



HAL
open science

Figure-Ground Discrimination by Mean Field Annealing

Laurent Herault, Radu Horaud

► **To cite this version:**

Laurent Herault, Radu Horaud. Figure-Ground Discrimination by Mean Field Annealing. 2nd European Conference on Computer Vision (ECCV '92), May 1992, Santa Margherita Ligure, Italy. pp.58–66, 10.1007/3-540-55426-2_7. inria-00590008

HAL Id: inria-00590008

<https://inria.hal.science/inria-00590008>

Submitted on 1 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Figure-Ground Discrimination by Mean Field Annealing *

Laurent Héroult¹ and Radu Horaud²

¹ CEA-LETI, avenue des Martyrs, 85X, 38041 Grenoble

² LIFIA-IRIMAG, 46, avenue Félix Viallet, 38031 Grenoble, FRANCE

Abstract. We formulate the figure-ground discrimination problem as a combinatorial optimization problem. We suggest a cost function that makes explicit a definition of shape based on interactions between image edges. These interactions have some mathematical analogy with *interacting spin systems* – a model that is well suited for solving combinatorial optimization problems. We devise a mean field annealing method for finding the global minimum of such a spin system and the method successfully solves for the figure-ground problem.

1 Introduction and background

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), mean field annealing (MFA), and microcanonical annealing (MCA) [3]. Here we concentrate on mean field annealing.

The figure-ground or shape/noise separation is best illustrated by an example. Fig. 1 shows a synthetic image. Fig. 2 shows the image elements that were labelled “shape” by the mean field annealing algorithm.

The interest for shape/noise separation stems from Gestalt psychologists’ figure-ground demonstrations [6]: 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. Edge detection is in general first performed. Edge grouping is done

* 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.

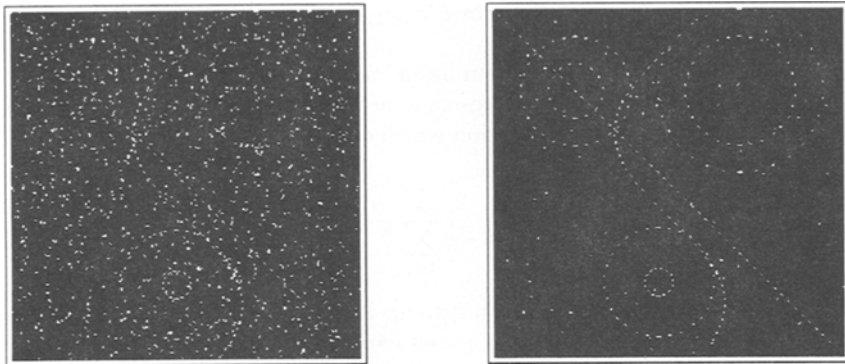


Fig. 1. A synthetic image with 1250 elements. **Fig. 2.** The result of applying mean field annealing to the synthetic image. The elements plunged into randomly generated elements. The elements shown on this figure were labelled “shape”.

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. There are two problems with the techniques just mentioned: (i) 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 and (ii) the notion of *noise* is not clearly specified with respect to the notion of shape.

Our work is related to the influential paper of Parent & Zucker [8] who introduced the notion of cocircularity between two edge elements and to the work of Sha’ashua & Ullman [11], and Sejnowski & Hinton [10].

The work described here has two main contributions. First, we suggest a mathematical encoding of the figure-ground discrimination problem that consists of separating shape from noise using a combinatorial optimization method. Second, we suggest mean field annealing, a deterministic global optimization method.

2 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: (i) A spin state-vector of N elements $\sigma = [\sigma_1, \dots, \sigma_N]$ whose components are described by discrete labels which correspond to *up* or *down* Ising spins: $\sigma_i \in \{-1, +1\}$. The components σ_i may well be viewed as the outputs of binary neurons. (ii) A symmetric matrix J describing the interactions between the spins. These interactions may well be viewed as the synaptic weights be-

tween neurons in a network. (iii) A vector $\delta = [\delta_1, \dots, \delta_N]$ describing an external field in which the spins are plunged.

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:

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

The main property of interacting spin systems is that at low temperatures the number of local minima of the energy function grows exponentially with the number of spins. Hence the adequation between the mathematical model of interacting spin systems and combinatorial optimization problems with many local minima is natural.

