PROGETTO FINALIZZATO
SISTEMI INFORMATICI E CALCOLO PARALLELO

SOTTOPROGETTO 2

Coordinatore Prof. Franco Denoth
Processori Dedicati

L.Bedini, I.Gerace, M.Pepe, E.Salerno, A.Tonazzini*

STOCHASTIC AND DETERMINISTIC ALGORITHMS FOR IMAGE RECONSTRUCTION WITH IMPLICITLY REFERRED DISCONTINUITIES

N. R/2/85 OTTOBRE 1992

Internal Report

* Istituto di Elaborazione della Informazione
Consiglio Nazionale delle Ricerche
Via S.Maria 46, 56126 Pisa

STOCHASTIC AND DETERMINISTIC ALGORITHMS FOR IMAGE RECONSTRUCTION WITH IMPLICITLY REFERRED DISCONTINUITIES

L.Bedini, I.Gerace, M.Pepe, E.Salerno, A.Tonazzini

Istituto di Elaborazione della Informazione
Consiglio Nazionale delle Ricerche
Via S.Maria, 46
I-56126 PISA (Italy)

ABSTRACT

The MRF-based approach to the solution of visual reconstruction problems, even if extremely flexible in describing the local behaviour of the image intensity and its discontinuities, generally needs stochastic relaxation algorithms to compute the MAP estimate. These algorithms, despite their asymptotic convergence properties, often present insurmountable computational costs.

Different methods, which consider the discontinuities implicitly rather than explicitly, have been recently proposed. In these methods, a general image reconstruction problem is formulated as the minimization of a regularized energy F(f) of the only intensity process f. Geman and Reynolds showed that, if the stabilizer employed has the attributes of concavity and finite asymptotic behaviour, the solution to this problem permits the recovery of discontinuities without introducing auxiliary variables. While this problem is equivalent to one involving an explicit line process, its complexity is greatly reduced. They minimized the non-convex energy F by using an optimization method based on simulated annealing.

In this paper, we investigate the possibility of solving the same problem through the Graduated Non-Convexity algorithm, that is by constructing a sequence of approximating energies and

using gradient descent techniques to minimize them. This deterministic method is compared with the direct minimization of the energy through stochastic relaxation, in terms of quality of the reconstructed images and of computational costs.

1. Introduction

Accounting for discontinuities in visual reconstruction problems is nowadays recognized to be fundamental, as it allows the regularizing smoothness constraint to be broken where it has no physical meaning. The most popular approach to deal with discontinuities makes use of Bayesian techniques, based on Markov Random Fields (MRF) models for describing the local behaviour of the image intensity and its discontinuities (the line process) [1], [6]. The solution to the reconstruction problem is the maximizer of the posterior probability, which is a function of both the intensity process and the line process. The computation of this solution employs stochastic relaxation algorithms, that, despite their asymptotic convergence properties, often present insurmountable computational costs.

Less expensive methods, both stochastic and deterministic, have been recently proposed [3], [4], [5], [7]. These methods consider the discontinuities implicitly rather than explicitly, and are based on designing a cost functional F(f) on the only intensity process f, which is a weighted sum of a prior constraint (regularization term) and a consistency constraint (data term). The reconstruction is the image f which minimizes F. If the regularization term satisfies certain conditions, the solution to this problem permits the recovery of discontinuities without introducing auxiliary variables and indeed this problem is equivalent to one involving an explicit line process.

Geman and Reynolds formalized the statement above developing a duality theory that relates a class of primal energies F(f), in which the discontinuities are implicitly referred through appropriate regularization terms, with a class of dual energies E(f,l), in which discontinuities l are explicitly marked and suitably constrained [5]. The primal and dual energies are equivalent in the sense that, if (f*,l*) is the global minimizer of E(f,l), then f* is the global minimizer of F(f). The dual theory states that the conditions under which this equivalence exists are the concavity and the finite asymptotic behaviour of the regularization term, and gives a tool to derive the dual energy function from the primal one, in the general case of a continuous positive-valued line process. In particular, Geman and Reynolds analyzed the

performance of regularization terms that are locally of the form $\phi(D^k f)$, where $\phi(t)=-(1+|t|)^{-1}$ and D^k, the k-th order derivative operator, k=1,2,3, determines the degree of smoothness assumed for the image. They minimize the still non-convex primal energy F by using an optimization method based on simulated annealing, that is coarse-to-fine in the order of the derivative.

Blake and Zisserman derived the cost functional F(f) eliminating the binary line process from the energy associated to a Weak Membrane system, for the reconstruction of a piecewise continuous image [12]. Since function F is still non-convex, they constructed a family of approximations $F^{(p)}$, depending upon a parameter p, $p \in [0,1]$, such that $F^{(0)}=F$ and $F^{(1)}$ is convex. Starting from p=1, a gradient descent algorithm is successively applied to the various approximations, for a prescribed decreasing sequence of values of p. This is the essence of the deterministic Graduated Non-Convexity algorithm (GNC). In general, such a procedure is not guaranteed to give the global minimum, but, since the energy function for image reconstruction problems does not have too many minima, the results are usually satisfactory even if not optimal. Moreover, Blake and Zisserman proved the convergence of the GNC algorithm for a particular class of data [3]. From the duality theory point of view, their function F can be seen as a particular primal energy and the Weak Membrane energy is the corresponding dual energy.

Geiger and Girosi proposed the mean field approximation theory to average out the binary line process from the Weak Membrane energy [4], [13]. In this way they obtained a set of deterministic equations, from which it is possible to estimate the statistic mean values of the intensity and the discontinuity fields. They provided a family of energy functions depending on temperature, and showed that the GNC, when applied to this family, can be seen as a deterministic annealing.

In this paper, we formulate the problem of the reconstruction of piecewise smooth images from sparse and noisy samples, relating the MRF approach with line process to the duality theory proposed by Geman and Reynolds. For simplicity, we consider a first order model for the image, i.e. we adopt the first derivative in the regularization term. We propose to minimize the primal

energy F(f) through the GNC algorithm, and construct a family of approximating function $F^{(p)}$, each minimized by a gradient descent technique. This deterministic algorithm is compared with the direct minimization of the primal energy through stochastic relaxation. With respect to the quality of the reconstructed images and the computational costs, the comparison evidences that both the two algorithms give satisfactory results in the limits imposed by the underlying image model. We observed that stochastic relaxation goes closer to the global minimum of energy F than GNC; this is paid with a significant increase in the computation time.

2. MRF formulation of the problem

Following the MRF-based approach, the original piecewise smooth image is modelled using three coupled first-order neighbourhood MRF's: a continuous-valued intensity field f, whose generic element corresponds to the intensity value $f_{i,j}$ of the pixel at location (i,j), and two line fields, h and v, whose variables are ideally associated to sites placed midway between each vertical or horizontal pair of pixels [14]. While traditionally these line elements were considered binary, here we let them assume continuous non-negative values in [0,M]. The role of the line variables is to weight the strength of the continuity constraint associated with each pair of adjacent pixels [5]. The local dependence among the elements of the fields is expressed by the Gibbsian joint probability distribution

$$P(f,h,v) = \frac{1}{Z} \exp\left(\frac{-U(f,h,v)}{\beta}\right)$$

$$U(f,h,v) = \sum_{c} V_c(f,h,v) \qquad (2.1)$$

where Z is the normalizing constant, β is a positive constant, U(f,h,v) is the energy function and the potentials $V_c(f,h,v)$ are functions supported on the cliques associated to the neighbourhood system. As we refer to a first order neighbourhood system, we only associate non-zero potentials to the mixed cliques

of the types $\{f_{i,j}, f_{i-1,j}, h_{i,j}\}$ and $\{f_{i,j}, f_{i,j-1}, v_{i,j}\}$, and to the line cliques constituted of single line elements, i.e. $\{h_{i,j}\}$ and $\{v_{i,j}\}$.

