Figure-Ground Discrimination: A Combinatorial Optimization Approach*

IEEE PAMI, Vol. 15, No. 9, pp. 899-914, September 1993

Laurent Hérault
CEA–LETI
avenue des Martyrs, 85X
38041 Grenoble

Radu Horaud
LIFIA–IMAG
46, avenue Félix Viallet
38031 Grenoble FRANCE

Abstract – *In this paper we attack the figure-ground discrimination problem from a combinatorial optimization perspective. In general the solutions proposed in the past solved this problem only partially: either the mathematical model encoding the figure-ground problem was too simple or the optimization methods that were used were not efficient enough or they could not guarantee to find the global minimum of the cost function describing the figure-ground model. The method that we devised and which is described in this paper is tailored around three main contributions.*

First, we suggest a mathematical model encoding the figure-ground discrimination problem that makes explicit a definition of shape (or figure) based on cocircularity, smoothness, proximity, and contrast. This model consists of building a cost function on the basis of image element interactions. Moreover, this cost function fits the constraints of a interacting spin *system which in turn is a well suited physical model to solve hard combinatorial optimization problems.*

Second, we suggest two combinatorial optimization methods for solving the figure-ground problem, namely (i) mean field annealing which combines mean field approximation theory and annealing and (ii) microcanonical annnealing. Mean field annealing may well be viewed as a deterministic approximation of stochastic methods such as simulated annealing. We describe in detail the theoretical bases of these methods, derive computational models, and provide practical algorithms.

Third, we provide a comparison of the efficiency of mean field annealing, simulated annealing, and microcanonical annealing algorithms. Within the framework of such a comparison, the figure-ground problem may well be viewed as a benchmark.

Index terms – low-level vision, figure-ground discrimination, feature grouping, thresholding, stochastic optimization theory, simulated annealing, mean field annealing, microcanonical annealing, recursive neural networks.

*This research has been sponsored in part by "Commissariat à l'Energie Atomique," in part by the ORASIS project, and in part by CEC through the ESPRIT-BRA 3274 (FIRST) project.

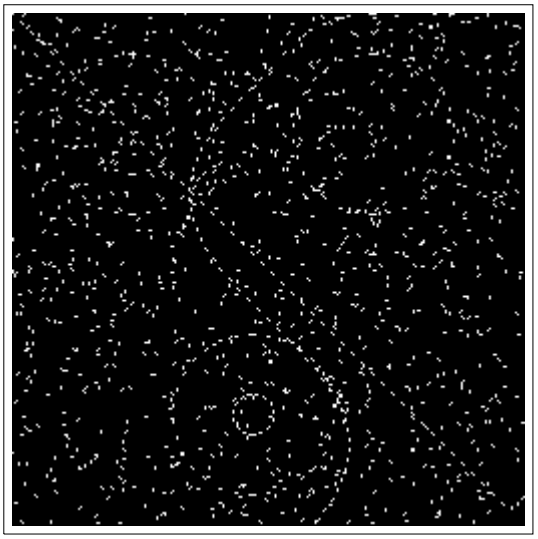


Figure 1: A synthetic image with 1250 elements. Circles, a straight line, and a sinusoid are plunged into randomly generated elements.

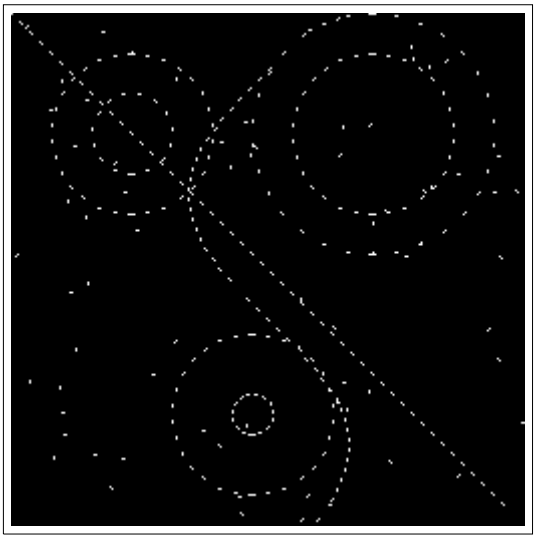


Figure 2: The result of applying mean field annealing to the synthetic image. The elements shown on this figure were labelled "shape".

I Introduction

The problem of separating figure from ground is a central one in computer vision. One aspect of this problem is the problem of separating shape from noise. Two-dimensional shapes are the input data of high-level visual processes such as recognition. In order to maintain the complexity of recognition as low as possible it is important to determine at an early level what is shape and what is noise. Therefore one needs a definition of shape, a definition of noise, and a process that takes as input *image elements* and separates them into shape and noise.

In this paper we suggest an approach whose goal is twofold: (i) it groups image elements that are likely to belong to the same (locally circular) shape while (ii) noisy image elements are eliminated. More precisely, the method that we devised builds a cost function over the entire image. This cost function sums up image element interactions and it has two terms, i.e., the first enforces the grouping of image elements into shapes and the second enforces noise elimination. Therefore the shape/noise discrimination problem becomes a combinatorial optimization problem, namely the problem of finding the global minimum for the cost function just described. In theory, the problem can be solved by any combinatorial optimization algorithm that is guaranteed to converge towards the global minimum of the cost function.

In practice, we implemented three combinatorial optimization methods: simulated annealing (SA) [15], mean field annealing (MFA) [18], [20], and microcanonical annealing (MCA) [7], [1]. The experimental results obtained with both real and simulated data tend to prove that these optimization methods are well suited for solving the figure-ground discrimination problem. The results also provide a comparison of the performances of the methods mentioned above. In all the examples the cost function converged to the same value. On a Sun Sparcstation, MFA is of the order of 5 times faster than MCA and of the order of 40 times faster than SA (see Table 1 and Table 2).

The figure-ground or shape/noise separation problem is best illustrated by an example. Figure 1

shows a synthetic image. In this image some elements belong to such shapes as circles, a straight line, and a sinusoid while some other elements are noise. Two independent sequences of random numbers were used to generate this noise: One sequence for the position in the image and another sequence for the orientation. This image has a total of 1250 elements. Figure 2 shows the result obtained with mean field annealing. This image contains 309 elements that were labelled "shape" by this algorithm. Notice that the circles, straight line, and sinusoid were correctly identified as shapes while most of the noise was thrown out.

A Background and contribution

The interest for shape/noise separation stems from Gestalt psychologists' figure-ground demonstrations [16]: certain image elements are organized to produce an emergent figure. Ever since the figure-ground discrimination problem has been seen as a side effect of feature grouping. The prerequisite of any method is the extraction from the raw data of the image elements that are either to be grouped or to be thrown out. Edge detection is in general first performed. Grouping edges is done on the basis of their connectivity or by using a clustering technique. Noise is eliminated by thresholding.

The connectivity analysis produces edge chains which are further fitted with piecewise analytic curves (lines, conics, splines, etc.). The clustering technique maps image edges into a parameter space and such a parameter space is associated with each curve type: this is the well known Hough transform. The problem of finding curves in the image is equivalent to a clustering problem in parameter space. With this technique, curves can be detected even if their constituting edges are not connected in the image. There are two major problems with the techniques just described:

1. One has to specify analytically the type of curve or curves that are to be sought in the image and this is done at a low-level of the visual process. At an early stage of an image interpretation system one may just want to know whether a set of edges form a shape or not, without any specific knowledge about the analytic curve that may fit this shape. For example, recognition of object prototypes needs such a vague description.

2. The notion of *noise* is not clearly specified with respect to the notion of shape. Hence noise elimination takes the form of adhoc methods. For example, edges associated with low contrast are usually thrown out. There are many examples that tend to prove that this is a bad heuristic. Indeed, a low contrasted shape may be present in an image in the same time as high-contrasted noise. Or, a shape may be composed of low-contrasted edges and high-contrasted ones. Simply eliminating the low-contrasted edges will have as an effect the truncation and/or elimination of shapes.

Parent & Zucker [19] define the notion of shape locally on the basis of curvature computed on a discrete grid and map the curve inference problem into a global optimization problem. The optimization itself is carried out by relaxation labelling. Finally, curve points are labelled "1" and noise points are labelled "0". We did not compare relaxation labelling with stochastic optimization and therefore we cannot assert which method is the best. However, relaxation labelling is a local optimization process which needs good initialization. The analysis of discrete curvature provided by Parent & Zucker is very interesting but the structure of the associated cost function is not well suited for such methods as mean-field annealing.

A similar approach was followed by Shashua & Ullman [25]. Curve inference takes the form of searching for the best sequence of edge elements in order to form the longest and the smoothest curves that pass through each image point. The search itself is carried out by dynamic programming which is not a global optimization technique. Therefore a global minimum cannot be guaranteed unless a good initialization is provided.

Curve/noise separation is also the concern of Gutfinger & Sklansky [12]. The coding of the problem is inspired from [19]. Curve/noise separation is then viewed as a classification problem. The classification of dots into curve-dots and noise-dots is carried out by "mixed adaptation", a method combining supervised and unsupervised training. The training stage computes statistics on noise images and on curve images. Unfortunately this a priori separation is possible only with simulated data and the method becomes impracticable when applied to real images.

The advantage of stochastic optimization over more classical approaches for solving the figure-ground discrimination problem was stressed by Sejnowski & Hinton [24]. They introduce a figure-ground model based on two possible labels for the image elements: region and edge. Starting with a random labelling, a gradient-descent procedure gets trapped in one of the many local minima of the energy landscape while simulated annealing converges to a solution where the region elements are "inside" the edge elements. The experimental results shown by the authors deal only with synthetic data.

The method proposed by Carnevali, Coletti, & Patarnello [6] uses simulated annealing and a simple pixel interaction model in order to classify the pixels of a binary image into object and noise. The pixel interaction model they propose uses pixel proximity on the premise that objects are dense sets of pixels while noise is formed of a sparsely distributed set of pixels.

The method proposed by Peterson [21] for tracking particles in high energy physics may be applied to the grouping problem. Peterson suggests a combinatorial optimization formulation based on mean field theory. However the model proposed by Peterson for track finding is not well suited when noise is present in the data. Moreover, Peterson's model maps n points onto n^2 variables and hence there are n^4 connections. Such a model is well suited when the number of variables is relatively small. Our model maps n points onto n variables and therefore only n^2 connections are necessary. More interesting, a VLSI implementation of the mean field theory is suggested in Peterson's paper.