We consider now 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 form the state vector $\mathbf{p} = [p_1, \dots, p_N]$. 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 , one may write by analogy with physics an interaction energy: $E_{saliency}(\mathbf{p}) = -\sum_{i=1}^N \sum_{j=1}^N c_{ij} p_i p_j$

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

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

$$E(\mathbf{p}) = E_{saliency}(\mathbf{p}) + \lambda E_{constraint}(\mathbf{p}) \quad (2)$$

In this expression λ is a positive real parameter that has to be adjusted and is closely related to the signal-to-noise ratio. With the substitution $p_i = (\sigma_i + 1)/2$ in eq. (2), the function to be minimized is given by eq. (1) where: $J_{ij} = (c_{ij} - \lambda)/2$ and $\delta_i = (\sum_{j=1}^N c_{ij} - N\lambda)/2$.

3 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 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:

$$g_x(x_i, y_i) = \frac{\partial I_f(x_i, y_i)}{\partial x} \quad \text{and} \quad g_y(x_i, y_i) = \frac{\partial I_f(x_i, y_i)}{\partial y}$$

I_f is the low-pass filtered image. From the gradient vector one can easily compute the gradient direction and magnitude: θ_i and g_i , e.g., Fig. 3:

$$\theta_i = \arctan \left(\frac{g_y(x_i, y_i)}{g_x(x_i, y_i)} \right) + \frac{\pi}{2} \quad \text{and} \quad g_i = (g_x^2(x_i, y_i) + g_y^2(x_i, y_i))^{1/2}$$

Let i and j be two edgels. We want that the interaction between these two edgels encapsulates the concept of shape. That is, if i and j 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) i belongs to one shape and j belongs to another one, (ii) i belongs to a shape and j is noise, or (iii) both i and j are noise. The interaction coefficient must therefore be a co-shapeness measure. In our approach, co-shapeness is defined by a combination of *cocircularity*, *proximity*, and *contrast*.

The definition of cocircularity is derived from [8] and it constrains the shapes to be as circular as possible, or as a special case, as linear as possible. Proximity restricts the interaction to occur in between nearby edgels. As a consequence, cocircularity is constrained to be a local shape property. The combination of cocircularity and proximity will therefore allow a large variety of shapes that are circular (or linear) only locally. Contrast enforces edgels with high gradient module to have a higher interaction coefficient than edgels with a low gradient module.

Following [8] and from Fig. 3 it is clear that two edgels belong to the same circle if and only if: $\lambda_i + \lambda_j = \pi$. 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 positions and 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. Moreover, linearity appears as a special case of cocircularity, namely when $\lambda_i = 0$ and $\lambda_j = \pi$.

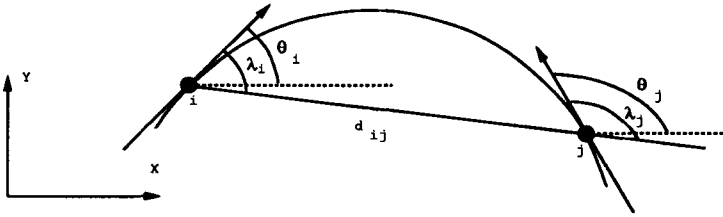


Fig. 3. The definition of cocircularity between two edgels (i and j).

From this cocircularity constraint we may derive a weaker constraint which will measure the closeness of a two-edgel configuration to a circular shape: $\Delta_{ij} = |\lambda_i + \lambda_j - \pi|$. Δ_{ij} will vary between 0 (a perfect shape) and π (no shape). Finally the cocircularity coefficient is allowed to vary between 1 for a circle and 0 for noise and is defined by the formula: $c_{ij}^{cocir} = (1 - \Delta_{ij}^2/\pi^2) \exp(-\Delta_{ij}^2/k)$. The parameter k is chosen such that the cocircularity coefficient vanishes rapidly for non-circular shapes.

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 multiplying the cocircularity coefficient with a coefficient that vanishes smoothly as the two edgels are farther away from each other: $c_{ij}^{prox} = \exp(-d_{ij}^2/(2\sigma_d^2))$ where d_{ij} is the distance between the two edgels and σ_d is the standard deviation of these distances over the image. Hence, the edgel interaction will adjust itself to the image distribution of the edgel population.

