Autonomous algorithms for image restoration

Meir Griniasty

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Autonomous algorithms for image restoration

Meir Griniasty

N° 2405

Novembre 1994

PROGRAMMES 4 et 5

Robotique, image et vision

Traitement du signal, automatique et productique

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Autonomous algorithms for image restoration

Meir Griniasty

Programmes 4 et 5 — Robotique, image et vision — Traitement du signal, automatique et productique
Projets Pastis et Mefisto

Rapport de recherche n ° 2405 — Novembre 1994 — 20 pages

Abstract: We describe a general theoretical framework for algorithms that adaptively tune *all* their parameters during the restoration of a noisy image. The adaptation procedure is based on a mean field approach which is known as "Deterministic Annealing", and is reminiscent of the "Deterministic Boltzmann Machine". The algorithm is less time consuming in comparison with its simulated annealing alternative. We apply the theory to several architectures and compare their performances.

Key-words: Image processing, Statistical Mechanics, Deterministic Annealing

(Résumé : tsvp)

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Algorithmes autonomes pour la restauration d'image

Résumé : Nous décrivons une famille d'algorithmes qui adaptent *tous* ses paramètres pendant la restauration d'une image bruitée. La procédure d'adaptation est basée sur une approche de type champ moyen, connu aussi sous le nom de "Recuit Déterministe" et qui s'apparente à une "Machine de Bolzmann Déterministe". Les algorithmes consomment moins de temps que leurs analogues de type "Recuit Simulée". Nous comparons les performances des algorithmes fondés sur des architectures différentes.

Mots-clé : Traitement d'image, Mécanique statistique, Recuit déterministe.

1 Introduction

In this paper we address the problem of image restoration. We are given a matrix of grey levels (real numbers), and should decide which label, out of a given set of possible labels, to assign to each pixel.

Our working assumption follows Geman and Geman [1]: we assume that our image labels are produced by a Markov random field, and then corrupted by an additive Gaussian noise. Each label type has a different corresponding distribution of grey levels. If the couplings (interaction parameters) of the Markov field and the noise parameters are given to us, then one can show that the restoration problem becomes a minimization problem of an energy function which is defined in the space of possible labeling configurations.

In this work we consider the more complicated case, where no information is given about the couplings and noise distribution parameters. We present algorithms that perform simultaneously both tasks: choosing the parameters and assigning labels to pixels.

In the theoretical part of the paper we shall describe a Bayesian approach, which leads to the formulation of the parameter estimation and labeling problem as an optimization problem in the combined space of parameters and labels. The minimization of the resulting energy function is approached by applying the machinery of statistical mechanics. More specifically, we use an approach which is called "Deterministic Annealing". The equations obtained in this approach are then translated into iterative algorithms. The translation is not unique or obvious. We discuss different possible algorithms, explain the difference in their results, and choose a preferable one.

The equations we derive and the resulting algorithm are general, and apply to different connectivity architecture of the underlying Markov field. In the second part of the work we consider several specific neighborhood architectures and show some results.

2 Theory

Following Geman and Geman [1], we assume that our image is a produced by a Markov random field, and then corrupted by an additive Gaussian noise.

The probability to get a set of image grey levels y is given therefore by:

$$P(y) = \sum_{\eta} P(y|\eta)P(\eta) \tag{1}$$

where η is a configuration of labels and y is a configuration of grey levels. Each η_i may take q different values, which correspond to q classes. Since we assume a Markov field in the origin of the image, the prior probability to get a pixel configuration η is the Bolzmann distribution

$$P(\eta) = \exp(-E_1(\eta))/Z_1 \tag{2}$$

E_1 is an energy function, which depends on the label configuration η, on the system of neighborhood, and on the size of coupling parameters. In this work we consider energy functions only of the form

$$E_1(\eta) = -\sum_{(ij)} W_{ij}\,\delta(\eta_i, \eta_j)\ . \tag{3}$$

Where δ is the Kronecker delta function, which is one if labels η_i and η_j are the same and zero otherwise. The sum runs over all pairs (ij). The different models that we shall consider have different structures of the matrix of couplings W.

