},$$

where $T_{V,n}$ is a set of configurations from T_V which are also configurations from $\Gamma_d(V,n)$, i.e. contain exactly n discs.

We assume that
1) the set T_V is finite and situated in $\Gamma_d(V)$,
2) for any configuration $\bar{\gamma}\in T_{V,n}$

$$\frac{\partial H_n}{\partial y_i^m}(\bar{x}_1,\dots,\bar{x}_n) = 0,$$

for any $i = 1,\dots,n,\ \bar{x}_i = (y_i^1, y_i^2)\in V,\ m = 1,2$, and the matrix

$$\mathcal{A}\ (\bar{\gamma}) = \left\{\frac{\partial^2 H_n}{\partial y_i^{m_1}\partial y_j^{m_2}}(\bar{x}_1,\dots,\bar{x}_n)\right\}$$

at point $\bar{\gamma} = (\bar{x}_1,\dots,\bar{x}_n)$ is strictly positively defined. Under the above assumptions the following theorem holds.

Theorem 1. [11] *Let $n_0 \in [0, \dots, N]$ be the minimal number for which the set $T_{V,n}$ is not empty. Then the Gibbs distributions μ_β weakly converge as $\beta \to \infty$ to a distribution μ_∞ on $\Gamma_d(V)$ of the form*

$$\mu_\infty \;=\; \sum_{\bar\gamma \in T_{V,n_0}} C_{\bar\gamma}\delta_{\bar\gamma} \;\textit{if}\; n_0 > 0, \quad \textit{and} \quad \mu_\infty \;=\; \delta_{\{\emptyset\}} \;\textit{if}\; n_0 = 0. \tag{23}$$

Here $\delta_{\bar\gamma}$ is the unit measure concentrated on the configuration $\bar\gamma$, and the coefficients $C_{\bar\gamma}$ hold the equality

$$\sum_{\bar\gamma \in T_{V,n_0}} C_{\bar\gamma} \;=\; 1.$$

5.2.2. A continuous-time equilibrium dynamics

Now we define a continuous time birth and death process via its generator as follows:

$$(L_\beta\; f)(\gamma) \;=\; \sum_{x \in \gamma} \mathrm{e}^{\beta E(x,\gamma\backslash x)}(f(\gamma\backslash x) - f(\gamma)) \;+\; z \int\limits_{V(\gamma)} (f(\gamma \cup y) - f(\gamma))\mathrm{d}y, \tag{24}$$

where

$$V(\gamma) \;=\; V\backslash D(\gamma), \quad D(\gamma) \;=\; (\cup_{x\in\gamma}\mathcal{B}_x(\epsilon)) \cap V,$$

$\mathcal{B}_x(\epsilon)$ is a disk with a center at point $x \in V$ and radius ϵ, and

$$E(x, \gamma\backslash x) \;=\; H(\gamma) - H(\gamma\backslash x).$$

In this context, the birth intensity $b(\gamma, x)$ in the unordered configuration γ at x and the death intensity $d(\gamma\backslash x, x)$ from the configuration γ at position x are respectively given by:

$$b(\gamma, x)\mathrm{d}x \;=\; z\; \mathrm{d}x, \qquad d(\gamma\backslash x, x) \;=\mathrm{e}^{\beta E(x,\gamma\backslash x)}.$$

Under this choice of the birth and death intensities, the detailed balance condition holds:

$$\frac{b(\gamma, x)}{d(\gamma\backslash x, x)} \;=\; \frac{p_V(\gamma)}{p_V(\gamma\backslash x)} \;=\; z\mathrm{e}^{-\beta E(x,\gamma\backslash x)},$$

and consequently, see for example [38], the corresponding birth-and-death process associated with the stochastic semigroup $T_\beta(t) = \mathrm{e}^{tL_\beta}$ is time reversible, and its equilibrium distribution is the Gibbs stationary measure μ_β with density (22).