The general form for U(f,h,v) is given by:

$$U(f,h,v) = \lambda^2 \sum_{i,j} (f_{i,j}-f_{i,j-1})^2 \xi(v_{i,j}) + \lambda^2 \sum_{i,j} (f_{i,j}-f_{i-1,j})^2 \xi(h_{i,j}) +$$

$$+ \sum_{i,j} \psi(h_{i,j}) + \sum_{i,j} \psi(v_{i,j}) \qquad (2.2)$$

where $\xi(b)$ is an increasing function such that $\xi(0)=0$ and $\psi(b)$ is a decreasing function. The first two terms express the interaction between the intensity process and the line process, and are conceived to make more likely as solution images which present decreasing line values where the absolute value of the horizontal or vertical gradient is high, and are smoothly varying where the line variables assume high values. In other words, a line element of low value weakens the smoothness constraint between the two pixels across it; the two pixels can thus assume very different values, that is a discontinuity in the image is created. Vice versa, a high value of the line element enforces the smoothness constraint. In the third and fourth terms the decreasing function ψ prevents the line elements to assume low values everywhere.

It is to note that this model for the image can be extended to higher order derivative models, to describe images that are locally smooth in the first or second derivative, away from visual boundaries and textured areas. In these cases, higher order neighbourhood systems must be adopted [5].

In the assumption of a noise process whose components are independent, white and Gaussian, with zero mean and variance σ^2, for the Bayes theorem the posterior probability of the image (f,h,v) given the data g is:

$$P(f,h,v|g) = \frac{1}{Z} (2\pi\sigma^2)^{-N^2/2} \exp\left(- \frac{||\mathbf{g}-\mathbf{H}\mathbf{f}||^2}{2\sigma^2} - \frac{U(f,h,v)}{\beta} \right) \qquad (2.3)$$

where **f** and **g** are the lexicographic notations for matrices f and g, N^2 is the dimension of the problem and H is a matrix that models the physical relationship between the data and the true image.

The Maximum A Posteriori (MAP) estimate maximizes $P(f,h,v|g)$ and thus coincides with the minimum of the posterior energy function

$$E(f,h,v) = \frac{\|g-Hf\|^2}{2\sigma^2} + \frac{U(f,h,v)}{\beta} \tag{2.4}$$

that, in consideration of (2.2) becomes:

$$E(f,h,v) = \frac{\|g-Hf\|^2}{2\sigma^2} + \lambda^2 \sum_{i,j} (f_{i,j}-f_{i,j-1})^2 \xi(v_{i,j}) +$$

$$+ \lambda^2 \sum_{i,j} (f_{i,j}-f_{i-1,j})^2 \xi(h_{i,j}) + \sum_{i,j} \psi(h_{i,j}) + \sum_{i,j} \psi(v_{i,j}) \tag{2.5}$$

where β has been set to 1. In this context, the positive parameter λ^2 determines a compromise between data fitting and smoothness constraints satisfaction in the reconstructed image. The Weak Membrane energy, as derived euristically by Blake and Zisserman [3], and in the MRF context by Geiger and Girosi [4], is given by eq. (2.5) when H=I, the line process is assumed binary, ξ and ψ are linear functions. The presence of the line process in E makes it a non-convex function whose minimization cannot be performed through ordinary gradient descent algorithms. In [1] D. Geman and S. Geman, for binary line elements, proposed the use of Monte Carlo techniques based on a Gibbs sampler coupled with simulated annealing. They proved the asymptotic convergence of this stochastic relaxation algorithm when the temperature is decreased according to a logarithmic schedule, and suggested a parallel implementation. Nevertheless, despite proposals of possible implementations on parallel general purpose or dedicated hardware [2], [10], stochastic relaxation algorithms often present prohibitive computational costs.

More recently, some authors have evidenced that the preliminary minimization of the energy E with respect to h and v, if it can be performed analytically, would eliminate the line process from E, thus reducing the computational complexity of the MAP problem. Moreover, this would make possible the design of ad hoc deterministic algorithms to solve it [3], [4].

3. Elimination of the line process

The possibility of eliminating the line process from the energy E(f,h,v) of eq. (2.5) is based on the following considerations: if we define a function F(f) such that

$$F(f) = \inf_{h,v} E(f,h,v) \tag{3.1}$$

and it is

$$(f^*,h^*,v^*) = \arg \min_{f,h,v} E(f,h,v) \tag{3.2}$$

then

$$\arg \min_{f} F(f) = f^* \tag{3.3}$$

In other words, the search for the global minimum of the energy function E(f,h,v) can be restricted to the set $\{f, h^*(f), v^*(f)\}$, where, for each f, $(h^*(f), v^*(f))$ is the minimizer of the energy function with respect to (h,v).

The computation of F(f) as in (3.1) is straightforward if the model adopted for the image does not include self-interactions of the line process. In our case this computation leads to:

$$F(f) = \frac{\|\mathbf{g}-\mathbf{H}\mathbf{f}\|^2}{2\sigma^2} + \sum_{i,j} \phi(f_{i,j}-f_{i,j-1}) + \sum_{i,j} \phi(f_{i,j}-f_{i-1,j}) \tag{3.4}$$

where ϕ, called neighbour interacting function, does not depend on the particular site (i,j). If interaction terms between two or more discontinuities are to be included in the image model, for example to describe local smoothness constraints on discontinuities, the minimization of the energy E(f,h,v) with respect to h and v is still possible, adopting preliminary minor approximations for the energy E itself. In [17], [22] an extension of the Weak Membrane model has been proposed to take into account for interacting discontinuities and a GNC-type deterministic algorithm has been developed with application to image reconstruction from sparse and noisy samples.

The energy F(f) of eq. (3.4) is a function of the only intensity variables f, and has the particularity that it addresses discontinuities implicitly rather than explicitly. Indeed, function ϕ retains all information about the local interaction between the intensity process and its discontinuities.

Geman and Reynolds in [5] reversed the problem and, using a regularization approach, derived the energy (3.4) as a cost functional which is a weighted sum of the data fitting term and a regularization term or stabilizer whose form express the a priori knowledge on the desired solution. They derived a general duality theory and proved that, from particular forms of function ϕ, it is possible to uniquely recover an energy E(f,h,v), in the form given by eq. (2.5) and such that (3.1) holds. In this sense, the problems of minimizing F and E are equivalent. They call F primal energy, while E is named dual energy, and proved the following

Theorem (existence of a dual)

Given a function $\phi(t)$ with the following properties on $[0, \infty)$:

1. $\phi(0)=0$
2. $\phi(\sqrt{t})$ is concave
3. $\lim_{t \to +\infty} \phi(t)=\alpha$

then there exist two functions $\xi(b)$ and $\psi(b)$ defined on an interval [0,M] such that

$$\phi(t) = \inf_{0 \le b \le M} (\lambda^2 t^2 \xi(b) + \psi(b)) \qquad (3.5)$$

and satisfying the following properties

1. $\psi(b)$ is decreasing
2. $\psi(0)=\alpha$
3. $\psi(M)=0$
4. $\xi(b)$ is increasing
5. $\xi(0)=0$

The geometric proof of the theorem is based on the fact that $\phi(\sqrt{t})$ is seen as the lower envelope of a one-parameter family of straight lines y=mt+q, where $m=\lambda^2\xi(b)$ and $q=\psi(b)$. Thus, if $\phi(\sqrt{t})$ is

strictly concave, then $\psi(b)$ results strictly decreasing, $\xi(b)$ strictly increasing and $M\lambda^2$ is the right-hand derivative of $\phi(\sqrt{t})$ at the origin. In [5], Geman and Reynolds proposed the two following neighbour interacting functions:

$$\phi(t) = \frac{\lambda^2 t^2}{\frac{\lambda^2}{\alpha} t^2 + 1} \tag{3.6}$$

$$\phi(t) = \frac{\lambda^2 |t|}{\frac{\lambda^2}{\alpha} |t| + 1} \tag{3.7}$$

whose graphical representations are reported in Figures 1 and 2, respectively, for given λ and α. The use of (3.6) has been reported in the literature, with application to single photon emission tomography [15], [16]. When used in (3.4), both (3.6) and (3.7) encourage neighbouring pixels to be of similar value until their differences are lower than a certain threshold. Beyond this value, further increase in these differences is allowed, at a relatively small increase in the penalty. The differences between neighbouring pixels within uniform regions are thus penalized without excessively penalizing the larger differences occurring at the boundaries between different regions of the image. Geman and Reynolds evidenced that this effect is actually due to the fact that such neighbour interacting functions somehow implicitly refer to a line process, which can be made explicit by deriving the forms of ξ and ψ. In particular they found that, for (3.6), the corresponding ξ and ψ are given by