Z_1 is the partition function associated with E_1:

$$Z_1 = \sum_{\eta} \exp(-E_1(\eta)) \tag{4}$$

and

$$P(y|\eta) = \prod_i \exp(-(y_i - \mu_i)^2/2\sigma_i^2)/\sqrt{2\pi}\sigma_i \tag{5}$$

is the Gaussian additive noise. Each label type has different μ and σ. μ_i and σ_i are the average and standard deviation of the distribution of the grey levels that corresponds to the class of the pixel i. In other words: $\mu_i = \mu(\eta_i)$, $\sigma_i = \sigma(\eta_i)$. The index i runs over all image pixels. In what follows we always consider image pixels that are organized in a two dimensional square lattice.

Until now we have described our assumptions about the probabilistic process which leads to the matrix of grey levels. However, our task is to do the

inverse: we are given the set of grey levels and we would like to find the most probable parameters and labelings. The quantity that we would like to maximize is the conditional probability $P(\eta, \theta|y)$. Using Bayes theorem we have:

$$P(\eta, \theta|y) = P(\eta, \theta, y)/P(y) = P(y|\eta, \theta)P(\eta|\theta)P(\theta)/P(y)$$

where θ stands for μ,σ and the coupling parameters. assuming that $P(\theta)$ is constant our problem becomes the maximization of the expression

$$P(y|\eta, \theta)P(\eta|\theta) \tag{6}$$

with respect to η and θ. The first term is given in Eq. 5 and the second in Eq. 2.

2.1 Prior models used in this work

The simplest model considered in this work is the two dimensional Potts model:

$$E_1(\eta) = -W \sum_{(ij)} \delta(\eta_i \ \eta_j) \tag{7}$$

where the sum is over the nearest neighbors on the square lattice. In this case all parameters W_{ij} are equal to W. The magnitude of W controls the tendency of nearby pixels to have the same label. The bigger W the bigger is the Bolzmann probability for homogeneous configurations. Generalizations of the Potts model allow for interactions between pixels that are farther away. We consider the following generalizations:

- A multirange model: Closest neighbors on the square lattice are coupled with strength W_1,second closest with strength W_2 etc.
- A hierarchical model: The image pixels are grouped into groups of four (each forming a square), each group interacts with a single pixel of a second group (second "layer") of pixels, with a coupling strength W_1. the second layer, which has four times less pixels, is also organized into groups of four, and each of these groups interacts with a single pixel of a third layer, with a coupling strength W_2 and so on. Note that only the first layer of pixels are image pixels, while the other are "service"

pixels, who serve as intermediators among the image pixels. This model is reminiscent of the model of Kato et al [5].

- Layered multirange model: each group of four pixels of the image interacts with one pixel of a second layer with coupling W_1. four groups of four pixels of the first layer (the image layer) interacts with one pixel of a third layer with strength W_2 etc. This model is different from the previous one since there is no direct interaction among "service" pixels. This model has been proposed originally by F. Heitz et al [2]

2.2 The effective energy function

Coming back to expression 6, and using Eqs. 2, 3 and 5, the maximization of the logarithm of this probability is equivalent to the minimization of the following effective energy function

$$E(\eta,\theta) = -\log P(y|\eta,\theta) - \log(\eta|\theta) = E_0 + E_1 + \log Z_1 + C^t \qquad (8)$$

E_0 comes from the Gaussian term

$$E_0(\eta) = -\sum_i (y_i - \mu_i)^2/2\sigma_i^2 - log(\sigma_i)$$

where $\mu_i = \mu(\eta_i)$, $\sigma_i = \sigma(\eta_i)$, as explained earlier. Z_1 is the partition function of the prior model (eq. 4) and C^t is a constant that does not depend on the labelings and parameters. This is similar to the expression arrived at [3] with the exception that here the parameters μ and σ should be estimated as well.