Recently, Kass, Witkin, and Terzopoulos [14] proposed the snake model: it is an energy-minimizing spline guided by external forces and images forces that pull the snake towards image edges. Nevertheless, snakes do not try to solve the entire problem of finding salient image contours. They rely on other mechanisms to place them nearby the desired contours.

The work described in this paper has the following contributions:

- We suggest a mathematical encoding of the figure-ground discrimination problem that consists of separating shape from noise using combinatorial optimization methods. The particular cost (or energy) function that we devised fits the constraints of a interacting spin system – a physical model well suited for solving hard combinatorial optimization problems.

- We suggest two combinatorial optimization methods for solving the figure-ground problem, namely (i) mean field annealing which combines mean field approximation theory and annealing and (ii) microcanonical annnealing. Mean field annealing may well be viewed as a deterministic approximation of stochastic methods such as simulated annealing.

- We provide a comparison of the efficiency of three optimization methods: two stochastic methods (simulated annealing and microcanonical annealing) and a deterministic one (mean field annealing). The figure-ground problem may well be viewed as a benchmark for this comparison. In the light of this comparison it appears that microcanonical annealing, as initially proposed by Barnard [1] is inefficient. Therefore we propose a substantial improvement of the annealing schedule associated with this algorithm.

In the past, the theoretical bases of mean field annealing were studied by Orland [18] and Peterson [20]. Mean field annealing has been used by Hérault & Niez to solve NP-complete graph problems [13], by Geiger & Girosi to solve the reconstruction problem [9], by Geiger & Yuille to solve the image segmentation problem [10], and by Zerubia & Chellappa to solve the edge detection problem [28].

The theoretical bases of microcanonical annealing were proposed by Creutz in 1983 [7] and it has been applied to vision for the first time by Barnard for solving the stereo matching problem [1].

Blake [2] provides a comparison of deterministic and stochastic optimization methods, namely graduated nonconvexity (GNC) [3] and simulated annealing. The benchmark for the comparison is the reconstruction of 1-D signals. Blake's conclusion is that the deterministic approach performs better. We extend Blake's comparison to mean field annealing and microcanonical annealing.

B Organization

The remainder of the paper is organized as follows. Section II briefly describes the mathematical structure of a class of cost functions that are well suited for stochastic optimization methods. This structure places strong constraints on the mathematical coding of the figure-ground discrimination problem. Next we formulate our problem in terms of such a cost function. This function involves interactions between the image elements that are considered. These interactions are made clear in Section III. Section IV describes in detail the physical basis of stochastic optimization and introduces three optimization methods. These methods are compared in terms of their efficiencies, that is, their performances on a sequential machine and their possible degrees of paralellisation. Section V shows the experimental results obtained with both synthetic and real images. Section VI contains a general discussion on figure-ground discrimination and gives some directions for future work.

II A combinatorial optimization formulation

We consider a particular class of combinatorial optimization problems for which the cost function has a mathematical structure that is analog to the global energy of a complex physical system, that is a *interacting spin* system. First, we briefly describe the state of such a physical system and give the mathematical expression of its energy. We also show the analogy with the energy of a recursive neural network. Second we suggest that the figure-ground discrimination problem can be cast into a global optimization problem of the type mentioned above.

The state of an interacting spin system is defined by:

1. A spin state-vector of N elements $\vec{\sigma} = [\sigma_1, \cdots, \sigma_N]$ whose components are described by discrete labels which correspond to *up* or *down* spins: $\sigma_i \in \{-1, +1\}$. Each vector $\vec{\sigma}$ is associated to a potential solution of the problem. Its components σ_i may well be viewed as the outputs of binary neurons.

2. A symmetric matrix J describing the interactions between the spins. An element J_{ij} of this matrix represents the interaction between the spins σ_i and σ_j. For all i, $J_{ii} = 0$. These interactions may well be viewed as the synaptic weights between neurons in a network.

3. A vector $\vec{\delta} = [\delta_1, \cdots, \delta_N]$ describing an external field in which the spins are plunged. δ_i is the value of this field viewed by the spin i (or the external input on neuron i).

Therefore, the interacting spin system has a "natural" neural network encoding associated with it which describes the microscopic behaviour of the system. A macroscopic description is given by the energy function which evaluates each spin configuration. This energy is given by:

$$E(\sigma_1, \sigma_2, \ldots, \sigma_N) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} \sigma_i \sigma_j - \sum_{i=1}^{N} \delta_i \sigma_i \tag{1}$$

In Section IV we will describe a few methods for approaching as near as possible the absolute minimum of the energy described by the equation above. Let us first formulate the figure-ground discrimination problem in terms of a cost function that has a similar structure, i.e., equation (1).

We consider N image elements. Each such element has a label associated with it, p_i, which can take two values: 0 or 1. The set of N labels forms the state vector $\vec{p} = [p_1, \cdots, p_N]$. We seek a state vector such that the "shape" elements have a label equal to 1 and the "noise" elements have a label equal to 0. If c_{ij} designates an interaction between elements i and j (this interaction will be specified in the next Section), one may write by analogy with physics an interaction energy:

$$E_{saliency}(\vec{p}) = -\sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} p_i p_j$$

Obviously, the expression above is minimized when all the labels are equal to 1. In order to avoid this trivial solution we introduce the constraint that some of the elements in the image are not significant and therefore should be labelled "noise":

$$E_{constraint}(\vec{p}) = \left(\sum_{i=1}^{N} p_i \right)^2$$

The function to be minimized could be something like the sum of these energies:

$$E(\vec{p}) = E_{saliency}(\vec{p}) + \alpha \; E_{constraint}(\vec{p}) \tag{2}$$

In this expression α is a positive real parameter that has to be adjusted and is closely related to the signal-to-noise ratio. With the substitution $p_i = (\sigma_i + 1)/2$, this formula can be written such that the analogy with the interacting spin system becomes obvious (compare with equation (1)):

$$E(\vec{\sigma}) = C - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{2} (c_{ij} - \alpha) \sigma_i \sigma_j \; - \sum_{i=1}^{N} \frac{1}{2} \left(-N\alpha + \sum_{j=1}^{N} c_{ij} \right) \sigma_i \tag{3}$$

C being the constant term:

$$C = \frac{1}{4}\left(-\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij} + \alpha N^2\right)$$

Minimizing equation (3) is therefore strictly equivalent to minimizing equation (1) with the following interaction coefficients and external fields:

$$J_{ij} = \frac{c_{ij} - \alpha}{2}$$

$$\delta_i = \frac{\sum_{j=1}^{N} c_{ij} - N\alpha}{2}$$

III Computing image interactions

An image array contains two types of information: changes in intensity and local geometry. Therefore the choice of the *image elements* mentioned so far is crucial. Edge elements, or edgels are the natural candidates for making explicit the two pieces of information just mentioned.

An edgel can be obtained by one of the many edge detectors now available. An edgel i is characterized by its position in the image (x_i, y_i) and by its gradient computed once the image has been low-pass filtered. The x and y components of the gradient vector are:

$$g_x(x_i, y_i) = \frac{\partial I_f(x_i, y_i)}{\partial x}$$

$$g_y(x_i, y_i) = \frac{\partial I_f(x_i, y_i)}{\partial y}$$

I_f is the low-pass filtered image. From the gradient vector one can easily compute the gradient direction and magnitude. The direction θ_i of the edgel is perpendicular to the gradient direction. It is also the direction of the tangent to the curve that may pass through this edgel, e.g., Figure 3. The magnitude of the gradient g_i is proportional to the height of the intensity change at the edgel location:

$$\theta_i = \arctan\left(\frac{g_y(x_i, y_i)}{g_x(x_i, y_i)}\right) + \frac{\pi}{2}$$

$$g_i = \left(g_x^2(x_i, y_i) + g_y^2(x_i, y_i)\right)^{1/2}$$

Let A and B be two edgels. We want that the interaction between these two edgels encapsulates the concept of shape. That is, if A and B belong to the same shape then their *interaction* is high. Otherwise their interaction is low. Notice that a weak interaction between two edgels has several interpretations: (i) A belongs to one shape and B belongs to another one, (ii) A belongs to a shape and B is noise, or (iii) both A and B are noise. The interaction coefficient must therefore be a co-shapeness measure. In our approach, co-shapeness is defined by a combination of *cocircularity, smoothness, proximity*, and *contrast*.

The definition of cocircularity is derived from [19] and it constrains the shapes to be as circular as possible, or as a special case, as linear as possible. Smoothness enforces shapes with low curvature. Proximity restricts the interaction to occur in between nearby edgels. As a consequence,

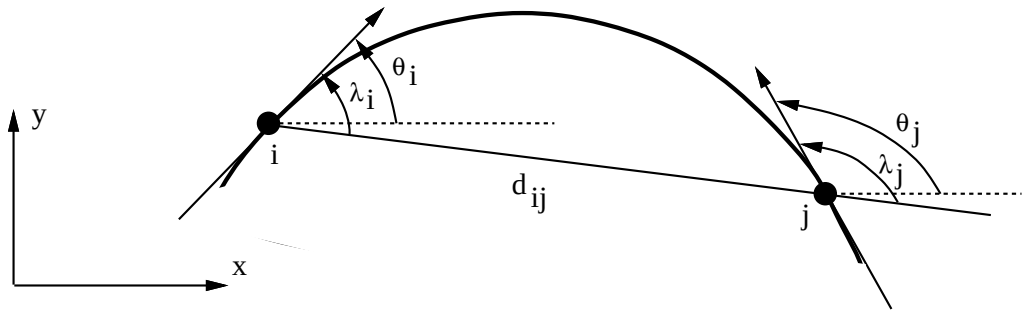


Figure 3: The definition of cocircularity between two edgels i and j.

cocircularity and smoothness are constrained to be local shape properties. The combination of cocircularity, smoothness, and proximity will therefore allow a large variety of shapes that are circular and smooth only locally. Contrast enforces edgels with a high gradient module to have a higher interaction coefficient than edgels with a low gradient module.