$$\xi(b)=b \tag{3.8}$$

$$\psi(b)=\alpha\,(1-2\sqrt{b}+1), \quad 0\le b\le 1 \tag{3.9}$$

from which b^*, the minimizer of (3.5), results

$$b^* = \frac{1}{\left(\frac{\lambda^2}{\alpha} t^2 + 1\right)^2} \tag{3.10}$$

For (3.7), they computed

$$\xi(b)= \frac{\lambda^2 b^{3/2}}{2\alpha(1-\sqrt{b})} \tag{3.11}$$

$$\psi(b)= \alpha\frac{b - 3\sqrt{b} + 2}{2}, \quad 0\le b\le 1 \tag{3.12}$$

and

$$b^* = \frac{1}{\left(\frac{\lambda^2}{\alpha}|t| + 1\right)^2} \tag{3.13}$$

The details of the derivations of these formulas can be found in [5] and [17]. Here, it is to be noted that the quantity α/λ^2 can be interpreted as a sort of threshold on the difference t between two adjacent pixels, above which the value of the line element b rapidly decreases, thus permitting a discontinuity in the image to be created.

When $\phi(t)$ is given as the truncated quadratic

$$\phi(t) = \begin{cases} \lambda^2 t^2 & \text{if } |t| < \sqrt{\alpha}/\lambda \\ \alpha & \text{otherwise} \end{cases} \tag{3.14}$$

whose graphical representation is given in Figure 3, for given values of λ and α, Geman and Reynolds showed that (3.5) is satisfied with

$$\xi(b)=b \tag{3.15}$$

$$\psi(b)= \alpha(1-b), \qquad 0\le b\le 1 \tag{3.16}$$

The infimum in (3.5) is achieved at b=1 whenever $|t| < \sqrt{\alpha}/\lambda$ and at b=0 whenever $|t| \ge \sqrt{\alpha}/\lambda$, that is

$$b^* = \begin{cases} 1 & \text{if } |t| < \sqrt{\alpha}/\lambda \\ 0 & \text{if } |t| \ge \sqrt{\alpha}/\lambda \end{cases} \tag{3.17}$$

The line process is thus here binary. If b is either $(1-v_{i,j})$ or $(1-h_{i,j})$, then the corresponding dual energy of eq. (2.5) results to be:

$$E(f,h,v) = \frac{\|g-Hf\|^2}{2\sigma^2} + \lambda^2 \sum_{i,j} (f_{i,j}-f_{i,j-1})^2 (1-v_{i,j}) +$$

$$+ \lambda^2 \sum_{i,j} (f_{i,j}-f_{i-1,j})^2 (1-h_{i,j}) + \alpha \sum_{i,j} h_{i,j} + \alpha \sum_{i,j} v_{i,j} \qquad (3.18)$$

that exactly corresponds to the Weak Membrane energy of Blake and Zisserman, if H=I [3].

For the three neighbour interacting functions analyzed above (eqs. (3.6), (3.7) and (3.14)), the primal energy function of eq. (3.4) is still non-convex, so that adequate algorithms must be devised to obtain satisfactory reconstructions. In this paper, with reference to the neighbour interacting function (3.7), we compare the performance of two different algorithms for the MAP estimate: a stochastic relaxation algorithm, based on simulated annealing, and a deterministic, GNC-type, algorithm, based on the construction of a family of functions that approximate the primal energy F(f).

4. Stochastic Relaxation

In [5], Geman and Reynolds studied the performance of the neighbour interacting function of eq. (3.7) with application to the restoration of images from blurred and noisy data. For a first order image model, the primal energy F of eq. (3.4), in consideration of (3.7), takes the form:

$$F(f) = \frac{\|g-Hf\|^2}{2\sigma^2} + \sum_{i,j} \frac{\alpha\,|f_{i,j}-f_{i,j-1}|}{|f_{i,j}-f_{i,j-1}| + \frac{\alpha}{\lambda^2}} +$$

$$+ \sum_{i,j} \frac{\alpha\,|f_{i,j}-f_{i-1,j}|}{|f_{i,j}-f_{i-1,j}| + \frac{\alpha}{\lambda^2}} \qquad (4.1)$$

To obtain the MAP estimate, they employed an optimization technique based on simulated annealing that is coarse-to-fine in the order of the model. Starting at data and using the first order model, reliable discontinuities are reconstructed even if the fine geometric structure of the smooth regions is missing; this first solution provides a better starting point for the second-order model than the data itself, in such a way that the existing discontinuities are preserved and planar and quadratic patches are recovered from the higher order models.

Stochastic relaxation algorithms with simulated annealing have the desirable feature of converging to a global minimum of any non-convex function [1]. However this is an asymptotic behaviour, so that there is no guarantee to escape from local minima with a finite amount of computation. Despite these recognized limitations, Geman and Reynolds still claim the superiority of stochastic relaxation over other existing methods, and propose a simplifying linear decay for the temperature, coupled with an approximated Gibbs sampler algorithm [1]. In practice, when updating the value of f at site s, instead of sampling from the actual conditional distribution of f_s, they reduce the support of the distribution to the values obtained by taking the union of small intervals around the current value at site s, the current values at the neighbours of s, and the data value g_s. They affirm that this approximation yields an order of magnitude decrease in the number of operations performed, thus giving considerable speed-up, with little apparent degradation in the quality of the reconstructed images.

With reference to the primal energy of eq. (4.1), we adopted a stochastic relaxation algorithm similar to that described by Geman and Reynolds, but employing a Metropolis algorithm [8] to update the single pixels [18]. For any temperature T_k of the annealing schedule, starting from an initial estimate $f^{(0)}$ of the image, at each step j the new estimate $f^{(j)}$ is computed from $f^{(j-1)}$ randomly updating the value of a single, predetermined pixel. Given $\Delta F = F(f^{(j-1)}) - F(f^{(j)})$, the configuration change is accepted if $\Delta F>0$ or, when $\Delta F\leq 0$, if $\exp(\Delta F/T_k) > \eta$, where η is a value randomly selected in the interval (0,1]. If this operation is

repeated indefinitely, visiting the entire set of pixels in some predetermined manner, the algorithm generates a Markov chain $f^{(j)}$, j=0,1,.. that converges to an equilibrium state, whose probability is given by $\pi_k(f)=\exp\left(-F(f)/T_k\right)$. As $T_k \rightarrow 0$ the stationary probability distribution of the Markov chain converges to the uniform probability measure over the set of global minima for F.

In order to define the annealing schedule we need to choose:

(i) an initial value T_0 for the temperature;

(ii) the number of steps of the Metropolis algorithm, performed for each fixed temperature, that is the length of the Markov chains;

(iii) a decay function for the temperature;

(iv) a stop criterion.

We adopted an annealing schedule that makes reference to the one proposed by Aarts and Van Laarhoven [19], with regard to points (i) and (ii), and to the one proposed by Kirkpatrick et al. [9], with regard to points (iii) and (iv).

Let us analyze point (i). The initial temperature value T_0 must be high enough to permit the configuration changes to be accepted with high probability. If we define:

$$\chi(T) = \frac{\text{number of accepted transitions}}{\text{number of proposed transitions}} \tag{4.2}$$

it must be $\chi(T_0)\approx 1$. Let us assume to execute a sequence of configuration changes for a given temperature T. Let x_1 be the number of transitions that decrease the energy and x_2 the number of transitions that increase the energy; moreover, let $<\Delta F>^+$ be the mean value of ΔF for these latter transitions. It is:

$$\chi(T) \approx \frac{x_1 + x_2 \exp\left[-\frac{<\Delta F>^+}{T}\right]}{x_1 + x_2} \tag{4.3}$$

from which:

$$T = \frac{\langle \Delta F \rangle^{+}}{\ln \left[\frac{x_2}{x_2 \chi - x_1 (1 - \chi)} \right]} \qquad (4.4)$$

A sequence of configuration changes is then executed for evaluating x_1 , x_2 and $\langle \Delta F \rangle^{+}$. The value of T_0 is computed from eq. (4.4), with χ fixed to a constant close to 1.