We explain now the significance of the different terms. The energy term E_0 tends to choose for each pixel the label whose associated average grey level (μ) is the closest to the given grey level y. The term E_1 is lower when neighboring pixels have similar labels. These two terms might be conflicting, and their competition defines the form of the restored image. The size of the couplings (W for the Potts case, W_1, W_2 etc. for other models) defines which of the two terms will win. If W is very big (and positive) then E_1 always wins over E_0 simply by choosing a configuration with all labels the same. This effect is controlled by the term $\log Z_1$. In contrast to the other terms, this term has no dependence on the label configuration, only on the coupling parameters.

2.3 Minimization using a Mean Field approach

In this work we approach the minimization of $E(\eta, \theta)$ (Eq. 8) using Deterministic Annealing ([6],[7],[8]). In this approach, one minimizes the mean field approximation to the free energy at finite temperature, while gradually decreasing the temperature. One hopes that in this way local minima will be avoided and the resulting parameter and labeling configuration at zero temperature will minimize the energy. This approach, although not proven to converge to the global minimum, has been found to be an effective, computer time saving, alternative to simulated annealing [4].

First we write the partition function, at inverse temperature β, that corresponds to the energy function at Eq. 8

$$Z(\beta) = \sum_{\eta,\theta} \exp(-\beta E(\eta, \theta)) = \sum_{\theta} Z(\beta, \theta) \tag{9}$$

the sum is over all labelling configurations, and possible parameters θ.

$$Z(\beta, \theta) = \sum_{\eta} \exp(-\beta E(\eta, \theta)) \ .$$

The free energy of the system is related to the partition function by

$$F(\beta) = -\log(Z(\beta))/\beta \ .$$

It is important to distinguish between the two types of variables: the "small" variables, η, and the "big" ones θ. The theta's are treated as *order parameters*, which means that $Z(\beta, \theta)$ is sharply peaked around some value $\theta^\star$, and one can consider only this value:

$$F(\beta) \approx \max_{\theta} F(\beta, \theta) = F(\beta, \theta^\star) = -\log(Z(\beta, \theta^\star))/\beta \ .$$

On the other hand, the summation over the labeling configurations should be actually performed. This summation (which can not be done exactly) can be evaluated approximately by a variational approach (see for example [9]) which leads to a mean field theory.

For completeness we outline here the variational approach. Given $Q(\eta)$, a probability distribution over the space of label configurations, and an energy

function $E(\eta)$, we define the following functional of the probability distribution:

$$F(Q) = \sum_{\eta} Q(\eta)E(\eta) - S(Q)/\beta \tag{10}$$

$$S(Q) = -\sum_{\eta} Q(\eta)\log Q(\eta)$$

$S(Q)$ is the entropy associated with $Q(\eta)$. One can straightforwardly prove that of all probability distributions, the one which minimizes F is the Bolzmann distribution $Q(\eta) = \exp(-\beta E(\eta))/Z$. Moreover, the value of the minimal F is just the free energy of the system, These relations serve as a basis for a variational approach: since we can not calculate Z accurately, we essay to minimize $F(Q)$ with respect to a restricted group of probability distributions, which have a factorization property: $Q(\eta_1....\eta_N) = \prod_i P_i(\eta_i)$.

The probability of η_i to have the label α is denoted P_i^α. Plugging this into eq. 10 and using the explicit form of our energy function, Eq. 8, we obtain

$$\begin{aligned} F(P,\theta) &= -\sum_{(ij)} W_{ij} \sum_{\alpha,\beta} P_i^\alpha P_j^\beta - \sum_{i,\alpha} P_i^\alpha \left((y_i - \mu_\alpha)^2/2\sigma_\alpha^2 + \log(\sigma_\alpha)\right) \\ &\quad + \sum_{i,\alpha} P_i^\alpha \log(P_i^\alpha)/\beta + \log(Z_1) \ . \end{aligned} \tag{11}$$

This expression holds for all the architectures considered in this work. The double sum runs over all pairs of labels (ij). In the first single sum i runs over all *image* pixels, which are those that obtain the y information. The second single sum runs over *all* pixels (the multirange model has only image pixels, while the hierarchical model, for example, has also pixels which do not receive grey level information).