The convergence to the stationary measure μ_β is guaranteed by the general result by C. Preston, see [37]. We consider a family $\mathbb{B}(\lambda)$ of measures ν on the space $\Gamma_d(V)$ with a bounded density $\tilde p_\nu(\gamma)$ with respect to Lebesgue-Poisson measure λ. This, in particular, implies that any density $p_\nu(\gamma)$ of the measure $\nu \in \mathbb{B}(\lambda)$ w.r.t. a Gibbs measure μ_β (for any β):

$$p_\nu(\gamma) \;=\; \frac{\mathrm{d}\nu}{\mathrm{d}\mu_\beta}(\gamma)$$

is also bounded, and consequently, $p_\nu(\gamma) \in L_2(\Gamma_d(V), \mu_\beta)$. Then we can define the evolution $\nu_t \equiv T(t)\nu$ of the measure $\nu \in \mathbb{B}(\lambda)$ as follows:

$$\langle \nu_t, F\rangle \;=\; (T_\beta(t)p_\nu, F)_{\mu_\beta}, \quad \forall F \in L_2(\Gamma_d(V), \mu_\beta).$$

Theorem 2. *Let $\nu \in \mathbb{B}(\lambda)$. Then for any $F \in L_2(\Gamma_d(V), \mu_\beta)$ we get*

$$\langle T_\beta(t)\nu, F\rangle \;\equiv\; \langle \nu_t, F\rangle \;=\; (T_\beta(t)p_\nu, F)_{\mu_\beta} \;\to\; \langle F\rangle_{\mu_\beta}. \tag{25}$$

The proof of theorem 2 follows from the general theorems, see e.g. [37].

5.2.3. Approximation process

Here we define a discrete time approximation of the proposed continuous birth and death process generated by (24).

We consider Markov chains $T_{\beta,\delta}(n), n = 0, 1, 2, \ldots$ on the same space $\Gamma_d(V)$. The process $T_{\beta,\delta}(n)$ can be described as follows: a configuration γ is transformed to a configuration $\gamma' = \gamma_1 \cup \gamma_2$, where $\gamma_1 \subseteq \gamma$, and γ_2 is a configuration of centers of discs such that $\gamma \cap \gamma_2 = \emptyset$ and is distributed w.r.t. the Poisson law with intensity z.

This transformation embeds a birth part given by γ_2 and a death part given by $\gamma \backslash \gamma_1$.

The transition probability for the death of a particle at x (i.e. a disc with the center at x) from the configuration γ is given by:

$$p_{x,\delta} = \begin{cases} \dfrac{\mathrm{e}^{\beta E(x,\gamma\backslash x)}\ \delta}{1\ +\ \mathrm{e}^{\beta E(x,\gamma\backslash x)}\ \delta} = \dfrac{a_x\delta}{1+a_x\delta}, & \text{if} \quad \gamma \to \gamma\backslash x, \\ \\ \dfrac{1}{1+a_x\delta}, \quad \text{if} \quad \gamma \to \gamma \quad (x \text{ survives}). \end{cases} \tag{26}$$

with $a_x = a_x(\gamma) = \mathrm{e}^{\beta E(x,\gamma\backslash x)}$. Moreover, all the particles are killed independently, and both configurations γ_1 and γ_2 are independent.

The transitions associated with the birth of a new particle in a small domain $\Delta y \subset V(\gamma)$ have the following probability distribution:

$$q_{y,\delta} = \begin{cases} z\Delta y\delta, \quad \text{if} \quad \gamma \to \gamma \cup y, \\ \\ 1 - z\Delta y\delta, \quad \text{if} \quad \gamma \to \gamma \quad (\text{ no birth in } \Delta y). \end{cases} \tag{27}$$

Finally, the transition operator $P_{\beta,\delta}$ for the process

$$T_{\beta,\delta}(n) = P^n_{\beta,\delta}$$

has the following form:

$$\begin{aligned}(P_{\beta,\delta}f)(\gamma) = {} & \sum_{\gamma_1\subseteq\gamma} \prod_{x\in\gamma_1} \frac{1}{1+a_x\delta} \prod_{x\in\gamma\backslash\gamma_1} \frac{a_x\delta}{1+a_x\delta} = \\ & \Xi_\delta^{-1}(\gamma_1) \sum_{k=0}^{N-|\gamma_1|} \int\limits_{V_k(\gamma_1)} \frac{(z\delta)^k}{k!} f(\gamma_1 \cup y_1 \cup \ldots \cup y_k)\mathrm{d}y_1 \ldots \mathrm{d}y_k\,, \end{aligned} \tag{28}$$

where $\Xi_\delta(\gamma) = \Xi_\delta(V(\gamma), z, \delta)$ is the normalizing factor for the conditional Lebesgue-Poisson measure under given configuration of discs γ.