Cocircularity. Following [19] and from Figure 3 it is clear that two edgels belong to the same circle if and only if:

$$\lambda_i + \lambda_j = \pi \tag{4}$$

In this formula, λ_i is the angle made by one edgel with the line joining the two edgels. Notice that a circle is uniquely defined if the relative orientations of the two edgels verify the equation above. This equation is also a local symmetry condition consistent with the definition of local symmetry of Brady & Asada [4]. Moreover, linearity appears as a special case of cocircularity, namely when $\lambda_i = 0$ and $\lambda_j = \pi$ or when $\lambda_i = \pi$ and $\lambda_j = 0$.

From this cocircularity constraint we may derive a weaker constraint which will measure the similarity between a two-edgel-configuration and a circular shape:

$$\Delta_{ij} = \mid \lambda_i + \lambda_j - \pi \mid .$$

Δ_{ij} will vary between 0 (a perfect circle) and π (no shape). Finally the cocircularity coefficient is constrained to vary between 1 for a circle and 0 for noise and is defined by the formula:

$$c_{ij}^{cocir} = \left(1 - \frac{\Delta_{ij}^2}{\pi^2}\right) \exp\left(-\frac{\Delta_{ij}^2}{k}\right) \tag{5}$$

The parameter k is chosen such that the cocircularity coefficient vanishes gently for non-circular interactions.

Smoothness. Consider an edgel and two nearby edgels. From Figure 4 it is clear that the cocircularity coefficient between edgels i and j is the same as the cocircularity coefficient between i and k. Indeed there are two circles, one passing through i and j and the other passing trhough i and k. One would like to give some preference to one of these configurations. If smooth curves (rather than rapidly turning curves) are prefered then the $i-j$ interaction should be stronger than the $i-k$ interaction. With the same definition for λ as above we define a smoothness coefficient

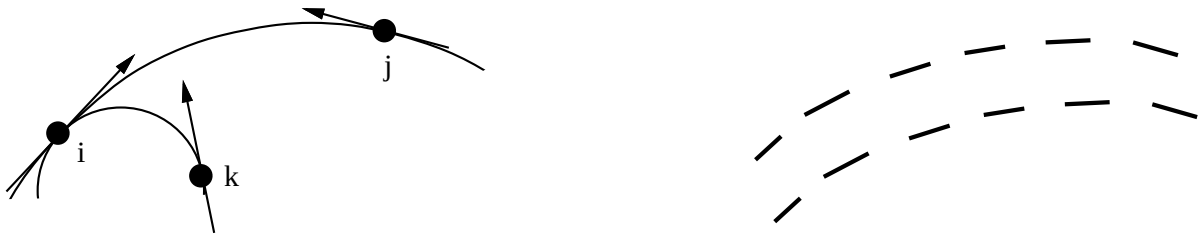


Figure 4: In the absence of the smoothness constraint, the interaction between i and j is the same as the interaction between i and k (left). The smoothness constraint reduces the interaction in between parallel curves (right).

between two image edgels. This coefficient varies between 0 (a sharp corner) and 1 (a straight line):

$$c_{ij}^{smooth} = \left(1 - \frac{\lambda_i(\pi - \lambda_i)}{\pi^2}\right)\left(1 - \frac{\lambda_j(\pi - \lambda_j)}{\pi^2}\right) \tag{6}$$

Obviously, for the example of Figure 4-left we obtain:

$$c_{ij}^{smooth} > c_{ik}^{smooth}$$

Figure 4-right illustrates another advantage of using a coefficient that enforces smoothness. Indeed, the absence of the smoothness constraint, edgel interactions occur between parallel curves: an edgel equally interacts with its longitudinal neighbours and with its *lateral* neighbours. The smoothness constraint reduces the lateral interactions occuring between such parallel curves.

Proximity. The surrounding world is not constituted only by circular shapes. Cocircularity must therefore be a local property. That is, the class of shapes we are interested to detect at a given scale of resolution are shapes that can be approximated by a sequence of smoothly connected circular arcs and straight lines. The proximity constraint is best described by a coefficient that vanishes smoothly as the two edgels are farther away from each other:

$$c_{ij}^{prox} = \exp\left(-\frac{d_{ij}^2}{2\,\sigma_d^2}\right) \tag{7}$$

where d_{ij} is the distance between the two edgels and σ_d is a fraction of the standard deviation of all these distances over the image. Hence, the edgel interaction will adjust itself to the image distribution of the edgel population.

Contrast. A classical approach to figure-ground discrimination is to compare the gradient value at an edgel against a threshold and to eliminate those edgels that fall under this threshold. An improvement to this simply-minded nonlinear filtering is to consider two thresholds such that edgel connectivity is better preserved [5]. Following the same idea, selection of shapes with high contrast

can be enforced by multiplying the interaction coefficient with a term whose value depends on contrast:

$$c_{ij}^{contrast} = \frac{g_i g_j}{g_{max}^2} \tag{8}$$

g_{max} is the highest gradient value over the edgel population.

Finally the interaction coefficient between two edgels becomes:

$$c_{ij} = c_{ij}^{cocir}\ c_{ij}^{smooth}\ c_{ij}^{prox}\ c_{ij}^{contrast} \tag{9}$$

IV Stochastic optimization

The states reachable by the system described in Section II (equation (3)) correspond to the vertices of a N-dimensional hypercube. We are looking for the state which corresponds to the absolute minimum of the energy function. Typically, when $N = 1000$, the number of possible configurations is $2^N \approx 10^{301}$. The problem of finding the absolute minimum is complex because of the large number of local minima of the energy function and hence this problem cannot be tackled with local minimization methods (unless a good initialization is available).

We already mentioned that the functional to be minimized has the same structure as the global energy of an interacting spin system. To find a near ground state of such a physical system we will use statistical methods. Two analyses are possible depending on the interaction of the system with its environment: either the system can exchange heat with its environment (case of the *canonical* analysis) or the system is isolated (case of the *microcanonical* analysis).

A Canonical analysis

This analysis makes the hypothesis that the physical system can exchange heat with its environment. At the equilibrium, statistical thermodynamics shows that the free energy F is minimized. The free energy is given by: $F = E - TS$, where E is the internal energy (the energy associated with the optimization problem) and S is the entropy (which measures the internal disorder). Hence, there is a competition between E and S. The equilibrium can be reached at any temperature T and hence F has a minimum at each temperature. At low temperatures, configurations with small entropy are predominant. On the contrary, at high temperatures, the system is in a configuration with a high entropy [23]. As a consequence at high temperatures and at equilibrium, since F is minimal, E must be high. At low temperatures and at equilibrium, F is minimal and TS is close to zero. Therefore, the internal energy E is minimized at low temperatures. However the minimum of E depends on how the temperature parameter decreases towards the absolute zero. It was shown that annealing is a very good way to decrease the temperature.

We are interested in physical systems for which the internal energy is given by equation (1). The remarks above are expressed in the most fundamental result of statistical physics, the Boltzmann (or Gibbs) distribution:

$$Pr(E(\vec{\sigma}) = E_i) = \frac{\exp\left(-E_i/(kT)\right)}{Z(T)}. \tag{10}$$

which gives the probability of finding a system in a state i with the energy E_i, assuming that the system is at equilibrium with a large heat bath at temperature $k\,T$ (k is the Boltzmann's constant). $Z(T)$ is called the *partition function* and is a normalization factor:

$$Z(T) = \sum_n \exp\left(-\frac{E_n}{kT}\right). \tag{11}$$

This sum runs over all possible spin configurations. The Boltzmann distribution expresses the fact that configurations with low internal energy are predominant at low temperatures, i.e, their probability is close to 1. Using equation (10) one can compute at a given temperature T the mean value over all possible configurations of some macroscopic physical parameter A:

$$\begin{aligned} \langle A \rangle &= \sum_n A_n \, Pr(E(\vec{\sigma}) = E_n) \\ &= \frac{\sum_n A_n \, \exp\left(-E_n/kT\right)}{Z(T)}. \end{aligned} \tag{12}$$

Unfortunately, the partition function $Z(T)$ is usually impossible to compute. Simulated annealing is a good way to avoid this computation in the general case. Nevertheless, when the system is described by equation (1) one can use equation (12) which, with an additionnal hypothesis, is a basic equation for mean field approximation and for mean field annealing.

A.1 Simulated annealing (SA)

In 1953, Metropolis and al. [17] described a Monte Carlo algorithm that generates at a given temperature a sequence of states which converges to the Boltzmann distribution at the limit. Hence, the system reaches at the limit a thermodynamic equilibrium at temperature T. For a given temperature T, this algorithm begins in an arbitrary state and successively generates candidate state transitions at random. Each such elementary transition has associated with it a change ΔE in the global energy of the system. An elementary transition is accepted with the following probability:

$$\begin{cases} 1 & \text{if } \Delta E < 0 \\ \exp(-\Delta E/T) & \text{otherwise.} \end{cases} \tag{13}$$

Let us define an elementary transition to be performed to solve our problem. A good way to define such a transition consists of picking up at random an edgel i_0 and then changing the output of the associated neuron, i.e., changing its labe. Before the transition, the state of the system was the following:

$$\vec{\sigma}_{before} = \left[\begin{array}{ccccccc} \sigma_1 & \cdots & \sigma_{i_0-1} & \sigma_{i_0} & \sigma_{i_0+1} & \cdots & \sigma_N \end{array} \right].$$

After the transition, σ_{i_0} is replaced by $-\sigma_{i_0}$. The system is now in the state:

$$\vec{\sigma}_{after} = \left[\begin{array}{ccccccc} \sigma_1 & \cdots & \sigma_{i_0-1} & -\sigma_{i_0} & \sigma_{i_0+1} & \cdots & \sigma_N \end{array} \right].$$

The number of possible transitions is N.

Let us calculate the energy variation associated with an elementary transition. The useful part of the energy in the calculation of the energy variation is (from equation (3)):

$$\sigma_{i_0} \sum_{j=1, j\neq i_0}^{N} (\lambda - c_{i_0 j}) \, \sigma_j + \left(N \, \lambda - \sum_{j=1}^{N} c_{i_0 j} \right) \, \sigma_{i_0} + \frac{\lambda}{2}.$$

The energy variation associated with an elementary transition is given by:

$$\Delta E_{\sigma_{i_0} \to -\sigma_{i_0}} = -2\,\sigma_{i_0} \left[\sum_{j=1,\, j \neq i_0}^{N} (\lambda - c_{i_0 j})\ \sigma_j + \left(N\,\lambda - \sum_{j=1}^{N} c_{i_0 j} \right) \right]. \tag{14}$$

Kirkpatrick [15] recognized a connection between the Metropolis technique and combinatorial optimization problems. In fact a good way to find low energy states of a complex physical system such as a spin system is to heat up the system to some high temperature and then to cool it down slowly. This *annealing* process forces the system's evolution into states with low energy while the system does not get trapped in local minima. The idea of simulated annealing is to express those concepts in terms of a numerical algorithm.