The resulting mean field equations for the probabilities are given by taking the derivative of F with respect to P_i^α and equating to zero. For image pixels we obtain the following equations

$$P_i^\alpha = \frac{\exp(\beta\, h_i^\alpha)}{\sum_\gamma \exp(\beta\, h_i^\gamma)} \tag{12}$$

with $h_i^\alpha = \sum_j W_{ij} P_j^\alpha + (y_i - \mu_\alpha)^2/2\sigma_\alpha^2 + \log(\sigma_\alpha)$

For additional pixels we have the same equations, but $h_i^\alpha = \sum_j W_{ij} P_j^\alpha$.

The mean field equations for the parameters θ are obtained by differentiating F with respect to μ, σ and W and equating to zero. Note that there is no entropy term associated with this variables in F. This is because we treat them as order parameters, and assume they have no fluctuations. The resulting equations for μ and σ are very natural:

$$\mu_\alpha = \sum_i y_i P_i^\alpha / \sum_i P_i^\alpha \ , \tag{13}$$

$$\sigma_\alpha^2 = \sum_i (y_i - \mu_\alpha)^2 P_i^\alpha / \sum_i P_i^\alpha \ . \tag{14}$$

The equations for the W's involve differentiation of $\log Z_1$ with respect to W. Z_1 is also a partition function, with an energy function E_1 and temperature $\beta = 1$, and is therefore calculated by the same variational approach.

The result is

$$F_1 = -\log(Z_1) = -\sum_{ij} W_{ij} \sum_{\alpha,\beta} M_i^\alpha M_j^\beta + \sum_{i,\alpha} M_i^\alpha \log(M_i^\alpha) \ . \tag{15}$$

The equations for M_i^α, the probability that pixel i has the label α, are

$$M_i^\alpha = \frac{\exp(\beta\, h_i^\alpha)}{\sum_\gamma \exp(\beta\, h_i^\gamma)} \tag{16}$$

where $h_i^\alpha = \sum_j W_{ij} M_j^\alpha$

In the models that we consider the W_{ij} are not all different (if we allow all W's to be different our model will be too flexible, and there will be no restoration of the image). For instance in the Potts model all W's are the same. In a hierarchical model with three layers all connections from the first layer to the second are the same, and all connections from the second to the third are also the same. Differentiating with respect to weights of type l we have

$$d\log(Z_1)/dW_l = \sum_{ij \in l,\alpha} M_i^\alpha M_j^\alpha$$

where the sum runs over all pairs of pixels that are connected by W_l. THe equation for W_l is obtained by differentiating F with respect to it and equating

to zero. We obtain finally the equations that determine the values of the W's:

$$\sum_{ij\in l,\alpha} P_i^\alpha P_j^\alpha = \sum_{ij\in l,\alpha} M_i^\alpha M_j^\alpha \ . \tag{17}$$

Note that $\sum_\alpha M_i^\alpha M_j^\alpha$ is the probability that pixel i and pixel j have the same label. The last equations are therefore reminiscent of the learning equations for the deterministic Bolzmann machine [10, 11]. One compares the (mean field approximation for the) correlations among pixels with (P's) and without (M's) the information about the image, which is conveyed by the y's.

2.4 Mean Field motivated Algorithms

The mean field equations 12-17 serve as a basis for our labeling and parameter estimation algorithm. The algorithms that we considered have two phases: in one phase the P's are varied, while the parameters θ are fixed, as we solve Eq.12. In the second phase, the P's are kept constant while the θ parameters are varied. The re-estimation of the μ and σ parameters is given by writing the mean field equations 13,14 as iterative equations

$$\mu_\alpha(n+1) = \sum_i y_i P_i^\alpha / \sum_i P_i^\alpha \tag{18}$$

$$\sigma_\alpha(n+1)^2 = \sum_i (y_i - \mu_\alpha)^2 P_i^\alpha / \sum_i P_i^\alpha \tag{19}$$

when we are at the $n+1$ estimation step, and the P parameters have been estimated using the $\theta(n)$ variables.