Using approximation technique, see [14], we proved that the approximation process $T_{\beta,\delta}(t) \equiv T_{\beta,\delta}\left(\left[\frac{t}{\delta}\right]\right)$ converges to the continuous time process $T_\beta(t)$ uniformly on bounded intervals $[0, t]$ as the discretization step δ tends to 0. Let us denote $\mathcal{L} = B(\Gamma_d(V))$ a Banach space of bounded functions on Γ_V with a norm

$$\|F\| = \sup_{\gamma\in\Gamma_d(V)} |F(\gamma)|.$$

Theorem 3 [11]. *For each* $F \in \mathcal{L}$

$$\|T_{\beta,\delta}(t)F - T_\beta(t)F\|_{\mathcal{L}} = \sup_\gamma |(T_{\beta,\delta}(t)F)(\gamma) - (T_\beta(t)F)(\gamma)| \to 0, \tag{29}$$

as $\delta \to 0$ *for all* $t \geqslant 0$ *uniformly on bounded intervals of time.*

We denote by $S_{\beta,\delta}(n)$ an adjoint to $T_{\beta,\delta}(n)$ semigroup acting on measures, such that for any $\nu \in \mathbb{B}(\lambda)$:

$$\langle S_{\beta,\delta}(n)\nu, F\rangle = (p_\nu, T_{\beta,\delta}(n)F)_{\mu_\beta}.$$

We now formulate the main result about convergence under a cooling procedure.

Theorem 4 [11]. *Let $F \in B(\Gamma_d(V))$ and an initial measure $\nu \in \mathbb{B}(\lambda)$. Then under relation*

$$\delta \, e^{\beta b} \ < \ const \tag{30}$$

with $b = \sup_{\gamma \in \Gamma_d(V)} \ \sup_{x \in \gamma} E(x, \gamma \backslash x)$ we have

$$\lim_{\beta \to \infty, \ t \to \infty, \ \delta \to 0} \langle F \rangle_{S_{\beta,\delta}([\frac{t}{\delta}])\nu} \ = \ \langle F \rangle_{\mu_\infty}, \tag{31}$$

where measure μ_∞ is defined in Theorem 1, and $\langle F \rangle_{S_{\beta,\delta}([\frac{t}{\delta}])\nu} \ = \ \langle S_{\beta,\delta}([\frac{t}{\delta}])\nu, F \rangle$.

5.3. Application to object detection from numerical images

5.3.1. Model

Let us consider a digital image $Y = (y_s)_{s \in S}$ on a finite subset of the lattice $S \subset \mathbb{Z}^2$. We consider configurations of center of discs $\gamma = \{x_i\} \in \Gamma_d(V)$. A hard core distance ϵ, which avoids two disks to be centered in the same pixel, is taken to be equal to the data resolution: $\epsilon = 1$ pixel. To define the energy H, we first consider some prior knowledge. We want to minimize the overlap between the objects. However, to obtain a more flexible model w.r.t. data and the object shape, we do not forbid but only penalize the overlapping objects. We define a pairwise interaction as follows:

$$H_2(x_i, x_j) = \max\left(0, \ 1 - \frac{||x_j - x_i||}{2r}\right) \quad \forall \{x_i, x_j\} \in \gamma^2 : \ ||x_j - x_i|| \geqslant 1, \tag{32}$$

where $||.||$ is the Euclidean norm and r is the radius of the underlying disc.