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 of this simply-minded nonlinear filtering is to consider two thresholds such that edgel connectivity is better preserved [1]. 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: $c_{ij}^{contrast} = g_i g_j / g_{max}^2$ where g_{max} is the highest gradient value over the edgel population. Finally the interaction coefficient between two edgels becomes:

$$c_{ij} = c_{ij}^{cocir} c_{ij}^{prox} c_{ij}^{contrast} \quad (3)$$

4 Mean field annealing (MFA)

The states reachable by the system described by eq. (1) 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 analysis 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) [3]. We will consider here the 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 . At low temperatures and at equilibrium, F is minimal and TS is close to zero. Therefore, the internal energy E is minimized. 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 eq. (1). The remarks above are expressed in the most fundamental result of statistical physics, the Boltzmann (or Gibbs) distribution:

$$Pr(E(\sigma) = E_i) = \frac{\exp(-E_i/(kT))}{Z(T)} \quad (4)$$

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 kT (k is the Boltzmann's constant). $Z(T)$ is called the *partition function* and is a normalization factor: $Z(T) = \sum_n \exp(-E_n/(kT))$. This sum runs over all possible spin configurations. Using eq. (4) one can compute at a given temperature T the mean value over all possible configurations of some macroscopic physical parameter A :

$$\langle A \rangle = \sum_n A_n Pr(E(\sigma) = E_n) = \frac{\sum_n A_n \exp(-E_n/(kT))}{Z(T)} \quad (5)$$

Unfortunately, the partition function $Z(T)$ is usually impossible to compute. Nevertheless, when the system is described by eq. (1) one can use eq. (5) which, with an additional hypothesis, is a basic equation for mean field approximation and for mean field annealing.

In order to introduce the mean field annealing algorithm, we first introduce the somewhat more classical mean field approximation method which has been used to solve optimization problems [7, 9, 4]: it is a simple analytic approximation of the behaviour of interacting spin systems in thermal equilibrium. We start by developing eq. (1) around σ_i :

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

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* (Φ_i) 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 of 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:

$$\langle \Phi_i \rangle = - \left(\sum_{j=1}^N J_{ij} \langle \sigma_j \rangle + \delta_i \right) \quad (6)$$

We introduce now the following approximation [12]: 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, eq. (4):

$$P(X_i = \sigma_i^0) \approx \frac{\exp(-\langle \Phi_i \rangle \sigma_i^0 / T)}{P(X_i = 1) + P(X_i = -1)}, \quad \sigma_i^0 \in \{-1, 1\} \quad (7)$$

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 eq. (5), (6), and (7), the mean state of σ_i can now be easily derived:

$$\langle \sigma_i \rangle \approx \frac{(+1) \exp(-\langle \Phi_i \rangle / T) + (-1) \exp(\langle \Phi_i \rangle / T)}{\exp(\langle \Phi_i \rangle / T) + \exp(-\langle \Phi_i \rangle / T)} = \tanh \left(\frac{\sum_{j=1}^N J_{ij} \langle \sigma_j \rangle + \delta_i}{T} \right) \quad (8)$$

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

$$\mu_i = \tanh \left(\frac{\sum_{j=1}^N J_{ij} \mu_j + \delta_i}{T} \right) \quad (9)$$

where $\mu_i = \langle \sigma_i \rangle$. The problem of finding the mean state of interacting spin system at thermal equilibrium is now mapped into the problem of solving a system of N coupled

non-linear equations, i.e., eq. (9). In the general case, an analytic solution is rather difficult to obtain. Instead, the solution for the vector $\boldsymbol{\mu} = [\mu_1, \dots, \mu_N]$ may well be the *stationary solution* of the following system of N differential equations:

$$\tau \frac{d\mu_i}{dt} = \tanh \left(\frac{\sum_{j=1}^N J_{ij} \mu_j + \delta_i}{T} \right) - \mu_i \quad (10)$$

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

$$\left. \frac{d\mu_i}{dt} \right|_{t_n} = \frac{\mu_i^{n+1} - \mu_i^n}{\Delta t} + o(\Delta t) \quad (11)$$

where μ_i^n is the value of μ_i at time t_n . By substituting in eq. (10) and by choosing $\tau = \Delta t$, we obtain an iterative solution for the system of differential equations described by eq. (10):

$$\forall i \in \{1, \dots, N\}, \mu_i^{n+1} = \tanh \left(\frac{\sum_{j=1}^N J_{ij} \mu_j^n + \delta_i}{T} \right), \quad n \geq 1 \quad (12)$$

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 $\boldsymbol{\mu}^0 = [\mu_1^0, \dots, \mu_N^0]$ the convergence is reached at the n 'th iteration such that $\boldsymbol{\mu}^n = [\mu_1^n, \dots, \mu_N^n]$ becomes stationary. The physicists would say that the thermal equilibrium has been nearly reached. When the vector $\boldsymbol{\delta}$ is the null vector, one can start with a vector $\boldsymbol{\mu}^0$ close to the obvious unstable solution $[0, \dots, 0]$. In practice, even when the vector $\boldsymbol{\delta}$ is not null one starts with an initial configuration obtained by adding noise to $[0, \dots, 0]$. For instance the μ_i^0 's are chosen randomly in the interval $[-10^{-5}, +10^{-5}]$. This initial state is plausible for an interacting 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 . During the iterative process, the μ_i 's converge to values in between -1 or +1.