the re-estimation of the W's is done by transforming eqs. 17 into iterative equations. Maximizing F with respect to the algorithm:

$$\Delta W_l = \epsilon dF/dW_l = \epsilon(C_l^0 - C_l)$$

where $C_l = \sum_{ij\in l,\alpha} P_i^\alpha P_j^\alpha$ and $C_l^0 = \sum_{ij\in l,\alpha} M_i^\alpha M_j^\alpha$ and ϵ is a small constant. However, this procedure consumes too much computer time since at each iteration we have to solve eqs. 12 anew for the P's. We therefore replace this algorithm by another one:

$$\Delta W_l = \epsilon(C_l^0(W) - C_L(W_n)) \ . \tag{20}$$

This means that we solve each time for the M's with different W's, but keep the P's constant, until we arrive at a solution $C_l^0(W) = C_l(W_n)$ for all l's, and choose these W to be W_{n+1}.

To summarize, our algorithm consists of iteratively solving eqs. 12,18,19,20, while decreasing the temperature. Our experience shows that there are several possible solutions to these equations, and that the type of solution obtained depends on the procedure of solution, and even on the initial conditions of the order parameters. Generally speaking, we observe three types of solutions: the first is what we call "W dominant". in this case, pixels of neighbors tend to have the same label, and effect of E_0 term, which conveys information about the grey levels, is secondary. The second type is "$\mu\sigma$ dominant": labels are determined mainly by the local y variable. In this case the resulting image is almost as noisy as the original. The third type is a *homogeneous* solution: the μ's and the σ's of all labels are found to be equal. The W's are zero.

Subjectively, the solution that we call a "good" one is the W dominant.

In our first attempt we tried to solve all equations 12, 18,19,20 while decreasing the temperature. The result of this algorithm was a *homogeneous* solution. This algorithm is comparable with the one used by Geman [3]. The later is a Simulated Annealing algorithm, where only the coupling parameters, but not the noise parameters, are estimated. Geman reports that his algorithm is successful. In our case, the freedom in the choice of the width of the noise leads to overestimation of the width of one dominant group.

After numerous trials we have found an efficient way to achieve a W dominant solution:

1. choose initially high values for W's, then solve for the P's, gradually decreasing the temperature, and estimate the μ's and σ's only at a very low temperature. Keep W's unchanged.

2. Start again at high temperature and cool down, solving only Eq. 12. μ, σ, and W parameters are re-estimated only as one arrives to low temperature by solving Eqs. 18,19,20.

3. Repeat step 2 until the order parameters stabilize.

We found that the values of the order parameters stabilize after a few (of order five) complete sweeps over the whole temperature range.

Our Mean Field algorithm offers advantages over the stochastic algorithm both in the re-labeling and parameter re-estimation phases. In the first phase

the mean field equations for the $P's$ may be solved in parallel, in contrast to the sequential update required by Simulated Annealing. In the second phase, equation 20 is solved more easily in the mean field approach. Since the M's correspond to a homogeneous system (no external input) one can solve one or a few mean field equations, in contrast to the simulated annealing approach, where one should solve for the full system.

3 Experimental results

The first architecture that we considered was hierarchical. The connections from the first layer to the second introduce effective interaction inside each group of 2x2 pixels of the first layer. The connections between the second and third layer introduce indirectly interaction between groups of 4x4 pixels of the first layer, etc. We expected that if homogeneous regions of our image are, for example, of size 8X8 then our adaptive algorithm will tend to enhance the connections between the first and second layer, second and third, third to fourth but not further.

The set of eqs. 20 that define the connections have the following form here:

$$1/N_L \sum_{i,\alpha} P^{\alpha}_{i,L} P^{\alpha}_{j(i),L+1} = \sum_{\alpha} M^{\alpha}_{L} M^{\alpha}_{L+1} \tag{21}$$

where the P's are solutions of the MF equations of the full system, N_L is the number of pixels is the L'th layer, $j(i)$ is the pixel in the $L+1$ layer that is connected the pixel i in the L'th layer. Taking advantage of the fact that the system, in the absence of the term E_0, is homogeneous, we calculate the M's (solve eq. 16) assuming that the solution $M^{\alpha}_{i,L}$ is equal for all members of the L'th layer. The number of equations that define the values of the M's is therefore reduced to the number of layers times the number of labels.