A data term is then added for each object to fit the disc configuration onto data. We consider that there is an object, modelled by a disc centered in pixel s, in the image, if the grey level values of the pixels inside the projection of the disc onto the lattice are statistically different from those of the pixels in the neighborhood of the disc. To quantify this difference we compute the Bhattacharya distance between the associated distributions. Denote $D_1(s)$ the projection of the disc with radius r centered in s onto the lattice S, and $D_2(s)$ the surrounding crown:

$$D_1(s) = \{t \in S : ||t - s|| \leqslant r\} \ \text{ and } \ D_2(s) = \{t \in S : r < ||t - s|| \leqslant r + 1\}. \tag{33}$$

We consider the mean and the variance of the data of these two subsets:

$$\mu_1(s) = \frac{\sum_{t \in D_1(s)} y_t}{\sum_{t \in D_1(s)} 1} \quad \text{and} \quad \mu_2(s) = \frac{\sum_{t \in D_2(s)} y_t}{\sum_{t \in D_2(s)} 1},$$
$$\sigma_1^2(s) = \frac{\sum_{t \in D_1(s)} y_t^2}{\sum_{t \in D_1(s)} 1} - \mu_1(s)^2 \quad \text{and} \quad \sigma_2^2(s) = \frac{\sum_{t \in D_2(s)} y_t^2}{\sum_{t \in D_2(s)} 1} - \mu_2(s)^2. \tag{34}$$

Assuming Gaussian distributions, the Bhattacharya distance between the distributions in $D_1(s)$ and in $D_2(s)$ is then given by:

$$B(s) = \frac{1}{4} \left(\mu_1(s) - \mu_2(s)\right)^2 \sqrt{\sigma_1^2(s) + \sigma_2^2(s)} - \frac{1}{2} \log \frac{2\sigma_1(s)\sigma_2(s)}{\sigma_1^2(s) + \sigma_2^2(s)}. \tag{35}$$

From this distance between the two distributions, a first order energy term is built:

$$\forall x_i \in \gamma, \ H_1(x_i) = \begin{cases} \left(1 - \frac{B(i)}{T}\right) & \text{if} \quad B(i) < T, \\ \left(\exp\left\{-\frac{B(i) - T}{3B(i)}\right\} - 1\right) & \text{if} \quad B(i) \geqslant T, \end{cases} \tag{36}$$

where i is the closest point to x_i on the lattice, and T is a threshold parameter. Finally, under the hard core constraint, the global energy is written as follows:

$$H(\gamma) = \alpha \sum_{x_i \in \gamma} H_1(x_i) + \sum_{\{x_i, x_j\} \in \gamma^2, \ x_i \neq x_j} H_2(x_i, x_j), \tag{37}$$

where α is a weighting parameter between the data term and the prior knowledge.

5.3.2. Algorithm

The algorithm simulating the process is defined as follows:

- **Computation of the data term:** For each site $s \in S$ compute $H_1(s)$ from the data

- **Computation of the birth map:** To speed up the process, we consider a non-homogeneous birth rate to favor birth where the data term is low (where the data tends to define an object):

$$\forall s \in S,\ b(s) = 1 + 9\frac{\max_{t \in S} H_1(t) - H_1(s)}{\max_{t \in S} H_1(t) - \min_{t \in S} H_1(t)}\,. \tag{38}$$

  The normalized birth rate is then given by:

$$\forall s \in S,\ B(s) = \frac{zb(s)}{\sum_{t \in S} b(s)}\,. \tag{39}$$

  This non-homogeneous birth rate refers to a non-homogeneous reference Poisson measure. It has no impact on the convergence to the global minima of the energy function but does have an impact on the speed of convergence in practice by favoring birth in relevant locations.

- **Main program:** initialize the inverse temperature parameter $\beta = \beta_0$ and the discretization step $\delta = \delta_0$, alternate birth and death steps

  – **Birth step:** for each $s \in S$, if $x_s = 0$ (no point in s) add a point in s ($x_s = 1$) with probability $\delta B(s)$ (note that the hard core constraint with $\epsilon = 1$ pixel is satisfied).

  – **Death step:** consider the configuration of points $x = \{s \in S : x_s = 1\}$ and sort it from the highest to the lowest value of $H_1(s)$. For each point taken in this order, compute the death rate as follows:

$$d_x(s) = \frac{\delta a_x(s)}{1 + \delta a_x(s)}, \tag{40}$$

  where:

$$a_x(s) = \exp\left\{-\beta\left(H(x/\{x_s\}) - H(x)\right)\right\} \tag{41}$$

  and kill s ($x_s = 0$) with probability $d(s)$.

  – **Convergence test:** if the process has not converged, decrease the temperature and the discretization step by a given factor and go back to the birth step. The convergence criterion is usually given in practice by a fixed number of iterations.