Let us now analyze point (ii). If the successive decrements of the temperature are small enough, a small number of transitions are sufficient to restore the quasi equilibrium. On this basis, we assume that, for each temperature, a small number of complete sweeps of the image are sufficient.

With regard to point (iii), we adopt a linear decay schedule for the temperature, that is:

$$T_{k+1} = \gamma \, T_k \qquad (4.5)$$

where γ is a constant, typically chosen between 0.88 and 0.9 [19].

We consider the algorithm ended (point (iv)) when the value of the cost function at the end of each Markov chain remains almost constant for a number of consecutive times.

5. Graduated Non-Convexity

A different, fully deterministic approach for the minimization of function (4.1) is to construct a family of approximating functions $F^{(p)}$, to be iteratively minimized according to the GNC strategy [3], [12]. Letting parameter p to vary in the interval [0,p*], these functions must satisfy the conditions that $F^{(0)}=F$ and $F^{(p^*)}$ is convex. The construction of the $F^{(p)}$s is usually performed by approximating the neighbour interacting function ϕ. For function $\phi(t)$ of eq. (3.7) we adopted the following approximations:

$$\phi^{(p)}(t) = \begin{cases} \dfrac{\lambda^2 |t|}{\frac{\lambda^2}{\alpha} |t| + 1} & \text{if } |t| \geq p \\ r\, t^2 + q & \text{otherwise} \end{cases} \qquad (5.1)$$

with

$$r = \frac{\alpha^2\lambda^2}{2p(\lambda^2 p+\alpha)^2} \quad ; \qquad q = \frac{\alpha\,\lambda^2 p - r\,p^2(\lambda^2 p+\alpha)}{\lambda^2 p+\alpha} \tag{5.2}$$

In Figure 4, a graphical representation of function (5.1) is reported, for given values of α, λ and p. It is straightforward to verify that, for p=0, eq. (5.1) reduces to eq. (3.7), so that condition $F^{(0)}=F$ is verified. An important issue in the GNC algorithm is the search for a value p* such that the resulting $F^{(p^*)}$ is convex. In [3], when H=I, this is done "balancing" the positive second derivative of $\|g-f\|^2/(2\sigma^2)$ against the negative second derivative of $\phi^{(p^*)}$. If $\phi^{(p^*)}$ is designed to satisfy:

$$\frac{\partial^2\phi^{(p^*)}(t)}{\partial t^2} \geq -c^* \quad \forall\ t \tag{5.3}$$

where $0< c^*< 1/(8\sigma^2)$, then the Hessian of $F^{(p^*)}$ is positive definite. In practice, the value c* can be chosen so that $\phi^{(p^*)}$ is as close as possible to ϕ, which leads to $c^*=1/(8\sigma^2)$. The application of this criterion to the neighbour interacting functions (5.1) leads to the condition [17]

$$2\,\frac{\lambda^2}{\alpha} \leq c^* \tag{5.4}$$

If λ and α can be chosen in such a way to verify this condition, then all the approximations $F^{(p)}$ are convex, $\forall\ p > 0$. For generic values of λ and α, it is possible to show that a value p* exists such that

$$p^* = \frac{\sqrt[3]{2\,\dfrac{\lambda^4}{\alpha c^*}} - 1}{\dfrac{\lambda^2}{\alpha}} \tag{5.5}$$

and for which the approximations $F^{(p)}$ are convex, for $p \in [p^*,+\infty]$ [17]. The GNC algorithm begins by minimizing $F^{(p^*)}$. Then p is decreased from p* to 0, which makes $\phi^{(p)}$ change steadily from $\phi^{(p^*)}$ to ϕ. For every p we minimize $F^{(p)}$ starting with the local

minimum obtained for the previous approximation. While in theory p should decrease continuously from p^* to 0, in practice, successive reductions $p \rightarrow p/2$ have been found to be quite acceptable. There are numerous ways to minimize each $F^{(p)}$. Blake and Zisserman proposed a direct descent algorithm, called Successive Over-Relaxation (SOR), while other authors propose the use of the optimal step conjugate gradient algorithm [11]. The results shown in this paper are obtained by means of a standard conjugate gradient algorithm.

6. Experimental results

The performance of the neighbour interacting function of eq. (3.7) was tested by minimizing the resulting primal energy of eq. (4.1) both via the stochastic relaxation algorithm and the GNC-type algorithm described in Section 4 and 5, respectively. As original images we considered both real and synthetic piecewise smooth images. The degraded images were obtained randomly selecting a percentage of the original images and adding an uncorrelated Gaussian noise of zero mean and variance σ^2. The intensity process was considered continuous for the GNC algorithm and quantized in 256 gray levels for the stochastic relaxation algorithm.

The value of the regularization parameter λ was chosen so as to balance the degree of smoothness and the consistency with the data in the reconstructed image. The most popular mathematical methods for estimating λ are the generalized cross validation method introduced by Wahba [20] and the standard regularization method described by Tikhonov [21]. The value of α should be chosen on the basis of the effective minimum value of the horizontal and vertical gradients, corresponding to the discontinuities across the true images. In general, an exact value of α is not available for real images; an approximate estimate can however be obtained on the basis of a priori information on the class of images to be reconstructed. It is to be noted that the best choice for α should be to consider it as a dynamic parameter, whose value changes over the different regions of the image.

The set of parameters that define the annealing schedule in the stochastic relaxation algorithm has been derived

experimentally, on the basis of the images treated. In particular, for all the trials, the initial temperature T_0 was chosen according to formula (4.4), with $\chi = 0.85$ and $x_1 + x_2 = N^2$, where N^2 is the dimension of the problem; the length of the Markov chains has been fixed to $30N^2$; the temperature has been decreased according to eq. (4.5), with $\gamma = 0.9$ and, finally, as stop criterion it has been assumed that the difference in cost between the last images of two consecutive Markov chains must remain lower than 5000 for 20 consecutive times.

In the GNC algorithm, the starting value p* of parameter p has been chosen according to inequality (5.5), in order to make the corresponding approximation of F convex. The schedule adopted for decreasing p was the following:

$$p_k = p^* - \frac{p^*}{k_{max}-1}(k-1) \qquad k=1,2,\ldots,k_{max} \qquad (6.1)$$

where k is the current iteration and k_{max} is the maximum number of iterations. In particular, we set $k_{max}=[p^*]$.

For each degraded image, the two algorithms have been applied with the same regularization parameters λ and α. The quality of the reconstructed images was evaluated by computing the root mean squared error (MSE) between the reconstruction itself and the original image; the value of the cost function in the reconstructed images has also been computed, as a measure of convergence for the algorithms.

The results of the trials performed are reported in a set of figures organized as follows: the top-left image is the original image, the top-right is the degraded image, the bottom-left is the stochastic relaxation reconstruction and the bottom-right is the GNC reconstruction. We considered three different images and four kinds of degradations. The first image is a synthetic piecewise smooth image of size 128×128, that we will refer as "Line"; the second image is a real image of printed characters, still of size 128×128, which can be roughly considered piecewise smooth and will be referred to as "Letters"; the third image is a face image (of the painter Caravaggio) of size 200×200, referred to as "Face"; this last image is hardly modellable as a piecewise smooth image, so that it serves to test the performance of the

adopted stabilizer when applied to images of general type. As first kind of degradation we considered the addition to the image of uncorrelated Gaussian noise of standard deviation $\sigma=12$; in the second kind of degradation σ is increased to 25; in the third and fourth kind of degradation we randomly select 50% of the original image and add uncorrelated Gaussian noise, first with $\sigma=12$ and then with $\sigma=25$. In Table I, to each degraded image is associated the corresponding figure. The values of the regularization parameters adopted for reconstructing a given image are reported in the caption of the corresponding figure, togheter with the obtained root mean squared errors.