We applied the algorithm on a chessboard image, Fig. 1. The size of the image is 128x128 pixels, and each single colored square is of size 32x32. The average grey level in the white region is .58 and the RMS is .1, in the black region the average grey level is .46 and the RMS fluctuation .1. we expected that $W_L, L = 1, 5$ will be positive, resulting in interactions inside each square. We have found, however, that iteration of eqs. 21 leads to a solution where only the W_1 is positive, while W_2 is zero. Since the structure is hierarchical, the

result is that only groups of four in the image layer (the first layer) interact, whatever is the value of W_3, W_4 etc. This interaction forces groups of four pixels to have the same label, therefore eliminating the smallest defects, but not eliminating defects of size four or bigger.

The problem of dependence of the interactions of longer range on the interaction of shorter range is overcome in the layered multirange model. Here groups of four pixels of the image layer interact among themselves effectively via a single pixel in the second layer with coupling $W2$, while groups of sixteen (four groups of four) interact via a pixel in the third layer with coupling $W3$ etc. equations 20 are written here

$$1/N_1 \sum_{i,\alpha} P^{\alpha}_{i,1} P^{\alpha}_{j(i),L} = \sum_{\alpha} M^{\alpha}_1 M^{\alpha}_L$$

where L runs from two to the number of layers. We show here results for the chessboard when the number of layers is three (Fig. 2) and four (Fig. 3). In the case of three layers, groups of four times four pixels of the image are forced to be the same, and we still find defects. In an architecture with four layers this defects disappear, and the restoration is perfect.

The disadvantage of this architecture is evident as we try to process Fig. 4 (again 128 by 128 pixels, grey level distribution in the white region: $.62 \pm .22$, in the black region: $.38 \pm .22$). Here we see that with increasing number of layers, the size of homogeneous regions increases, and the result is crude. Figures 5, 6, 7 correspond to architectures with three four and five layers respectively. The conclusion is that the layered multirange model finds connections which are too strong, and forces all pixels inside a square (whose size depends on the number of layers) to have the same label, even if a "natural border" passes through it. Another disadvantage of this model is the lack of translation invariance: the division of the image into square regions of interacting pixels is arbitrary, and when the interconnected squares are shifted the resulting processed image is different.

We turn now to a description of the results for the simple short range model. The results are surprisingly good. In Fig. 8 we see the processed chessboard, and in Fig. 9 the quarter circle. The boundaries in both images are not in right angles, in contrast to the previous model, and the results for the circle are much finer.

We have also experimented with a nonhierarchical multirange model. We have found that in this case the system tends, in a manner reminiscent of the hierarchical model, to diminish the long range interactions, and the final results are the same as the short range case.

4 Discussion

In this paper we have shown how to transform the mean field equations into iterative algorithms which effectively estimate parameters and restore an image. There are two new ingredients in our approach. The first is the fact that *all* the parameters are found autonomously by the algorithm. The second is the use of deterministic annealing.

Our approach has two shortcomings. One is that the algorithms tend to choose short range interactions in models where long range are possible (with the exception of the layered multirange model). A second problem is that we have not succeeded to classify correctly images with more than two types of labels. In this case our algorithms tend to merge two or more groups into one. The second problem can be solved if the widths of the distributions of grey levels in each class are supplied. In this case, however, the algorithm is no more "autonomous".

Acknowledgements I have benefited from instructive and useful discussions with Fabien Campillo, Josiane Zerubia and Zoltan Kato.

References

[1] S. Geman and D. Geman "Stochastic relaxation, Gibbs Distributions, and the Bayesian Restoration of Images" IEEE Trans. PAMI, Vol. 6 No. 6, P. 721 (1984)

[2] F. Heitz, P. Perez, and P. Bouthemy. Parallel Visual Motion Analysis Using Multiscale Markov Random Fields. In *Proc. Workshop on Motion*, Princeton, Oct. 1991.