5.3.3. Results

The first application concerns tree crown extraction from aerial images. We consider 50cm resolution images of poplars. An example of the obtained results is given in figure 10. The results are satisfactory. One can remark a few false alarms on the border of the plantation, due to shadows.

The second application concerns the problem of flamingo population counting. A generalization of this work to ellipses was done to better fit the flamingo shape [8]. The obtained result is given in figure 11 for the initial image of the flamingo colony, and in figure 12 for a fragment of the whole image of the detected birds. The birds were correctly detected. This colony was automatically counted in about 20 minutes, and we got 3682 detected flamingos while the expert manually counted 3684 individuals. This represents the main advantage with respect to more standard optimization techniques based on a RJMCMC sampler. Indeed, the speed of convergence and the computational efficiency of the proposed algorithm allows us to deal with huge datasets in a reasonable delay.

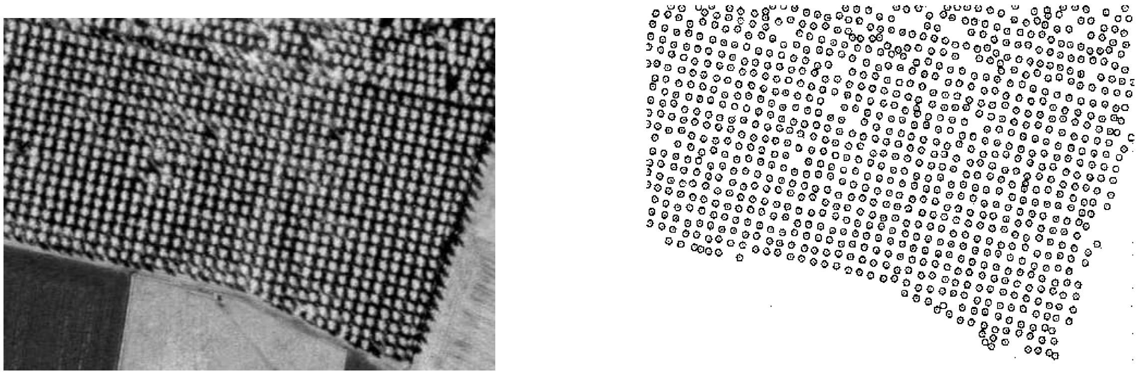


Figure 10. The result on a poplar plantation (left: initial image © IFN; right: detected trees).

Figure 11. Bird population © Station Biologique Tour du Valat.

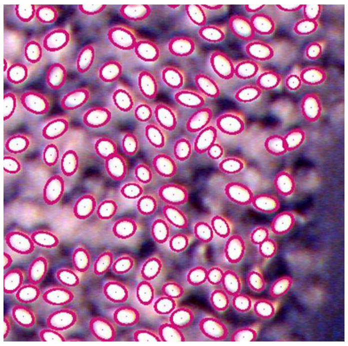


Figure 12. Extract of the detected birds from the image shown in figure 11.

6. Conclusions

In this paper, we have given a brief overview of Gibbs field approaches in image processing. This survey is far from being complete. Indeed, numerous different topics, such as movement detection, stereo-matching, texture synthesis or shape from shading, can be addressed within this framework. We have divided the models in two classes. The first class represents a discrete modelling of the solution, i.e. when we are searching for a numerical image obtained from the data by a restoration or a segmentation process. We have shown that Gibbs fields represent an elegant way to introduce prior information on the solution, such as regularity and smoothing constraints. In the recent years, new models have been proposed to overcome the posterior modelling, consisting in defining independently a prior and a likelihood, by using for example the Conditional Random Fields or the Couple Fields approaches [3,29]. One can directly model the joint distribution as a Gibbs field. In this case local interactions can depend on data, for example, smoothing constraint can depend on the data gradient. The second class contains the models defined in a continuous plane. The main characteristic of these models is that they manipulate objects instead of pixels. Therefore, they make it possible to take into account the geometrical information in the data and in the expected results. These models are more recent. We have described some recent developments regarding the dynamics associated with the optimization problem in order to speed up the convergence. In this paper, we have not addressed the parameter estimation problem. In case of lattice based models, different efficient approaches have been proposed. This point is still an open issue for marked point processes. Recent developments of high resolution sensors generate a new class of image analysis problems focusing on automatic data flow handling. And, this new context, Gibbs fields approach appears to be just as promising and useful, as it used to be 30 years ago.