The algorithm starts at a high initial temperature. At this temperature, nearly 100% of the tested transitions are accepted. The Metropolis procedure is applied at constant temperature in order to reach a near thermal equilibrium state at this temperature. Then the temperature is slightly decreased and the Metropolis procedure is applied again. This process is repeated until the system is frozen, i.e., when the energy cannot significantly be reduced anymore. There are results by Geman & Geman [11] showing the existence of annealing schedules that guarantee to the system to reach ground states (global energy minima) in finite time. Unfortunately, these schedules are not usable in practice. Nevertheless, we use a geometric schedule.

The simulated annealing algorithm is given in Appendix A.

A.2 Mean field annealing

In order to introduce the mean field annealing algorithm, we first introduce the somehow more classical mean field approximation method: it is a simple analytic approximation of the behaviour of interacting spin systems in thermal equilibrium. We start by developing equation (1) around σ_i:

$$E(\sigma_1, \sigma_2, \ldots, \sigma_N) = \Phi_i\,\sigma_i - \frac{1}{2} \sum_{k=1, k \neq i}^{N} \sum_{j=1, j \neq i}^{N} J_{kj} \sigma_k \sigma_j - \sum_{k=1, k \neq i}^{N} \delta_k \sigma_k.$$

Φ_i is the total field which affects the spin σ_i:

$$\Phi_i = -\left(\sum_{j=1}^{N} J_{ij}\,\sigma_j + \delta_i \right).$$

The *mean field* $\langle \Phi_i \rangle$ affecting σ_i is computed from the sum of the fields created on the spin σ_i by all the other spins "frozen" in their mean states, and from the external field δ_i viewed by the spin σ_i. The mean state of a spin $\langle \sigma_i \rangle$ is the mean value of the σ_i's computed over all possible states that may occur at the thermal equilibrium. We obtain:

$$\langle \Phi_i \rangle = -\left(\sum_{j=1}^{N} J_{ij}\,\langle \sigma_j \rangle + \delta_i \right). \tag{15}$$

We introduce now the following approximation [26]: The system composed of N interacting spins is viewed as the union of N systems each composed of a single-spin. Such a single-spin

system $\{\sigma_i\}$ is subject to the mean field $\langle\Phi_i\rangle$ created by all the other single-spin systems. Let us study such a single spin system. It has two possible states: $\{-1\}$ or $\{+1\}$. The probability for the system to be in one of these states is given by the Boltzmann distribution law, equation (10):

$$P(X_i = \sigma_i^0) \approx \frac{\exp\left(-\langle\Phi_i\rangle\,\sigma_i^0/T\right)}{P(X_i = 1) + P(X_i = -1)},\ \sigma_i^0\ \in\ \{-1, 1\}. \tag{16}$$

X_i is the random variable associated with the value of the spin state. Notice that in the case of a single-spin system the partition function (the denominator of the expression above) has a very simple analytical expression. By combining equations (12), (15), and (16), the mean state of σ_i can now be easily derived:

$$\begin{aligned}
\langle\sigma_i\rangle &\approx \frac{(+1)\,\exp\left(-\langle\Phi_i\rangle/T\right) + (-1)\,\exp\left(\langle\Phi_i\rangle/T\right)}{\exp\left(\langle\Phi_i\rangle/T\right) + \exp\left(-\langle\Phi_i\rangle/T\right)} \\
&= \tanh\left(\frac{-\langle\Phi_i\rangle}{T}\right) \\
&= \tanh\left(\frac{\sum_{j=1}^{N} J_{ij}\,\langle\sigma_j\rangle + \delta_i}{T}\right).
\end{aligned} \tag{17}$$

We consider now the whole set of single-spin systems. We therefore have N equations of the form:

$$\mu_i = \tanh\left(\frac{\sum_{j=1}^{N} J_{ij}\,\mu_j\ +\ \delta_i}{T}\right). \tag{18}$$

where $\mu_i = \langle\sigma_i\rangle$. It was shown by Peterson & Anderson [22] that the final value of $1/2(\mu_i + 1)$ approximates the probability that the corresponding spin i has a value $\sigma_i = 1$ at thermal equilibrium as reached by the Metropolis procedure. The problem of finding the mean state of the spin system at thermal equilibrium is now mapped into the problem of solving a system of N coupled non-linear equations, i.e., equation (18). In the general case, an analytic solution is rather difficult to obtain. Instead, the solution for the vector $\vec{\mu} = [\mu_1, \cdots, \mu_N]$ may well be the *stationary solution* of the following system of N differential equations:

$$\tau\,\frac{d\mu_i}{dt} = \tanh\left(\frac{\sum_{j=1}^{N} J_{ij}\,\mu_j\ +\ \delta_i}{T}\right) - \mu_i. \tag{19}$$

where τ is a time constant introduced for homogeneity. In the discrete case the temporal derivative term can be written as:

$$\left.\frac{d\mu_i}{dt}\right)_{t_n} = \frac{\mu_i^{n+1} - \mu_i^n}{\Delta t} + o(\Delta t). \tag{20}$$

where μ_i^n is the value of μ_i at time t_n. By substituting in equation (19), we obtain:

$$\tau\,\frac{\mu_i^{n+1} - \mu_i^n}{\Delta t} = \tanh\left(\frac{\sum_{j=1}^{N} J_{ij}\,\mu_j^n\ +\ \delta_i}{T}\right) - \mu_i^n. \tag{21}$$

Hence, by choosing $\tau = \Delta t$, we obtain an iterative solution for the system of differential equations described by equation (19):

$$\forall\, i \in \{1, \ldots, N\},\ \mu_i^{n+1} = \tanh\left(\frac{\sum_{j=1}^{N} J_{ij}\,\mu_j^n\ +\ \delta_i}{T}\right),\ n \geq 1. \tag{22}$$

where μ_i^n is an estimation of $\langle \sigma_i \rangle$ at time t_n or at the n'th iteration.

Starting with an initial solution $\vec{\mu^0} = [\mu_1^0, \cdots, \mu_N^0]$, the convergence is reached at the n'th iteration such that $\vec{\mu^n} = [\mu_1^n, \cdots, \mu_N^n]$ becomes stationary. The physicists would say that the thermal equilibrium has been nearly reached.

Two convergence modes are possible:

- *Synchronous mode.* At each step of the iterative process **all** the μ_i^n's are updated simultanously using the μ_i^{n-1}'s previously calculated.
- *Asynchronous mode.* At each step of the iterative process a unique spin μ_i^n is randomly selected and updated using the μ_i^{n-1}'s.

In practice, the asynchronous mode produces better results because the convergence process is less subject to oscillations frequently encountered in synchronous mode. In order to obtain a solution for the vector $\vec{\sigma}$ from the vector $\vec{\mu}$, one simply looks at the signs of the μ_i^n's. A positive sign implies that the probability that the corresponding spin has a value of $+1$ is greater than 0.5:

$$\begin{cases} \text{if } 1/2(1+\mu_i^n) > 0.5 \text{ then} & \sigma_i = +1 \\ \text{else} & \sigma_i = -1. \end{cases}$$

A practical difficulty with mean field approximation is the choice of the temperature T at which the iterative process must occur. To avoid such a choice one of us [13] and other authors [27] have proposed to combine the mean field approximation process with an annealing process giving rise to mean field annealing. Hence, rather than fixing the temperature, the temperature is decreased during the convergence process: thus the μ_i's tend to $+1$ or -1 as the system converges.

A.3 Estimating the MFA parameters and good initial conditions

The MFA equations contain two parameters to be determined: the initial temperature T_{init} and the decreasing factor of the temperature between two steps, $decT$. Moreover, the initial values of the μ_i^n's should be independent of the image. The aim is to avoid a trial-and-error process: the algorithm should be a black box from a user's perspective.

The initial temperature. The spin system onto which we have mapped the figure/ground discrimination problem typically has two phases; at high enough temperatures, according to the Boltzmann distribution, all the states reachable by the system are equally likely. As the temperature is lowered down, a phase transition occurs at $T = T_c$ and as $T \to 0$ the μ_i values represent a specific decision made as to the solution to the problem. At convergence, these values verify:

$$\frac{1}{N} \sum_{i=1}^{N} \mid \mu_i \mid = 1 \tag{23}$$

This is illustrated in Figure 11. Untill now, we don't know how to calculate the critical temperature of an interacting spin system when the external field is not null (i.e., when there is at least one δ_i that is not null). Practically, we use random asynchronous dynamics (which is the dynamics the most similar to the behaviour of a spin system). We start with a high enough initial temperature determined by the μ_i distribution: Each μ_i has to be around 0 at the beginning and during several iterations; otherwise, the initial temperature is increased.

Annealing schedule. As already mentioned, two different annealing schedules are possible:

- Van den Bout & Miller [27] tried to estimate theoretically the critical temperature and then they perform two sets of iterations: one iteration process at this critical temperature until a near equilibrium state is reached and another iteration process at a temperature value that is close to 0. However, the critical temperature is quite difficult to estimate and the behaviour of the spin system predicted by their analytical approximation is not in accordance with the computational experiments.

- Instead of estimating the critical temperature, we prefer the following schedule. Initially the temperature has a high value and as soon as every spin has been updated at least once, the temperature is decreased to a smaller value. Then the temperature continues to decrease at each step of the convergence process. This does not guarantee that a near equilibrium state is reached at each temperature value but when the temperature is small enough then the system is frozen in a good stable state. Consequently, the convergence time is reduced since at low temperatures the convergence to a stationary solution is accelerated. One of us has successfully used this strategy to solve hard, np-complete graph combinatorial problems [13].