[3] D. Geman, "Bayesian Image Analysis by Adaptive Annealing" proceedings of IGARSS conference, Amherst, US, P. 269 (1985)

[4] D. Geiger and F. Girosi "Paralel and Deterministic Algorithms for MRF's: Surface Reconstruction and Integration", Lecture Notes in Computer Science, No. 427, P. 89. Editor: O. Faugeras (1990)

[5] Z. Kato, M. Berthod and J. Zerubia, "A Hierarchical Markov Random Filed Model and Multi-Temperature Annealing for Parallel Image Classification" INRIA report No. 1938, 1993

[6] K. Rose, E. Gurewitz and G. Fox, "Vector Quantization by Deterministic Annealing" IEEE Trans. Inf. Theo., Vol. 38 No. 4 P. 1249 (1992)

[7] J. Buhmann and H. Kuhnel "Vector Quantization with Complexity Costs" IEEE Trans. Inf. Theo., Vol. 39 No. 4 P. 1133 (1993)

[8] A. L. Yuille and J.J. Kosowsky "Statistical Physics Algorithms that converge" Neural Comp. Vol 6 No. 3 P. 341 (1994)

[9] G. Parisi "Statistical Field Theory", Editors: Edisson Wesly (1988)

[10] G.E. Hinton and T.J. Sejnowsky, "Learning and Unlearning in Bolzmann Machines". Parallel Distributed Processing, Vol 1, eds: D.E. Rumelhart, J. L. McLelland. Cambridge MA: MIT press (1986)

[11] C. Peterson and J.R. Anderson "Exploration of the Deterministic Bolzmann Machine" Complex Systems **1**, P. 995 (1987)

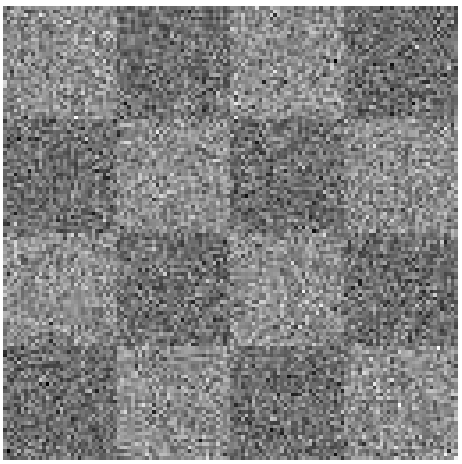

Figure 1: Noisy chessboard image.

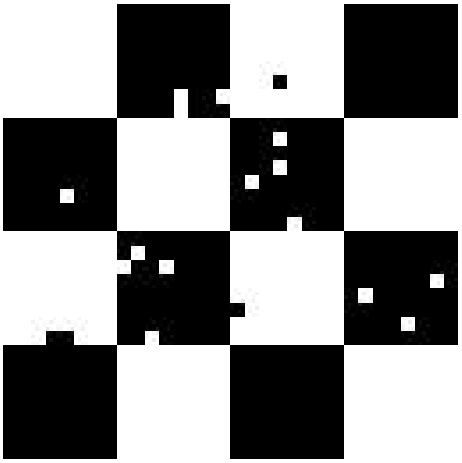

Figure 2: Noisy chessboard treated with the layered multirange model with three layers.

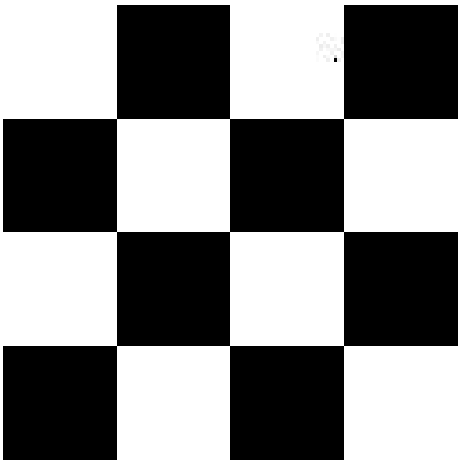

Figure 3: Noisy chessboard treated with the layered multirange model with four layers.