This set of experiments allows to make some comparison between the two algorithms. In the most part of the cases, we found that the value of the cost functional in the reconstructed image is lower for stochastic relaxation than for GNC. From a theoretical point of view, this could mean that stochastic relaxation approaches closer the effective global minimum of the energy. A confirmation of this hypothesis can be found in the qualitative analysis of the results. Indeed, it is possible to note that the images reconstructed with stochastic relaxation are generally more "stylized" than those produced by the GNC, that is they fit more closely the a priori information of piecewise smoothness expressed by the adopted stabilizer. This effect is particularly evident in the "Face" image. Indeed, as already said, this image cannot be adequately described through local smoothness constraints on its intensity values, and would require higher order derivative models. With regard to the values obtained for the root mean squared errors, we can summarize that, for both the two algorithms, they increase with the level of degradation and with the level of misfit between the a priori model adopted and the actual underlying image model. Thus, for the "Line" image, the root mean squared errors are lower than for the "Face" image. According to what observed about the closer fitting of the stochastic relaxation reconstructions with the a priori information, it can happen that, especially for the "Letters" and "Face" images, the root mean squared error results lower with the GNC. With regard to the computational time, for all the trials performed, and for the chosen annealing schedule, the stochastic

relaxation algorithm takes about 1000 CPU seconds to run on an IBM-3081, for a 128×128 image. The computational time of the GNC algorithm is dependent on the number of iterations executed, other than on the image size; for a 128×128 image and 10 iterations the algorithm stops after about 400 seconds of CPU.

7. Final remarks

It should be observed that both the two algorithms are naturally suitable for a parallel implementation that would greatly decrease the computation time. A parallel version of the GNC has been proposed by Blake and Zisserman [3], when the various approximations $F^{(p)}$ of the primal energy F are minimized according to the Successive Over-Relaxation (SOR) algorithm. This is a gradient descent algorithm that uses local quadratic approximations to determine optimal step sizes. In our case, for the iterative minimization of $F^{(p)}$, the n^{th} iteration is:

$$f_{i,j}^{(n+1)} = f_{i,j}^{(n)} - \omega \frac{1}{T_{i,j}} \frac{\partial F^{(p)}}{\partial f_{i,j}} \qquad (7.1)$$

where ω is the "SOR parameter", governing convergence speed, and $T_{i,j}$ is an upper bound on the second derivative.

From the form of $F^{(p)}$ it is straightforward to observe that its derivative with respect to $f_{i,j}$ only depends on four neighbouring pixels, so that a parallel version of the iterative scheme (7.1) could be easily obtained; Blake and Zisserman propose a "chequer-board" updating scheme and prove that both successive and simultaneous schemes are convergent for $\omega \in (0,2)$.

With regard to a possible parallel implementation of stochastic relaxation it can be observed that, at each step, for each pixel, the computation of ΔF only depends on interactions between the pixel itself to be updated and its four adjacent pixels; a parallel implementation can thus be obtained simultaneously executing configuration changes for those pixels that are not adjacent.

References

1 **Geman, S and Geman, D** 'Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images' *IEEE Trans. Pattern Anal. Machine Intell.* Vol 6 (1984) pp 721-741.

2 **Bedini, L and Tonazzini, A** 'Image Restoration Preserving Discontinuities: the Bayesian Approach and Neural Networks' *Image and Vision Computing* Vol 10 no 2 (March 1992) pp 108-118.

3 **Blake, A and Zisserman, A** *Visual Reconstruction* The MIT Press, Cambridge, Massachusetts (1987).

4 **Geiger, D and Girosi, F** 'Parallel and deterministic algorithms for MRFs: surface reconstruction', *IEEE Trans. Pattern Anal. Machine Intell.* Vol 13 no 5 (1991) pp 401-412.

5 **Geman, D and Reynolds, G** 'Constrained Restoration and the Recovery of Discontinuities', *IEEE Trans. Pattern Anal. Machine Intell.* Vol 14 no 3 (1992) pp 367-383.

6 **Marroquin, J L** 'Surface Reconstruction Preserving Discontinuities' Artificial Intell. Lab., AI Memo 792 (August 1984).

7 **Blake, A** 'Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction' *IEEE Trans. Pattern Anal. Machine Intell.* Vol 11 (1989) pp 2-12.

8 **Metropolis, N, Rosenbluth, A W, Rosenbluth, M N and Teller, E** 'Equations of state calculations by fast computing machines' *J. Chem. Phys.* Vol 21 (1953) pp 1087-1091.

9 **Kirkpatrick, S, Gellatt, C D and Vecchi, M P** 'Optimisation by simulated annealing *Science* Vol 220 (1983) pp 671-680.

10 **Hopfield, J J** 'Neurons with graded response have collective computational properties like those of two-state neurons' *Proc. Natl. Acad. Sci. USA* Vol 81 (1984) pp 3088-3092.

11 **Chellappa, R, Simchony, T and Lichtenstein, Z** "Image Estimation using 2D noncausal gauss-Markov random Field Models" in *Digital Image Restoration*, A.K.Katsaggelos ed. Springer-Verlag Berlin (1991).

12 **Blake, A and Zisserman, A** 'Localising discontinuities using weak continuity constraints' *Pattern Recogn. Letters* Vol 6 (1987) pp 51-59.

13 **Geiger, D and Girosi, F** 'Parallel and Deterministic Algorithms for MRFs: Surface Reconstruction and Integration' Artificial Intell. Lab., MIT, AI Memo 1114 (May 1989).

14 **Marroquin, J L, Mitter, S and Poggio, T** 'Probabilistic Solution of Ill-posed Problems in Computational Vision", *Journal of the American Statistical Association* Vol 82 no 397 Theory and Methods (March 1987) pp 76-89.

15 **Hebert, T and Leahy, R** 'A Generalized EM Algorithm for 3-D Bayesian Reconstruction from Poisson Data Using Gibbs Priors', *IEEE Trans. Med. Imag.*, Vol 8, No 2, 1989, pp 194-202.

16 **Geman, S and McClure, D E** 'Bayesian Image Analysis: an Application to Single Photon Emission Tomography', *Proc. Am. Stat. Ass., Stat. Comp. Sect.*, 1985, pp 12-18.

17 **Gerace, I** 'Algoritmi Deterministici per la Ricostruzione di Immagini che presentano Discontinuità' Tesi di Laurea, Università degli Studi di Pisa, Facoltà di Scienze Matematiche, Fisiche e Naturali, corso di Laurea in Scienze della Informazione, A.A. 1991-92.

18 **Pepe, M** 'Stime Bayesiane per la ricostruzione di immagini: analisi di alcuni algoritmi stocastici' Tesi di Laurea, Università degli Studi di Pisa, Facoltà di Scienze Matematiche, Fisiche e Naturali, corso di Laurea in Scienze della Informazione, A.A. 1991-92.

19 **Aarts E., Korst J.** *Simulated annealing and Boltzmann machines*, John Wiley & Sons (1989).

20. **Wahba, G** 'Practical Approximate Solutions to Linear Operator Equations when the Data are Noisy', *SIAM J. Numer. Anal.* Vol 14 (1977).

21. **Tikhonov, A N and Arsenin, V Y** *Solution of ill-posed problems* Winston-Wiley, Washington, USA (1977).

22. **Bedini, L, Gerace, I and Tonazzini, A** 'An extended GNC algorithm for image reconstruction with interacting discontinuities', Progetto Finalizzato 'Sistemi Informatici e Calcolo Parallelo', Sottoprogetto 2 Processori Dedicati, Internal Report N. R/2/62, 1992.