The initial values of the μ_i's. When the vector $\vec{\delta}$ (i.e., the external field) is the null vector, one can start with a vector $\vec{\mu^0}$ close to the obvious unstable solution $[0, \cdots, 0]$. In practice, when the vector $\vec{\delta}$ is not null one starts with an initial configuration obtained by adding noise to $[0, \cdots, 0]$. For instance the μ_i^0's are chosen randomly in the interval $[-10^{-5}, +10^{-5}]$. This initial state is plausible for a spin system in a heat bath at high temperature. In fact the spin values (+1 or –1) are equally likely at high temperatures and hence, $\langle \sigma_i \rangle = 0$ for all spin i (see equation (10)). During the iterative process, the μ_i's converge to values in between -1 or $+1$.

The mean field annealing algorithm is outlined in Appendix B.

B Microcanonical Analysis

This analysis makes the hypothesis that the physical system is isolated from its environment. Then the main property of such a system is that its total energy is constant whatever the dynamic evolution is. In practice the total energy E_{total} lies in the range $[E_0 - \delta E, E_0 + \delta E]$. From statistical physics we know that in an equilibrium situation as described above, all the states are equally likely [23]. The probability of finding a system in state i with total energy E_i is given by:

$$Pr(E_{total} = E_i) = \begin{cases} B & \text{if } E_0 - \delta E \leq E_i \leq E_0 + \delta E \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

where B is a constant which can be determined by the normalization condition that the sum of the probabilities over all accessible states is equal to 1. Hence the entropy of the system, which measures uncertainty on the final state is maximized at equilibrium.

The total energy is the sum of a potential energy E_p which encodes the energy associated with the optimization problem (equation (1)) and of a kinetic energy E_c which has always a positive value:

$$E_{total} = E_p + E_c.$$

Standard statistical mechanics arguments show that at equilibrium the probability of finding a system in a state with a kinetic energy E is given by a Boltzmann distribution law:

$$Pr(E_c = E) = \frac{\exp\left(-E/(kT)\right)}{Z}. \tag{25}$$

where Z is a normalization factor. In this expression, the temperature expresses the internal turbulence of the system. From this distribution, one derives a fundamental result of the microcanonical analysis: it can be shown that the temperature is directly related to the mean value of the kinetic energy $< E_c >$ over time [7]:

$$\frac{1}{kT} = \frac{1}{4} \ln\left(1 + \frac{4}{< E_c >}\right). \tag{26}$$

This can be simplified by supposing that E_c can take any positive value. In fact, from equation (25), one computes $< E_c >$:

$$\begin{aligned} < E_c > &= \frac{\int_0^\infty E \exp\left(-E/(kT)\right) dE}{\int_0^\infty \exp\left(-E/(kT)\right) dE} \\ &= kT. \end{aligned} \tag{27}$$

Hence, the mean value of the kinetic energy over time $\langle E_c \rangle$ has the same role as the temperature parameter in the canonical analysis.

B.1 Microcanonical annealing

In 1983 Creutz [7] described a Monte Carlo algorithm that generates a sequence of states which converges at the limit to an equally likely distribution of the total energy. This is analogous to the Metropolis algorithm applied to the microcanonical analysis.

The principle of the algorithm is the following: an extra degree of freedom is added to the spin system which corresponds to the kinetic energy E_c. The algorithm starts with an arbitrary state of the spin system (described by its potential energy E_p given by eq. (1)) and a high value for E_c. The kinetic energy is in fact a parameter which plays the role of the thermal noise in the canonical case. State transitions are randomly generated and they are identical to the ones generated in simulated annealing. Each transition has associated with it a change in the potential energy ΔE_p. The transition is accepted according to the following criterion: if $\Delta E_p \leq 0$ then the transition is accepted and one adds to the kinetic energy the quantity $-\Delta E_p$ in order to maintain the total energy unchanged. If $\Delta E_p > 0$ then the transition is accepted only if there is enough kinetic energy available, that is, if $E_c > \Delta E_p$. In this case the kinetic energy is decreased by ΔE_p in order, once again, to maintain the total energy unchanged.

Creutz called the extra degree of freedom corresponding to the kinetic energy a *demon*. It is defined by a sack which contains energy. This demon randomly travels around the system, i.e., the lattice of spins. Figure 5 shows an example of the assignment of the demon to the spins. Upon reaching a spin, the demon attempts to flip the corresponding spin (from $+1$ to -1 or vice versa). It succeeds only if it lowers the potential energy or if it has enough energy in its sack.

This algorithm maximizes the entropy of the system at the limit, which corresponds to the thermodynamical equilibrium. Notice that while it is difficult to estimate when an equilibrium is actually reached with the Metropolis algorithm, this is relatively simple in microcanonical analysis.

The kinetic energy being distributed at equilibrium according to equation (25), one may notice that the, at equilibrium, standard deviation $\rho(E_c)$ equals $< E_c >$. Thus, at equilibrium we have:

$$\frac{< E_c >}{\rho(E_c)} = 1. \tag{28}$$

In the basic version of the Creutz algorithm, each decision to accept or reject a state transition depends on the current value of E_c and therefore, on the previous decision. Moreover, the temperature T (eq. (27)) is measured using the distribution of E_c over time. We will see that this basic version is not convenient to proceed an annealing.

We currently use a parallel version of the algorithm by using a large number of demons. Collectively, they carry an appreciable amount of energy. Theoretically, it can be shown that when the number of demons becomes large compared with the number of spins, the Creutz algorithm reduces to that of Metropolis [7]. Each demon i has associated with it a part E_c^i of the total kinetic energy E_c. In our model, we use the same number of demons as the number of spins in the system, and at any time each demon is associated with a single spin. Figure 6 shows an exemple of the assignment of demons to the spins when $N = 6$. The amount of kinetic energy carried by the set of demons is defined by:

$$E_c = \sum_{i=1}^{N} E_c^i. \tag{29}$$

At the beginning of the parallel algorithm, the demons are evenly distributed on the lattice of spins, each with the same high kinetic energy. Then, in synchronous or in asynchronous mode, each demon will try to flip its associated spin. After every lattice update[1], we perform a complete random permutation of the demons (see Figure 6 for an example).

A great advantage of this parallel version of the Creutz algorithm is that, because of the ergodicity hypothesis (nonergodic behavior would represent a fascinating exception to the generic case), statistics can be sampled over space rather than time. Hence, the mean energy of a demon over time is equivalent to the mean energy over the set of demons at a fixed time. In the algorithm, after each lattice update, we can compute kT and $\rho(E_c)$ and compare them in order to find out whether the equilibrium has been reached (the equilibrium being described by equation (28)) or not. We have:

$$kT =< E_c >_{all\,the\,demons} = \frac{1}{N} \sum_{i=1}^{N} E_c^i. \tag{30}$$

and

$$\rho(E_c) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (E_c^i - kT)^2}. \tag{31}$$

Microcanonical annealing, as defined for the first time by Barnard [1], roughly replaces the Metropolis method with the Creutz method in the simulated annealing algorithm. Instead of explicitely reducing the temperature as in simulated annealing, Barnard reduces the kinetic energy and hence, the total energy. Thus the constant energy surface in the state space on which evolves the state of the system shrinks to a set of states of low energy.

The annealing schedule associated with the kinetic energy, as proposed by Barnard is linear: he suggested to reduce the energy of each demon by three units after the lattice of spins has been

[1] That is, once each individual spin in the lattice has been updated once

updated once. There is no theoretical reason to make such a linear choice. This linear schedule has been proved to be inefficient in the canonical analysis [11]. Moreover, updating the lattice of spins only once at each temperature step unables the system to reach an equilibrium configuration at every temperature step. This explains why we obtained unsatisfactory experimentals results with the annealing schedule proposed by Barnard: the system converges to a local minimum of high energy. This phenomenon was not visible in Barnard's implementation because of his use of a multiscale representation of images which reduces dramatically the number of local minima of the cost function at each scale.

We propose to use an annealing schedule comparable to the one which is used in simulated annealing. By reducing geometrically the temperature from a high initial value (at this value nearly 100% of the tested transitions are accepted) and by letting the system evolve to a near equilibrium configuration (by using the Creutz algorithm) at each temperature, the results are very good. At every temperature, we know whether the system has reached a near equilibrium configuration by computing the ratio given in equation (28). Once the system has converged at temperature T, each demon energy E_c^i is multiplied by a coefficient Λ such that $\Lambda < 1$ and $\Lambda \approx 1$. Then, the temperature decrease is geometric:

$$\begin{aligned} T^{n+1} &= \frac{1}{kN} \sum_{i=1}^{N} \left(\Lambda \, E_c^i \right) \\ &= \Lambda \, T^n. \end{aligned} \tag{32}$$

The microcanonical algorithm is given in Appendix C.

C Relative efficiency of SA, MFA, and MCA

The three algorithms just described and outlined in the Appendix derive directly from statistical physics and therefore, in order to converge, must properly simulate the behaviour of interacting spin systems.

The annealing schedules are the same for the three algorithms and they approximate the one proposed by Geman & Geman [11] for simulated annealing.

In the case of stochastic methods (simulated annealing and microcanonical annealing) a large number of iterations is necessary such that the statistics become meaningful. That is, at fixed temperature, or at fixed total energy, a large number of elementary transformations (spin flips) must be tested before the system is guaranteed to converge towards a Boltzmann distribution. We recall that N is the number of spins (variables) and this number is equal to the number of possible transformations of a state. Then, the number of iterations at each temperature may be of the order of $100N$ at small temperatures. Nevertheless, in tha case of microcanonical analysis, it is possible to estimate at each iteraton of the annealing process whether the equilibrium has been reached or not, i.e., eq. 28. Therefore, it is expected that the number of iterations at each temperature associated with microcanonical annealing is smaller than the number of iterations associated with simulated annealing.

Mean field annealing is a deterministic approximation of the average behaviour of the system's variables in simulated annealing. One iteration of the MFA algorithm consists of updating the value of a system variable according to an updating rule described by eq. (22). At fixed temperature N iterations are necessary in order to update the whole state vector. In asynchronous mode the state

	Annealing	Computation of energy variation	Computation of of probability	Required accuracy of computations
SA	geometric	yes	yes	high
MFA	geometric	no	no	high
MCA	geometric	yes	no	low

Table 1: A qualitative comparison of the efficiency of simulated annealing, mean field annealing, and microcanonical annealing.

vector is sequentially updated, each variable being considered in a random order. In a synchronous mode all the variables are updated in parallel. Therefore, on a parallel machine, the latter mode is in theory N times faster than the former mode. In practice, as we already mentioned, the convergence process associated with the latter mode is subject to oscillations.