References

1. Baddeley A., Moller J., Nearest-neighbour Markov point processes and random sets, International Statistical Review 57, 1989, 89–121.
2. Baddeley A., Van Lieshout M.N., Stochastic geometry models in high-level vision. Statistics and Images, 1993, **20**, 235–256, 1993.
3. Benboudjema D., Pieczynski W., Unsupervised image segmentation using triplet Markov fields. Computer Vision and Image Understanding, 2005, **99**, 3, 476–498.
4. Bertini L., Cancrini N., Cesi F., The spectral gap for a Glauber-type dynamics in a continuous gas. Ann. Inst. H. Poincare Probab. Statist., 2002, **38**, 91–108.
5. Besag J., Spatial interaction and the statistical analysis of lattice systems (with discussion). J. Roy. Statist. Soc. Ser. B, 1974, **36**, 192–236.
6. Cross A., Jain K. Random Markov Field Textures. IEEE Trans. Pattern Anal. Machine Intelligence, 1983, **5**.
7. Boykov Y., Kolmogorov V., An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, **26**, 9, 1124–1137.
8. Descamps S., Descombes X., Béchet A., Zerubia J., Détection de flamants roses par processus ponctuels marqués pour l'estimation de la taille des populations (In French), Research Report 6328, INRIA, 2007.
9. Descombes X., Application of stochastic techniques in image processing for automatic tissue classification in MRI and blood vessel restoration in MRA, Tech. Rep. KUL/ESAT/MI2/9603, Laboratory for Medical Imaging Research (ESAT-Radiology), K.U.Leuven, Belgium, 1996.
10. Descombes X., Mangin J.F., Pechersky E., Sigelle M., Fine structures preserving model for image processing. Proc. 9th SCIA 95 Uppsala, Sweden, 1995, 349–356.
11. Descombes X., Minlos R., Zhizhina E., Object extraction using a stochastic birth-and-death dynamics in continuum, Research Report 6135, INRIA, 2007.
12. Descombes X., Sigelle M., Preteux F., GMRF Parameter Estimation in a non-stationary Framework by a Renormalization Technique: Application to Remote Sensing Imaging. IEEE Transactions on Image Processing, 1999, **8**, 4, 490–503.
13. Descombes, X., Zhizhina E., Applications of Gibbs fields methods to image processing problems. Problems of Information Transmission, 2004, **40** (3), 108–125.
14. Ethier S.N., Kurtz T.G., Markov processes. Characterization and convergence. NY, 1986.
15. Geman D., Random fields and inverse problems in imaging. Lecture Notes in Math., vol. 1427, 113–193, Springer-Verlag, 1991.
16. Geman S., Geman D., Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. IEEE Trans. Pattern Anal. Machine Intelligence, 1984, **6**, 6, 721–741.
17. Geman S., Reynolds G., Constrained restoration and recovery of discontinuities. IEEE Trans. Pattern Anal. Machine Intelligence, 1992, **14**, 3, 367–383.
18. Geyer C.J., Likelihood inference for spatial point processes. Stochastic Geometry, Likelihood and Computation, Eds. Barndorff-Nielsen, Kendall, van Lieshout, 1996.
19. Geyer C.J., Moller J., Simulation and likelihood inference for spatial point process. Scandinavian Journal of Statistics B, 1994, **21**, 359–373.
20. Gimel'farb G. Image Textures and Gibbs Random Fields. Kluwer Academic Publishers: Dordrecht, 1999.
21. Green P.J., Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 1995, **82**(4), 711–132.
22. Green P.J., Richardson S., On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Stat. Society B, 1997, **59**(4), 731–792.
23. Grenander U., Miller M., Representations of knowledge in complex systems. Journal of the Royal Stat. Soc. B, 1994, **56**(4), 549–603.
24. Haario H., Saksman E., Simulated annealing process in general state space. Adv. Appl. Probab., 1991, **23**(4), 866–893.
25. Hajek B., Cooling schedules for optimal annealing. Math. Operat. Res., 1988, **13**, 311–329.
26. Kirkpatrick S., Gellatt C.D., Vecchi M.P., Optimization by simulated annealing. Science, 1983, **220**, 671–680.
27. Kondratiev Y.G., Minlos R.A., Zhizhina E.A., One particle subspace of the Glauber dynamics generator for continuous particle systems. Reviews in Math. Physics, 2004, **16**, 9, 1073–1114.
28. Lacoste C., Descombes X., Zerubia J., Point Processes for Unsupervised Line Network Extraction in