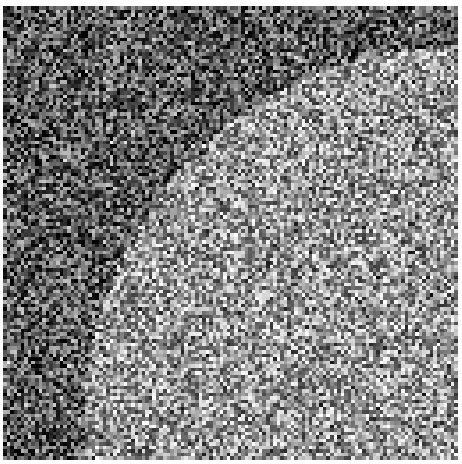

Figure 4: Noisy quarter of a circle image.

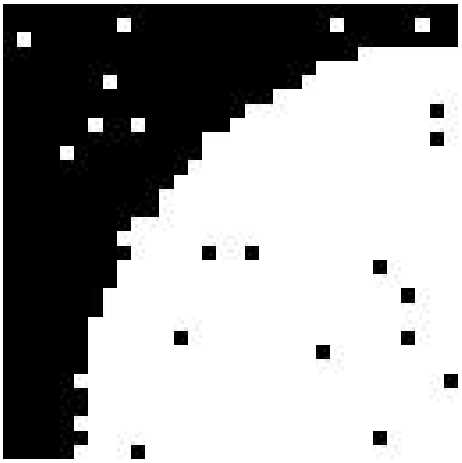

Figure 5: Noisy quarter–circle treated with the layered multirange model with three layers.

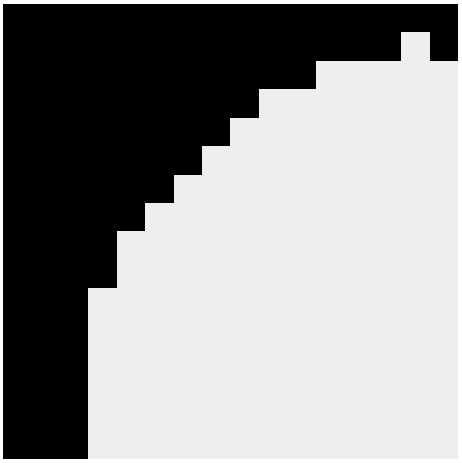

Figure 6: Noisy quarter–circle treated with the layered multirange model with four layers.

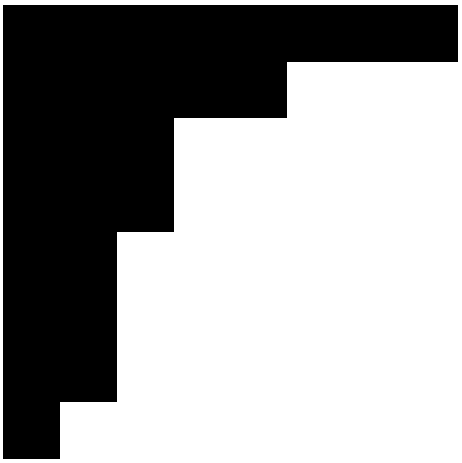

Figure 7: Noisy quarter–circle treated with the layered multirange model with five layers.

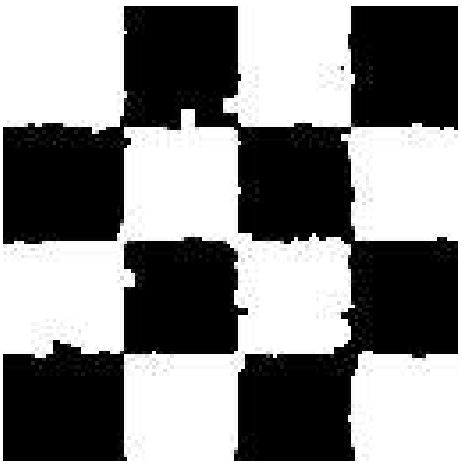

Figure 8: Noisy chessboard treated with the short range model.

Figure 9: Noisy quarter–circle treated with the short range model.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399