We analyse now the amount of computation required at each iteration. In the case of SA, an energy variation must be first calculated, eq. (14) at each iteration (for each spin flip). Second the probability of accepting or rejecting this flip is estimated according to eq. (13). If the energy variation is positive, $\exp\left(-\frac{\Delta E}{T}\right)$ is evaluated and compared to a random number r, $0 \leq r \leq 1$. Third, whenever the transition is accepted, the global energy is updated.

In the case of MCA, an energy variation must be computed as well at each iteration. However, the decision of whether to accept or reject the transition is taken by simply comparing the value of the energy variation with the kinetic energy currently available – no probability need be computed. In the case of acceptance both the kinetic and potential energies are updated. Nevertheless, a random number has to be generated once each demon has been updated in order to determine its new location on the lattice.

In the case of MFA no energy computation is needed at all. The only computaton involved at each iteration is described by eq. (22).

SA and MCA operate on discrete-valued variables (-1 or $+1$) while MFA operates on real-valued variables taken in the interval $[-1, +1]$. Moreover, since MCA doesn't have to evaluate any mathematical function, highly accurate computations are not required. In conclusion, MFA in either asynchronous or synchronous modes compares favourably with MCA and SA. At first glance, in terms of computation time, MCA appears to be more efficient that SA. Nevertheless, Creutz [7] noticed that when the size N of the lattice is too small the system is more subject to be trapped in local minima of high energy than in the case of canonical analysis: finite size effects differ from those in the canonical approach. Table 1 summarizes this comparison.

V Examples

We tested these algorithms over a wide variety of images. Figure 7 and Figure 8 show two such images. These images are preprocessed as follows. Edges are first extracted using the Canny/Deriche operator [8]. A small set of connected edges are grouped to form an edgel as follows: The tangent direction associated with such an edgel is computed by fitting a straight line in the least-square sense to the small set of connected edges. Then this small set of edges is replaced by an edgel,

	λ (eq. (2))			Energy minimum			CPU time (s)			Nb. of iterations		
	#1	#2	#3	#1	#2	#3	#1	#2	#3	#1	#2	#3
SA	1.7	2.0	2.5	–2720.2	–94.9	–251.7	9311	4898	9523	7741077	6163313	7967099
MFA	1.7	2.0	2.5	–2718.2	–94.9	–252.4	287	103	224	131250	77700	115500
MCA	1.7	2.0	2.5	–2715.5	–94.8	–252.7	1103	962	1520	892682	915486	1220494

Table 2: A summary of the results obtained with the three images (see text). The CPU times are those obtained with a Sun 4/330.

i.e., the fitted line. The position of the edgel is given by its midpoint, its direction is given by the direction of the line, and its contrast is given by the average contrast of the edges forming the edgel. Figure 9 and Figure 10 show the input data of our experiments.

We ran the three algorithms over the three sets of available data: Image #1 (the synthetic image of Figure 1), image #2 (Figure 9), and image #3 (Figure 10). Table 2 summarizes the experimental results. Obviously, the quality of the results are nearly independent of the method that is used: the energy minima obtained with the various methods are almost the same. Nevertheless, the convergence times obtained when the algorithms are sequentially implemented differ from one method to another. MFA is from 32 to 47 times faster than SA and from 3.8 to 9 times faster than MCA. This is due to the small number of iterations that is needed in MFA as compared with the other methods. The number of iterations given in Table 2 is the number of tested transitions in the case of SA and MCA and the number of μ_i updates in the asynchronous MFA. This number is 59 to 79 times smaller in MFA than in SA and 6.8 to 11.8 times smaller in MFA than in MCA. Figure 11 shows a typical evolution of the μ_i variables in the MFA. The number of iterations is smaller in MCA than in SA because one easily determines when the equilibrium is reached in MCA. Figures 12 and 13 show the number of iterations as a function of the number of steps of the annealing process in SA and MCA. Figure 14 shows the value of $\mathrm{Log}\left(\frac{<E_c>}{\rho(E_c)}\right)$ as a function of the number of energy steps once the equilibrium has been reached at each step: at high energies, the thermodynamic equilibrium is not reached but this does not affect the convergence because at those energies all the tested transitions are actually accepted. At lower energies, the equilibrium is reached with a precision of less than 1%.

Figure 2 shows the result of applying mean field annealing to the synthetic data (image #1). Figure 15 and Figure 16 show the results of applying simulated annealing and microcanonical annealing to the synthetic data. Figures 17 and 18 show the corresponding evolutions of the cost functions in the SA and MCA methods.

Finally Figure 19 through Figure 24 show the results of applying the three optimization algorithms to image #2 and image #3.

VI Discussion

In this paper we attacked the problem of figure-ground discrimination with special emphasis on the problem of separating image data into curve and noise. We proposed a global approach using combinatorial optimization. We suggested a mathematical encoding of the problem which takes

into account such image properties as cocircularity, proximity, and contrast and which fits the constraints of the statistical modelling of interacting spin systems.

We tested our approach with three algorithms: We described in detail these algorithms, compared them and implemented them. Not surprisingly, both the comparison and the experimental results allow one to conclude that simulated annealing is the less efficient method and that MFA is the most efficient one. Nevertheless, since the SA algorithm that we implemented was proved to converge to a state that approximates very closely the fundamental state (the overall minimum of the energy function), the results obtained with SA are useful for evaluating the results obtained with the other two algorithms.

Our results are farther validated by the use of similar annealing schedules in all the experiments, that is, a geometric temperature (or energy) decrease. A linear decrease would have obviously been more efficient but with no guarantee that the global minimum is properly approached.

We conclude that the interacting spin model is well suited for encoding the figure-ground problem. Moreover, the analogy of the energy of such a model with the energy of a recursive neural network allows one to assert that the optimization algorithms proposed here are implementable on a fine-grained parallel machine. In such an implementation, a processor is associated with an edgel (a spin or a neuron) and each processor communicates with all the other processors.

In the future we plan to continue to try to improve our method in order to be able to eliminate all the noisy edgels, even those that are close to a shape and included in this shape by our current encoding. We also plan to try to solve other aspects of the image segmentation problem such as the feature grouping problem. We intend to extend the approach advocated in this paper to other vision problems such as matching and reconstruction.

Appendix

A The simulated annealing algorithm

1. Begin with the system in an arbitrary state: $\vec{\sigma_0}$.
2. Fix the initial temperature T_0. In practice $T_0 = 100$.
3. Make a small change in the state, i.e., a state transition.
4. Evaluate the resulting change in energy ΔE.
5. Accept the transition to the new state with the probability defined by equation (13).
6. Repeat steps 3 through 5 until the system reaches an equilibrium (each possible state transition is tested, at the most 100 times). There are two possible situations.
   (a) If 10% of the tested transitions are actually accepted, then the system is considered to have reached the equilibrium at this temperature. In this case go to step 7, or
   (b) If less than 1% of the tested transitions are actually accepted, then the system is "frozen". In this case go to step 8.
7. Update the temperature according to an annealing schedule that approximates a geometric temperature decrease ($T_{n+1} = 0.93\, T_n$) and go to step 3.

8. Stop.

B The mean field annealing algorithm

1. Fix the convergence mode (synchronous or asynchronous).
   Fix the initial temperature: $T = T_{init} = 100$.
   Fix the temperature decreasing factor between two consecutive iterations: $decT = 0.9995$.
   For all i, μ_i^0 is set to a random value in the interval $[-10^{-5}, 10^{-5}]$.
   Initialize an iteration counter: $iter \leftarrow 0$.

2. $iter \leftarrow iter + 1$.
   Case of an asynchronous convergence mode:

   (a) $\forall i$, $\mu_i^{old} = \mu_i^{iter-1}$.

   (b) Scan at random the spins μ_i and iteratively update each spin μ_i see n in the scan (see equation (22)):

      i. Calculate
      $$\mu_i^{new} = \tanh\left(\frac{\sum_{j=1}^{N} J_{ij}\,\mu_j^{old} + \delta_i}{T}\right).$$

      ii. Update μ_i: $\mu_i^{old} \leftarrow \mu_i^{new}$.

   (c) $\forall i$, $\mu_i^{iter} = \mu_i^{old}$.

   Case of a synchronous convergence mode: for all spins μ_i,
   $$\mu_i^{iter} = \tanh\left(\frac{\sum_{j=1}^{N} J_{ij}\,\mu_j^{iter-1} + \delta_i}{T}\right).$$

3. If $iter \leq 5$,

   (a) decrease the temperature: $T \leftarrow decT \times T$,

   (b) goto step 2.

4. Test if the algorithm has converged into a configuration different from the initial one: if $\sum_{i=1}^{N} \sum_{l=iter-5}^{iter} \mid \mu_i^l \mid < 0,99 \times 6 \times N$, then the system has not converged:

   (a) decrease the temperature: $T \leftarrow decT \times T$,

   (b) goto step 2.

5. The system has converged: assign the final values of the σ_i's: $\forall i$, if $\mu_i^{iter} > 0$ then $\sigma_i = 1$ else $\sigma_i = -1$.

In practice we take $T_{init} = 10$ and $decT = 0.9995$. At convergence the values of the μ_i's are very close to either -1 or $+1$. In our experiments we have noticed that this is not the case with mean field approximation in which case the final values of the μ_i's are less disriminative.

C The microcanonical annealing algorithm

1. Begin with a system in an arbitrary state $\vec{\sigma_0}$.
2. Fix the initial kinetic energy of the demons: for all i, $E_c^i = 100$.
3. Sweep at random the lattice of spins and for each visited spin i:
   (a) Perform a spin flip.
   (b) Evaluate the change in potential energy associated with this flip, ΔE_p.
   (c) If $\Delta E_p < 0$, accept the transition and increase the energy of the demon located at this spin: $E_c^i \leftarrow E_c^i - \Delta E_p$.
   (d) If $\Delta E_p \geq 0$ accept the transition according to the following conditions:
   - If $\Delta E_p < E_c^i$, accept the transition and decrease the demon energy: $E_c^i \leftarrow E_c^i - \Delta E_p$.
   - Otherwise, reject the transition.
4. Perform a complete random permutation of the demons.
5. Repeat step 3 and 4 until the system reaches an equilibrium (the lattice of spins is visited at the most 100 times). There are two possible situations.
   (a) If the ratio $\frac{<E_c>}{\rho(E_c)} \in [0.99, 1.01]$ then the equilibrium is reached. Go to step 6.
   (b) If 10% of the tested transitions are actually accepted, then the system is considered to have reached the equilibrium at this total energy value. In this case go to step 6, or
   (c) If less than 1% of the tested transitions are actually accepted, then the system is almost in its fundamental state. In this case go to step 7.
6. Update the kinetic energy according to an annealing schedule that follows a geometric law: for all demon i, $E_c^i \leftarrow 0.93\, E_c^i$. Go to step 3.
7. Stop.