Table I

List of Figures

	$\sigma=12$		$\sigma=25$	
	complete data	sparse data	complete data	sparse data
"Line"	Fig. 5	Fig. 6	Fig.7	Fig. 8
"Letters"	Fig. 9	Fig.10	Fig. 11	Fig. 12
"Face"		Fig.13		

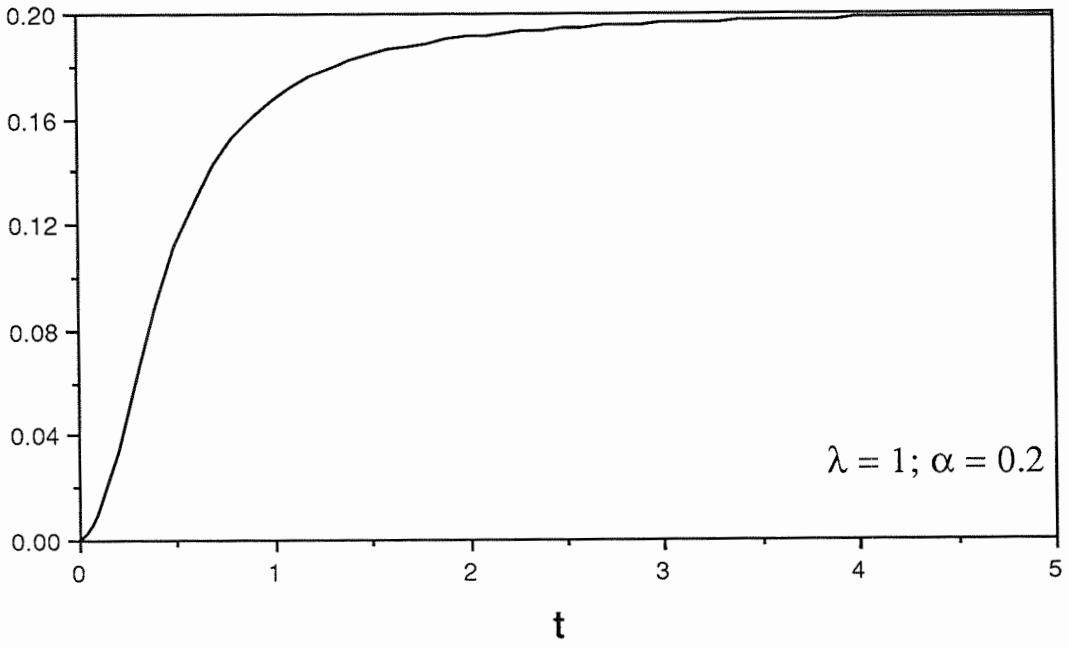

Figure 1 Graphical representation of the neighbour interacting function of eq.(3.6) (λ=1, α=0.2).

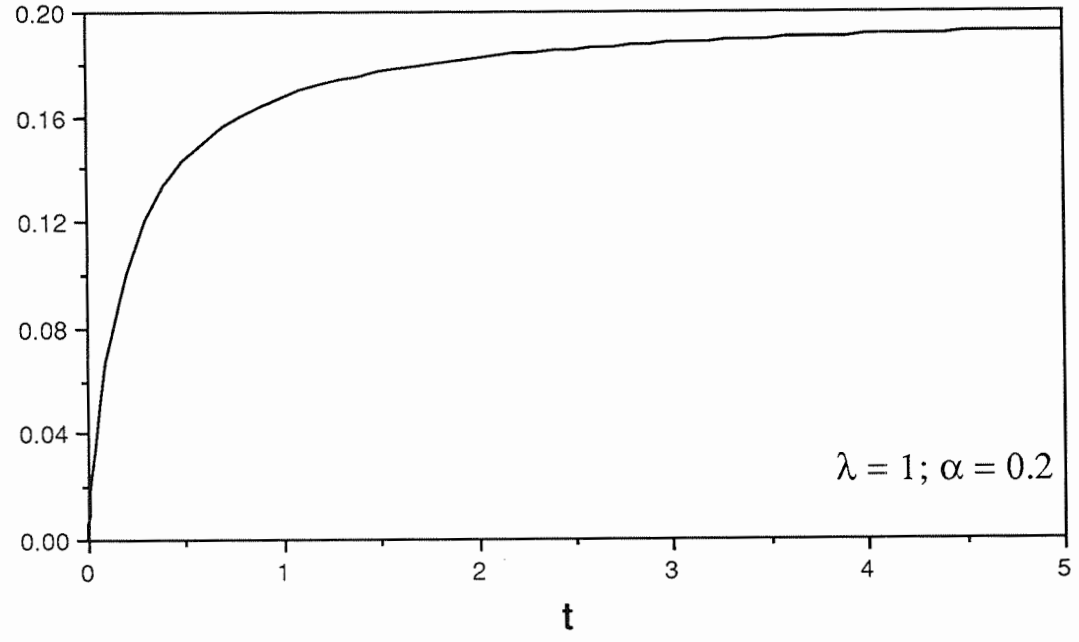

Figure 2 Graphical representation of the neighbour interacting function of eq.(3.7) (λ=1, α=0.2).

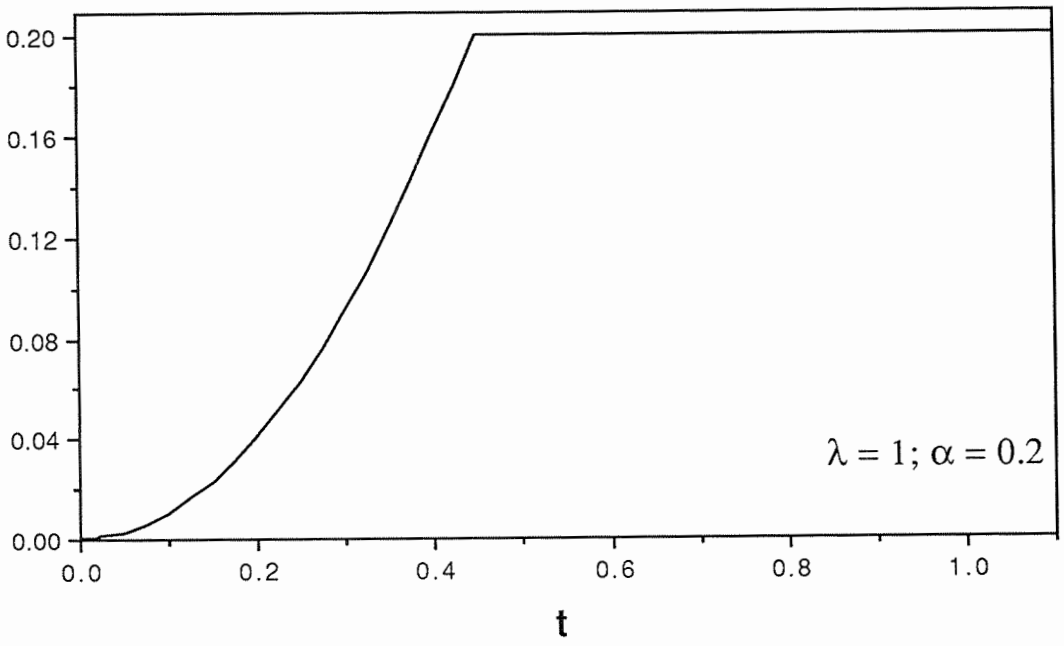

Figure 3 Graphical representation of the neighbour interacting function of eq.(3.14) (λ=1, α=0.2).

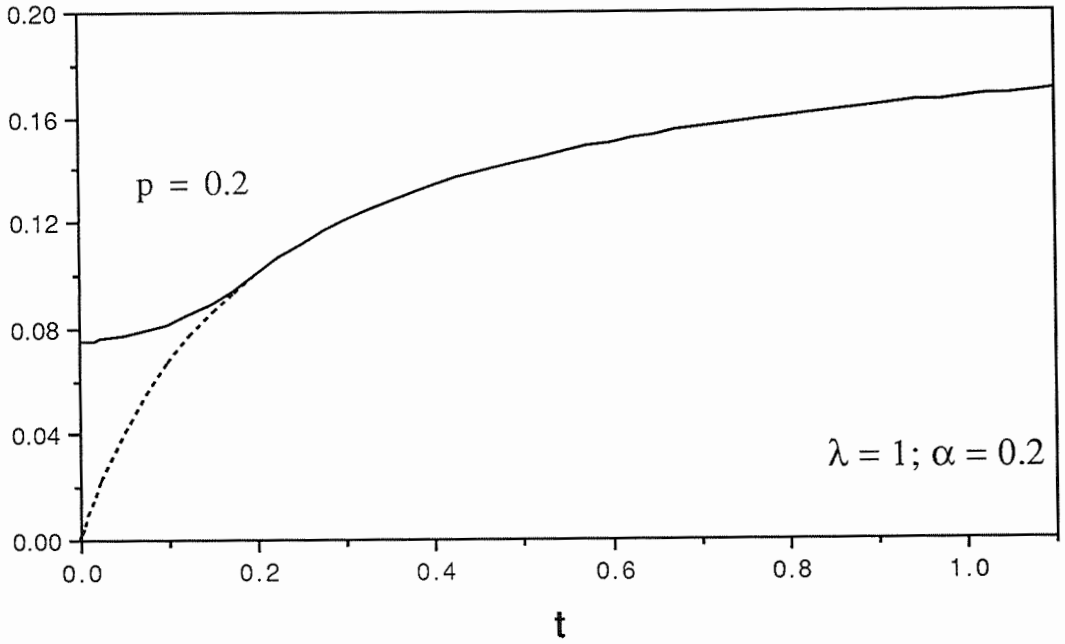

Figure 4 Graphical representation of the neighbour interacting function of eq.(5.1) (λ=1, α=0.2, p=0.2).