Remote Sensing. IEEE Trans. Pattern Analysis and Machine Intelligence, 2005, **27**(10), 1568–1579.
29. Lafferty J., McCallum A., Pereira A. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data – In Proceedings of the Eighteenth International Conference on Machine Learning (ICML–2001).
30. Lorette A., Descombes X., Zerubia J., Texture Analysis through a Markovian Modelling and Fuzzy Classification: Application to Urban Area Extraction from Satellite Images. International Journal of Computer Vision, 2000, **36**, 3, 219–234.
31. Maruani A., Pechersky E., Sigelle M., On Gibbs Fields in Image Processing. Markov Processes and Related Fields, 1995, **1**(3), 419–442.
32. Morris R.D., Descombes X., Zerubia J., The Ising/Potts model is not well suited to segmentation tasks IEEE Digital Signal Processing Workshop, Sept. 1–4 1996, Norway.
33. Nikolova M., Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. SIAM Journal on Multiscale modelling and Simulation, 2005, **4**, 3, 960–991.
34. Ortner M., Descombes X., Zerubia J., Building Outline Extraction from Digital Elevation Models using Marked Point Processes. International Journal of Computer Vision, 72(2): pages 107–132, April 2007.
35. Perez P., Markov random fields and images // CWI Quarterly, Vol. 11(4), 1998, p. 413–437.
36. Perrin G., Descombes X., Zerubia J., 2D and 3D Vegetation Resource Parameters Assessment using Marked Point Processes. – In Proc. International Conference on Pattern Recognition (ICPR), Hong-Kong, August 2006.
37. Preston C.J., Spatial birth-and-death processes. Bulletin of the International Statistical Institute, 1977, **46**, 371–391.
38. Ripley B.D., Kelly F.P., Markov point processes. Journal of the London Mathematical Society, 1977, **15**, 188–192.
39. Stoica R., Descombes X., Zerubia J., A Gibbs point process for road extraction inremotely sensed images. International Journal of Computer Vision, 2004, **57**(2), 121–136.
40. Stoyan D., Kendall W.S., Mecke J. Stochastic Geometry and its Applications, second edition. John Wiley and Sons, 1995.
41. Tjelmeland H., Besag J., Markov Random Fields with Higher-order Interactions. Scandinavian Journal of Statistics, 1998, **25**(3), 415–433.
42. Tupin F., Maitre H., Mangin J-F., Nicolas J-M., Pechersky E., Detection of Linear Features in SAR Images: Application to the Road Network Extraction. IEEE Trans. Geosci. and Remote Sensing, 1998, **36**, 2, 434–453.
43. Winkler G., Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. A Mathematical Introduction. Springer, NY, 2003.
44. Wolberg G., Pavlidis T., Restoration of binary images using stochastic relaxation with annealing. Pattern Recognition Letters, 1985, **3**, 375–388.

Підхід до перетворення образів з використанням гіббсівських полів і пов'язаної з ними динаміки

К.Декомб[1], О.Жижина [2]

[1] Аріана, Об'єднана група CNRS/INRIA/UNSA, Франція
[2] Інститут проблем перетворення інформації, Москва, Росія

Отримано 31 січня 2008 р.

Стаття містить стислий огляд застосувань полів Гіббса і відповідних динамік в теорії розпізнавання образів. Обговорюється як класична піксельна модель так і новітні просторові точкові моделі в термінах теорії полів Гіббса.

Запропоновано новий мультиоб'єктний адаптивний алгоритм для визначення об'єктів, що базується на процесі народження і знищення та його апроксимації.

Ключові слова: *гіббсівські поля, зворотні задачі в перетворенні образів, стохастична динаміка, виявлення об'єктів, марковані точкові процеси*

PACS: *05.20.Gg, 02.50.Ey, 05.10.-a, 05.90.tm*