Acknowledgments. The authors would like to thank the anonymous reviewers for their insightfull comments and suggestions, their colleagues for many enlightful dicussions and suggestions, and especially Jean-Jacques Niez, Pierre Puget, and Gérard Giraudon.

References

[1] S. T. Barnard. Stochastic Stereo Matching over Scale. *International Journal of Computer Vision*, 3(1):17–32, May 1989.

[2] A. Blake. Comparison of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-11(1):2–12, January 1989.

[3] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, 1987.

[4] M. Brady and H. Asada. Smoothed local symmetries and their implementation. *International Journal of Robotics Research*, 3(3):36–61, 1984.

[5] J. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, November 1986.

[6] P. Carnevali, L. Coletti, and S. Patarnello. Image Processing by Simulated Annealing. *IBM Journal of Research and Development*, 29(6):569–579, November 1985.

[7] M. Creutz. Microcanonical Monte Carlo Simulation. *Physical Review Letters*, 50(19):1411–1414, May 1983.

[8] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.

[9] D. Geiger and F. Girosi. Parallel and deterministic algorithms from mrf's: Surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):410–412, May 1991.

[10] D. Geiger and A. Yuille. A common framework for image segmentation. *International Journal of Computer Vision*, 6(3):227–243, 1991.

[11] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and Bayesian Restorarion of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, 1984.

[12] D. Gutfinger and J. Sklansky. Robust classifiers by mixed adaptation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13(6):552–567, June 1991.

[13] L. Hérault and J.J. Niez. Neural Networks and Combinatorial Optimisation: A Study of NP-Complete Graph Problems. In E. Gelembe, editor, *Neural Networks: Advances and Applications*, pages 165–213. North Holland, 1991.

[14] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, January 1988.

[15] S. Kirkpatrick, C.D. Gelatt, and Vecchi M.P. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[16] W. Kohler. *Gestalt Psychology*. Meridian, New-York, 1980.

[17] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[18] H. Orland. Mean field theory for optimization problems. *Journal de Physique – Lettres*, 46(17):L–763–L–770, September 1985.

[19] P. Parent and S.W. Zucker. Trace Inference, Curvature Consistency, and Curve Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):823–839, August 1989.

[20] C. Peterson. A new method for mapping optimization problems onto neural networks. *International Journal of Neural Systems*, 1(1):3–22, 1989.

[21] C. Peterson. Track finding with neural networks. *Nuclear Instruments and Methods in Physics Research*, A279:537–545, 1989.

[22] C. Peterson and J. R. Anderson. A Mean Field Learning Algorithm for Neural Networks. *Complex Systems*, 1:995–1019, 1987.

[23] F. Reif. *Fundamental of Statistical and Thermal Physics.* McGRAW-HILL, 1965.

[24] T. J. Sejnowski and G. E. Hinton. Separating Figure from Ground with a Boltzmann Machine. In Michael Arbib and Allen Hanson, editors, *Vision, Brain, and Cooperative Computation*, pages 703–724. MIT Press, 1988.

[25] A. Shashua and S. Ullman. Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network. In *Proc. IEEE International Conference on Computer Vision*, pages 321–327, Tampa, Florida, USA, December 1988.

[26] H.E. Stanley. *Introduction to Phases Transitions and Critical Phenomena.* Oxford University Press, 1971.

[27] D.E. Van den Bout and T. K. Miller. Graph Partitioning Using Annealed Neural Networks. In *Int. Joint Conf. on Neural Networks*, pages 521–528, Washington D.C., June 1989.

[28] J. Zerubia and R. Chellappa. Mean field approximation using compound Gauss-Markov random field for edge detection and image restoration. In *Proceedings of ICASSP*, pages 2193–2196, Alburquerque, New Mexico, USA, April 1990.

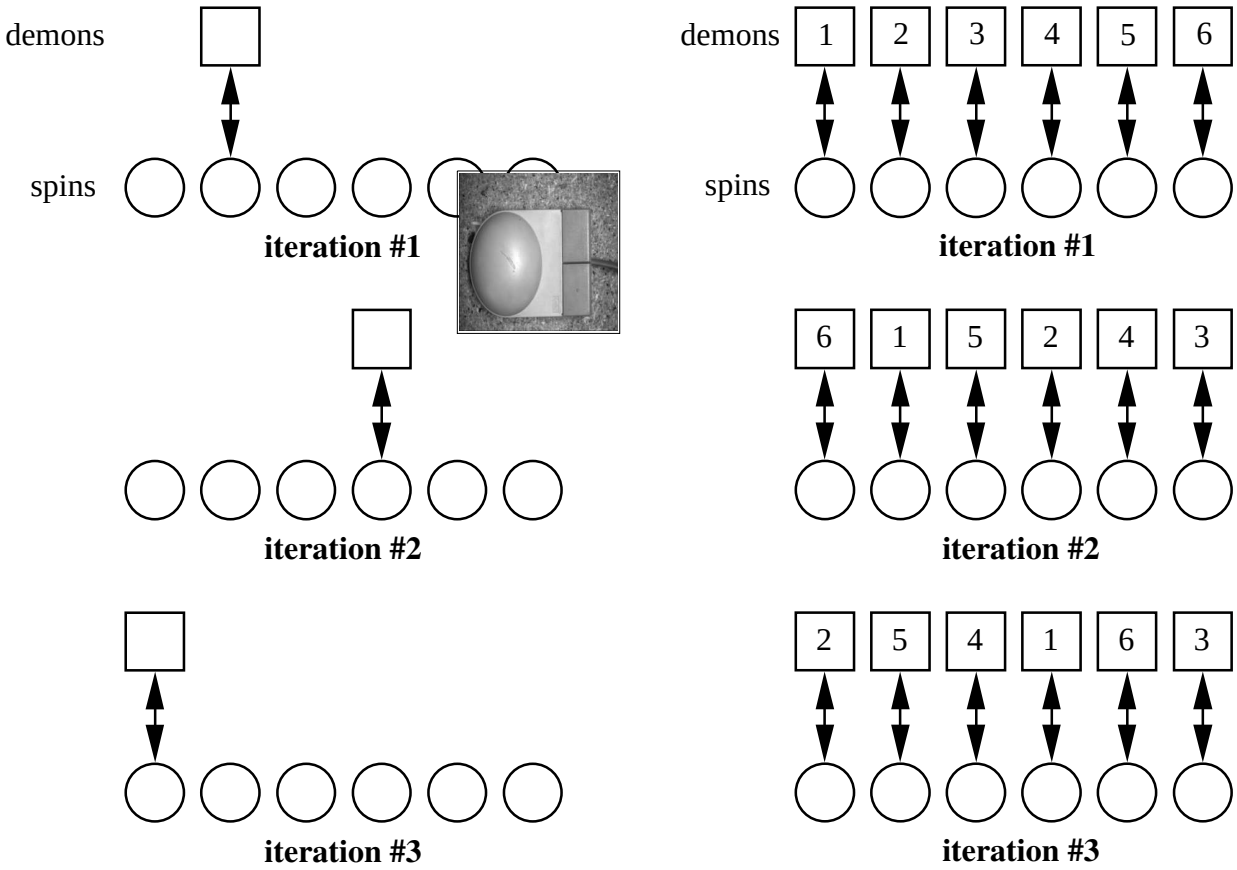
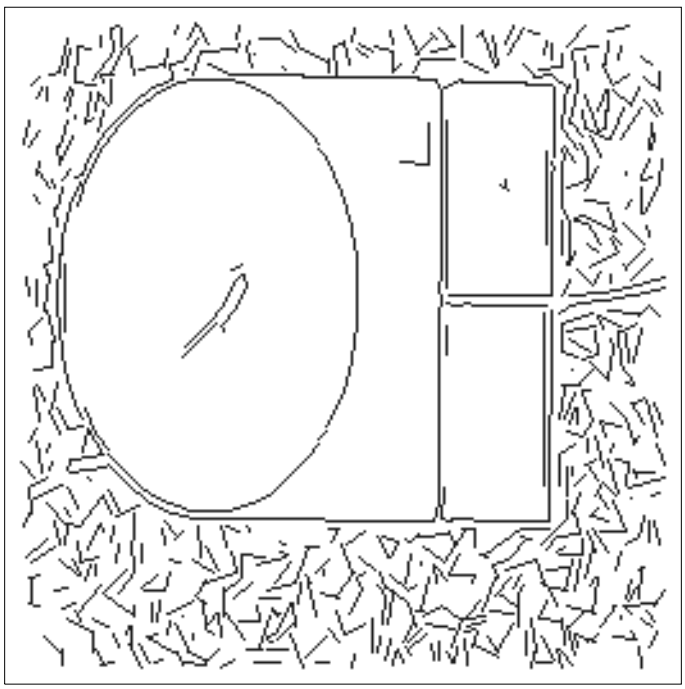
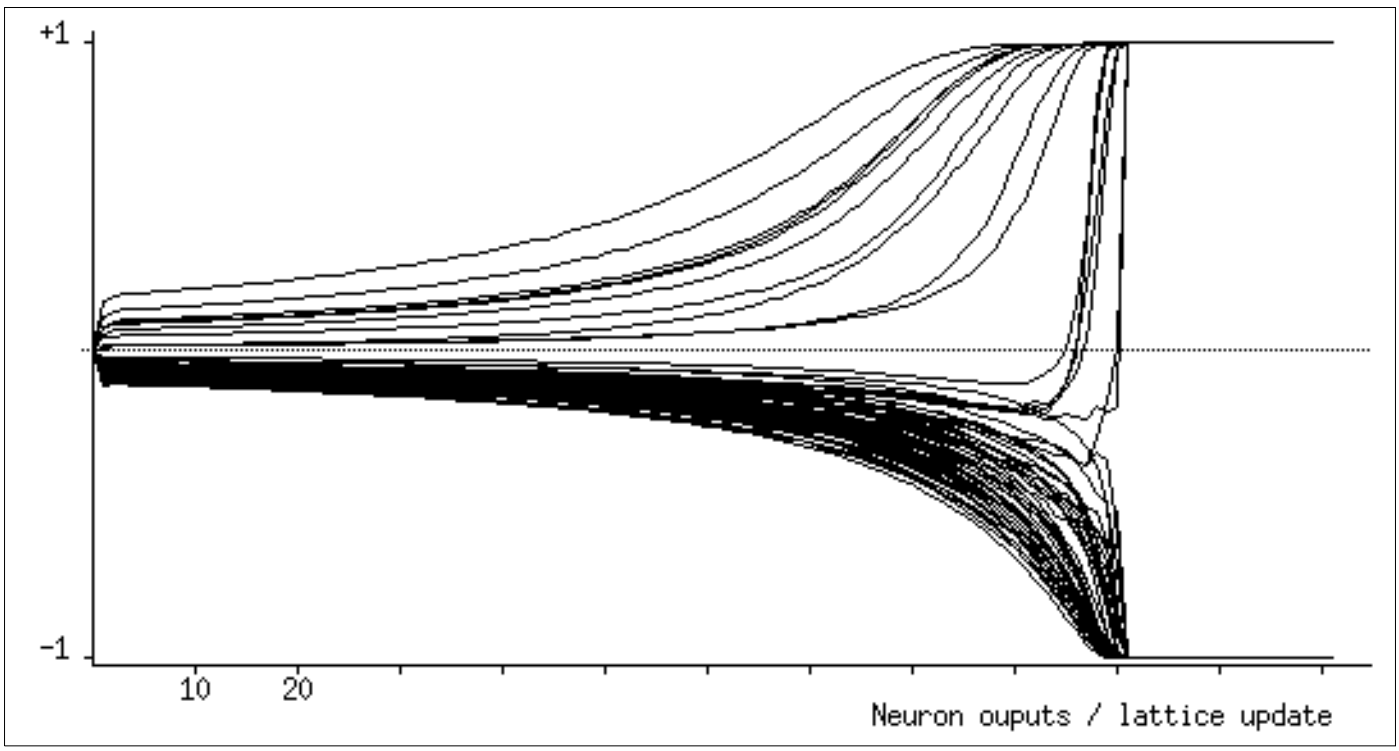