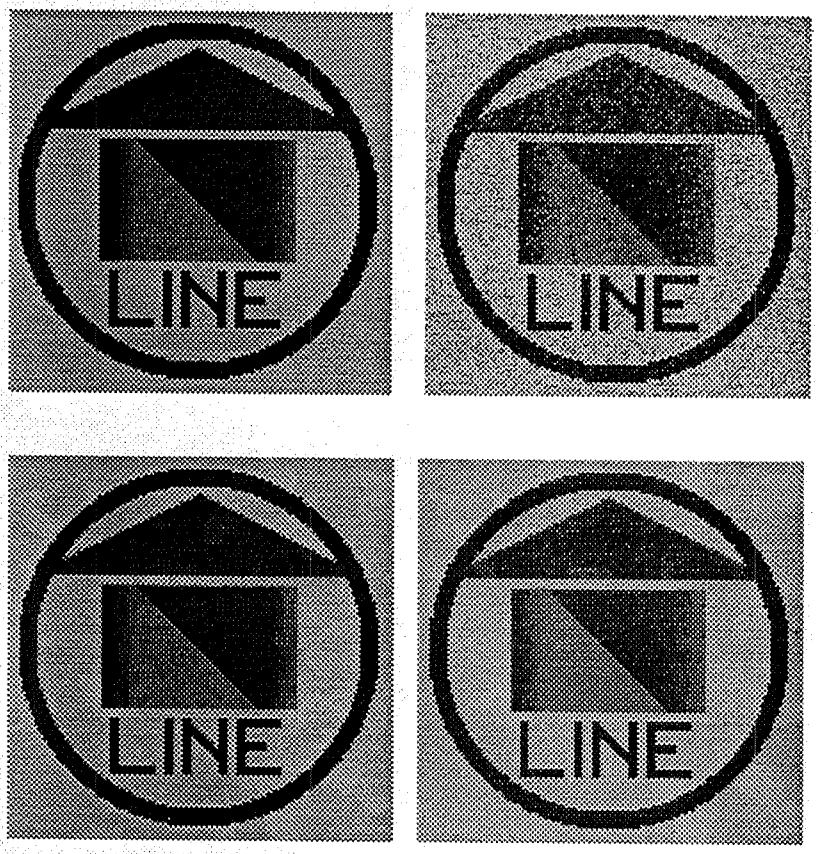

Figure 5: 128×128 "Line" image. Top-left: original image; top-right: image degraded adding noise (σ=12); bottom-left: stochastic relaxation reconstruction (λ=5.5, α=2300, MSE=2.2); bottom-right: GNC reconstruction (λ=5.5, α=2300, MSE=2.2).

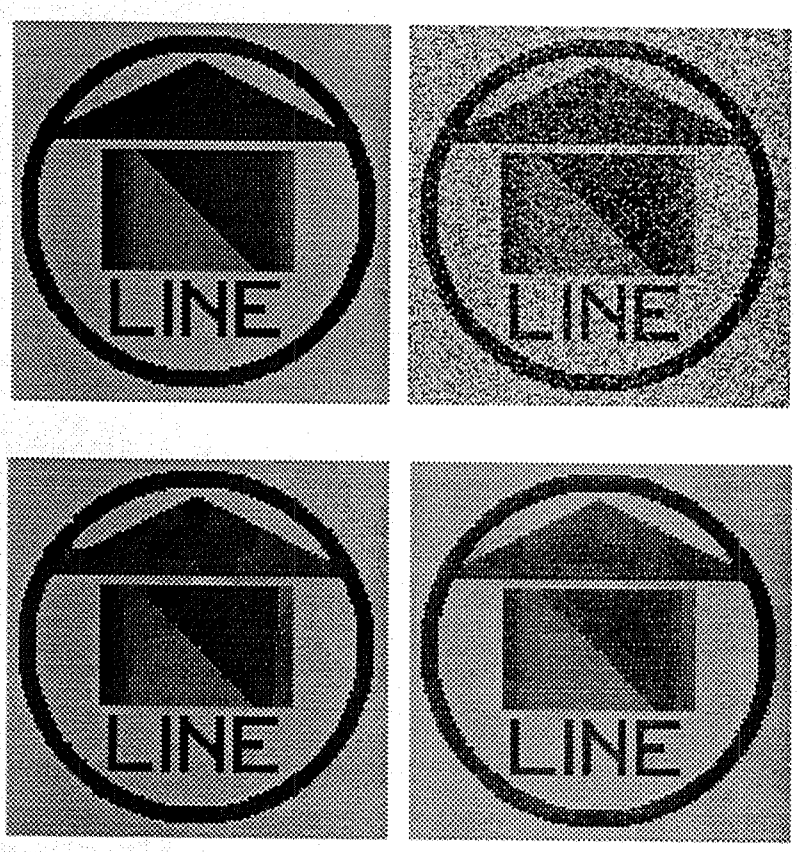

Figure 6: 128×128 "Line" image. Top-left: original image; top-right: image degraded adding noise (σ=25); bottom-left: stochastic relaxation reconstruction (λ=15, α=6000, MSE=5.9); bottom-right: GNC reconstruction (λ=15, α=6000, MSE=5.4).

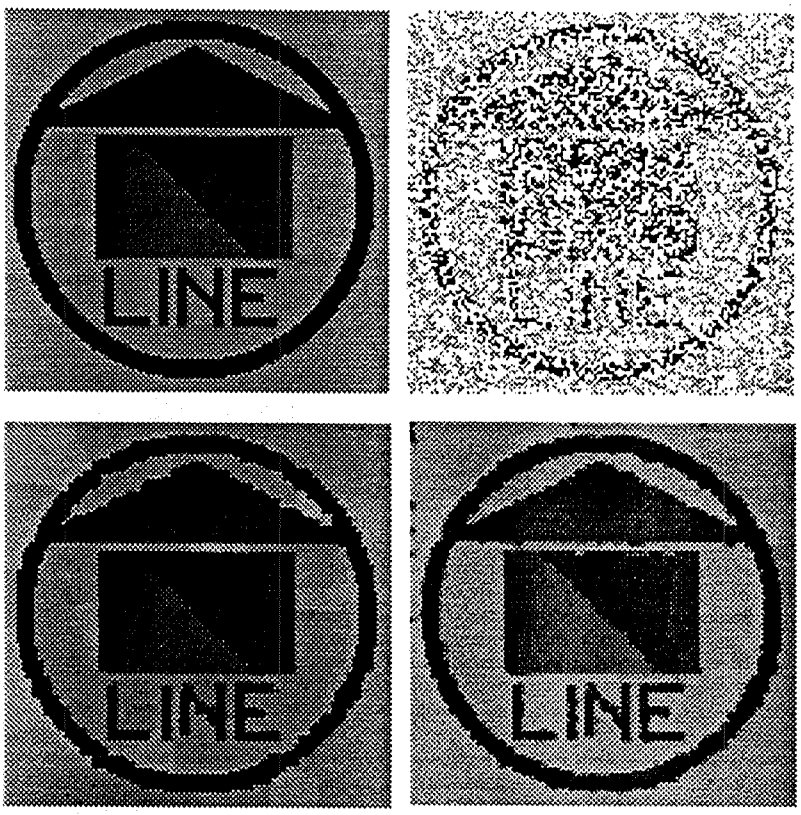

Figure 7: 128×128 "Line" image. Top-left: original image; top-right: randomly selected 50% of the original image plus noise (σ=12) - for display purposes the remaining 50% is filled with white dots; bottom-left: stochastic relaxation reconstruction (λ=20, α=2200, MSE=12.5); bottom-right: GNC reconstruction (λ=20, α=2200, MSE=12.5).