Two convergence modes are possible:

- *Synchronous mode.* At each step of the iterative process **all** the μ_i^n 's are updated using the μ_i^{n-1} 's previously calculated.
- *Asynchronous mode.* At each step of the iterative process a 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 $\boldsymbol{\sigma}$ from the vector $\boldsymbol{\mu}$, 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: *if* $1/2(1 + \mu_i^n) > 0.5$ *then* $\sigma_i = +1$ *else* $\sigma_i = -1$.

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 [5] and other authors [13] 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 according to two possible annealing schedules:

- 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 slightly 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. This strategy was successfully used to solve hard np-complete graph combinatorial problems [4], [5];
- Van den Bout & Miller [13] tried to estimate the critical temperature T_c . At this temperature, some of the mean field variables μ_i begin to move significantly towards either -1 or $+1$. Hence their strategy consists performing two sets of iterations: one iteration process at this critical temperature until a near state equilibrium is reached and another iteration process at a temperature value that is close to 0. However, the critical temperature is quite difficult to estimate.

We currently use the first of the annealing schedules described above.

5 Example and discussion

We tested these algorithms over a wide variety of images. The images are preprocessed as follows. Edges are first extracted using the Canny/Deriche operator [2]. The tangent direction associated with such an edge is computed by fitting a straight line in the least-square sense to a small set of connected edges. Then this small set of edges is replaced by an edgel, 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. Fig. 4 shows the input data of our example. Fig. 2 and Fig. 5 show the result of applying mean field annealing.



Fig. 4. Set of edgels obtained with no noise elimination.

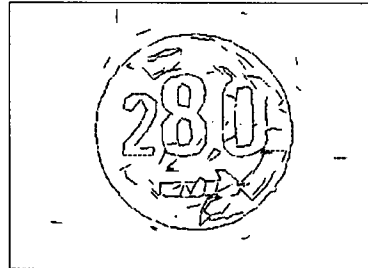


Fig. 5. The result of applying MFA to the image on the left.

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 this encoding fits the constraints of the statistical modelling of interacting spin systems.

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

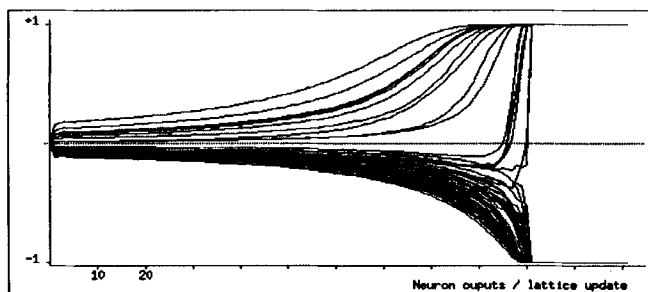


Fig. 6. Evolution of the μ_i variables in MFA for the image above.

a recursive neural network allows one to assert that the MFA algorithm proposed here is implementable on a fine-grained parallel machine. In such an implementation, a processor is associated with an edgel (or a spin, or a neuron) and each processor communicates with all the other processors.

References

1. J. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, November 1986.
2. R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.
3. L. Héroult and R. Horaud. Figure-ground discrimination: a combinatorial optimization approach. Technical Report RT 73, LIFIA-IMAG, October 1991.
4. L. Héroult and J.J. Niez. Neural Networks and Graph K-Partitioning. *Complex Systems*, 3(6):531–576, December 1989.
5. L. Héroult and J.J. Niez. Neural Networks and Combinatorial Optimisation: A Study of NP-Complete Graph Problems. In E. Gelembé, editor, *Neural Networks: Advances and Applications*, pages 165–213. North Holland, 1991.
6. W. Kohler. *Gestalt Psychology*. Meridian, New-York, 1980.
7. H. Orland. Mean field theory for optimization problems. *Journal Physique Lettres*, 46:L-763–L-770, 1985.
8. 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.
9. C. Peterson. A new method for mapping optimization problems onto neural networks. *International Journal of Neural Systems*, 1(1):3–22, 1989.
10. 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.
11. A. Sha'ashua 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.
12. H.E. Stanley. *Introduction to Phases Transitions and Critical Phenomena*. Oxford University Press, 1971.
13. 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.