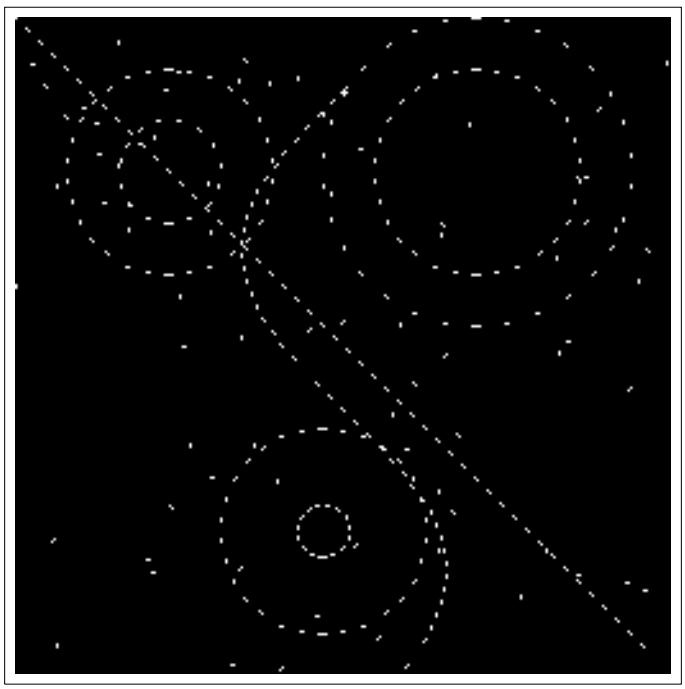
Figure 5: Example of the assignment of a demon to the spins in the Creutz algorithm.

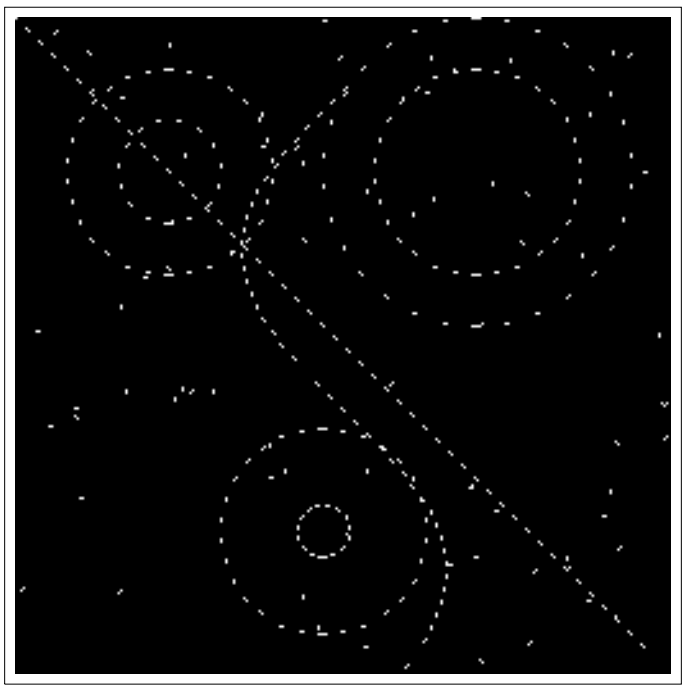
28.0

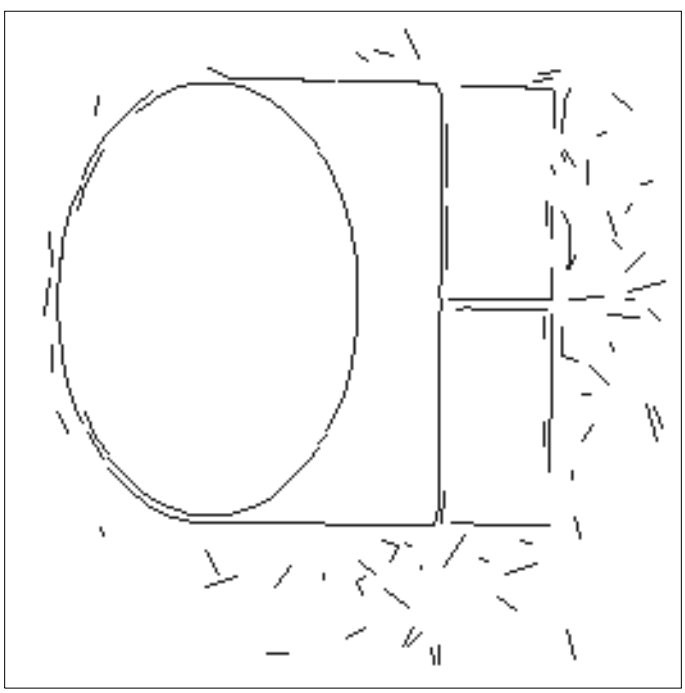


280

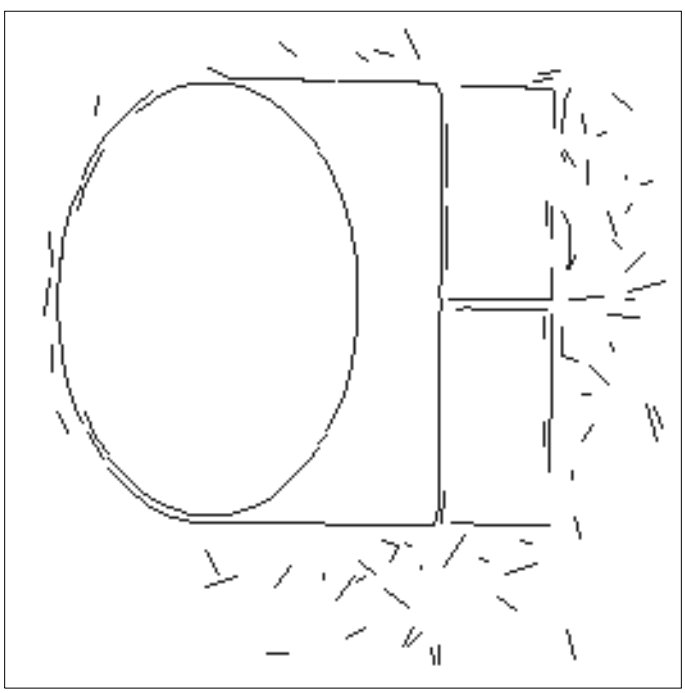

+1
-1
10
20
Neuron ouputs / lattice update



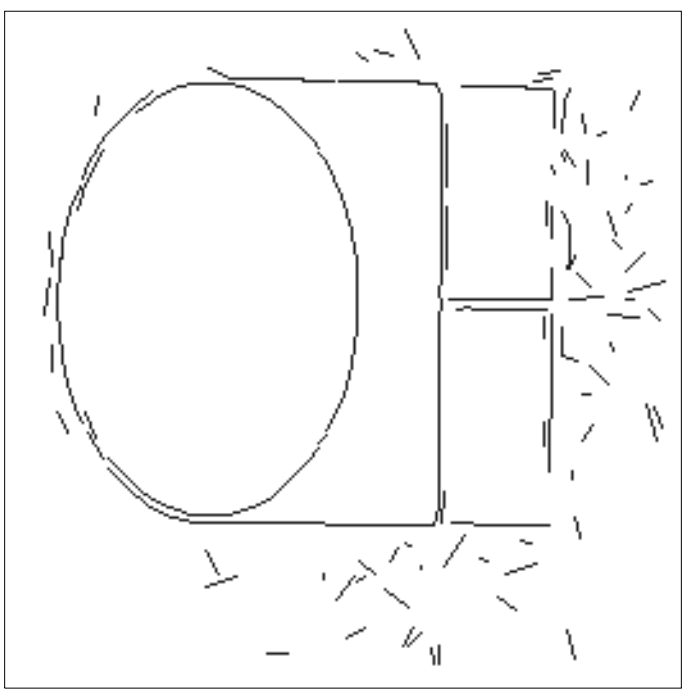




28.0



28.0



28.0