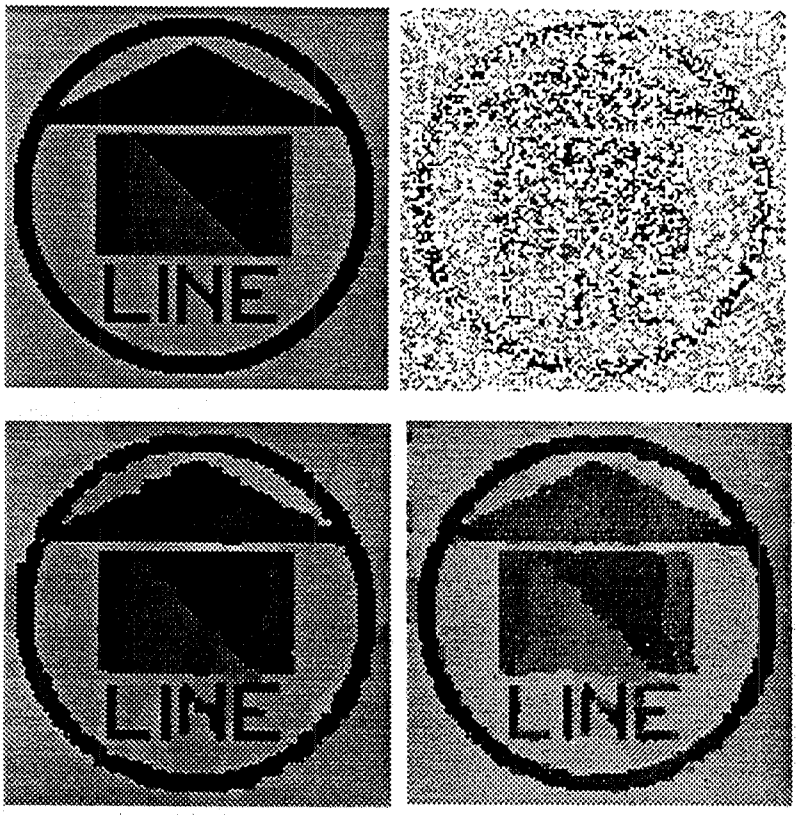

Figure 8: 128×128 "Line" image. Top-left: original image; top-right: randomly selected 50% of the original image plus noise (σ=25) - for display purposes the remaining 50% is filled with white dots; bottom-left: stochastic relaxation reconstruction (λ=15, α=2200, MSE=13.5); bottom-right: GNC reconstruction (λ=15, α=2200, MSE=14.8).

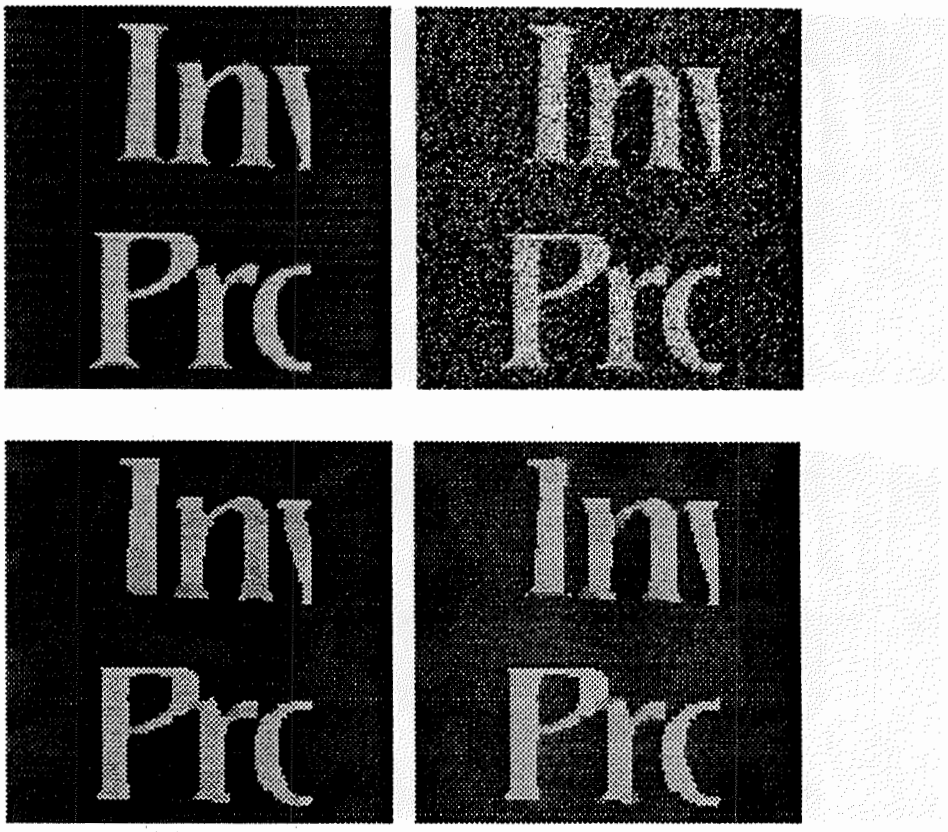

Figure 9: 128×128 "Letters" image. Top-left: original image; top-right: image degraded adding noise (σ=12); bottom-left: stochastic relaxation reconstruction (λ=15, α=4000, MSE=8.0); bottom-right: GNC reconstruction (λ=15, α=4000, MSE=6.1).

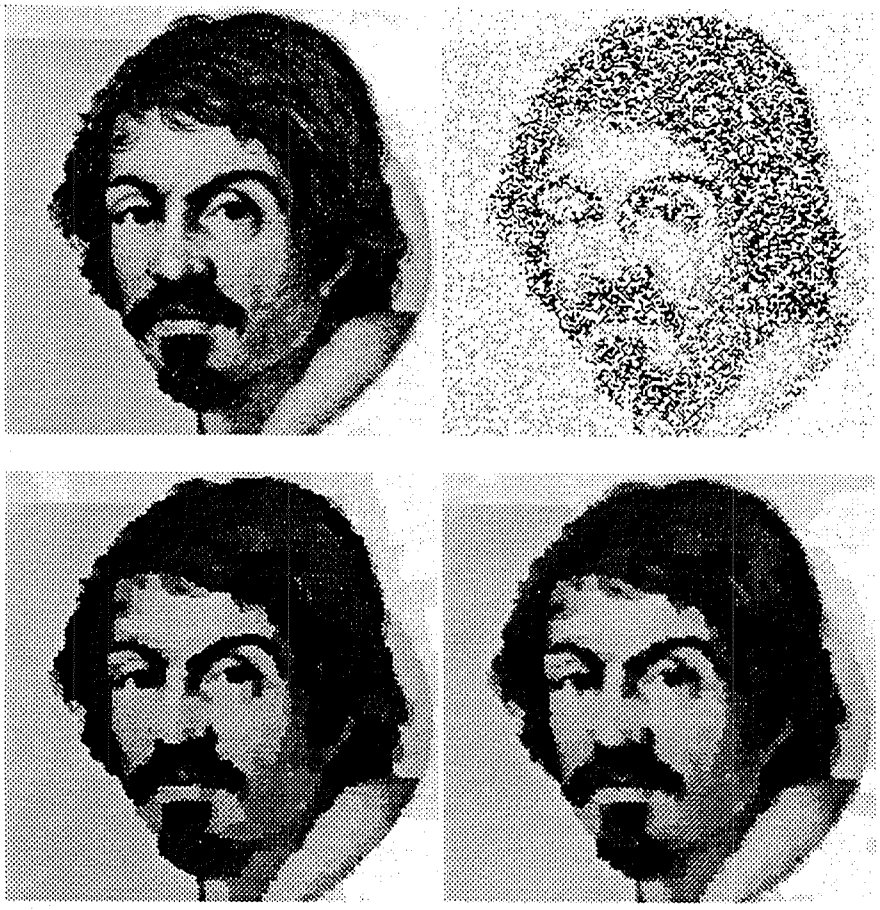

Figure 13: 200×200 "Face" image. Top-left: original image; top-right: randomly selected 50% of the original image plus noise (σ=12) - for display purposes the remaining 50% is filled with white dots; bottom-left: stochastic relaxation reconstruction (λ=10, α=1000, MSE=12.9); bottom-right: GNC reconstruction (λ=10, α=1000, MSE=11.5).