

## IMAGE SELECTIVE SMOOTHING AND EDGE DETECTION BY NONLINEAR DIFFUSION. II\*

LUIS ALVAREZ†, PIERRE-LOUIS LIONS‡, AND JEAN-MICHEL MOREL‡

**Abstract.** A stable algorithm is proposed for image restoration based on the “mean curvature motion” equation. Existence and uniqueness of the “viscosity” solution of the equation are proved, a  $L^\infty$  stable algorithm is given, experimental results are shown, and the subjacent vision model is compared with those introduced recently by several vision researchers. The algorithm presented appears to be the sharpest possible among the multiscale image smoothing methods preserving uniqueness and stability.

**Key words.** mean curvature motion of level sets, image restoration, viscosity solutions, multiscale vision models

**AMS (MOS) subject classifications.** 49F22, 53A10, 82A60, 76T05, 49A50, 49F10, 80A15

**Introduction.** In this paper, we propose and study a class of nonlinear parabolic differential equations for image processing of the following kind:

$$(1) \quad \frac{\partial u}{\partial t} = g(|G * Du|) |Du| \operatorname{div} \frac{Du}{|Du|}, \quad u(0, x, y) = u_0(x, y),$$

where  $u_0(x, y)$  is the grey level of the image to be processed,  $u(t, x, y)$  is its smoothed version depending on the “scale parameter”  $t$ ,  $G$  is a smoothing kernel (for instance, a Gaussian),  $G * Du$  is therefore a local estimate of  $Du$  for noise elimination, and  $g(s)$  is a nonincreasing real function which tends to zero as  $s \rightarrow \infty$ . Roughly speaking, the interpretation of the terms of the equation are as follows.

(a) The term  $|Du| \operatorname{div} (Du/|Du|) = \Delta u - D^2u(Du, Du)/|Du|^2$  represents a degenerate diffusion term, which diffuses  $u$  in the direction orthogonal to its gradient  $Du$  and does not diffuse at all in the direction of  $Du$ . (Here and everywhere below,  $D^2u$  denotes the Hessian of  $u$ .) The aim of the degenerate diffusion term is to make  $u$  smooth on both sides of an “edge” with a minimal smoothing of the edge itself. (An edge is defined as a line along which the gradient is “large.”)

(b) The term  $g(|G * Du|)$  is used for the “enhancement” of the edges. Indeed, it controls the speed of the diffusion: if  $Du$  has a small mean in a neighborhood of a point  $x$ , this point  $x$  is considered the interior point of a smooth region of the image and the diffusion is therefore strong. If  $Du$  has a large mean value on the neighborhood of  $x$ ,  $x$  is considered an edge point and the diffusion spread is lowered, since  $g(s)$  is small for large  $s$ .

Thus, the proposed model is a selective smoothing of the image, where the “edges” are relatively enhanced and preserved as much as possible.

The present model may seem complicated. However, it has the minimal number of parameters required by any image processing model: a “contrast” function, represented by  $g$ , which allows us to decide whether a detail is sharp enough to be kept, and a “scale parameter,” given by the variance of  $G$ , which fixes the minimal size of the kept details in the processed picture.

\* Received by the editors December 19, 1990; accepted for publication (in revised form) May 24, 1991. This work has been partially supported by U.S. Army contract DAJA45-88-C-0009.

† Departamento de Informatica y Sistemas, Universidad de Las Palmas, Avenida Maritima del Sur s/n, 35016 Las Palmas, Spain.

‡ Ceremade, Université Paris-Dauphine, Place de Lattre de Tassigny, 75775 Paris Cedex 16, France.

In many applications, these parameters can remain constant for all processed images.

On the other side, this model generalizes or specializes most of the models that have been proposed for image smoothing and edge detection. As we shall see, among the partial differential equation (PDE) models, this model seems to be at the frontier of the stable models which preserve  $L^\infty$  norms of the image and can therefore be associated with efficient numerical schemes.

In § 1, we shall first present the new aspect of the model: the degenerate diffusion. We shall then present the whole model and discuss its relation to the classical models of image processing. In § 2, we present an approximate model whose numerical analysis will be easier. In § 3 we prove the mathematical validity of the model and of its approximated one. In other words, we prove the existence and uniqueness of the solution of the associated parabolic equation. Finally, in § 4, we show the numerical scheme and some experimental results.

## 1. The model.

**1.1. Quasilinear anisotropic models of edge detection.** The “low level” analysis of images presents two opposite requirements. It is generally desirable to smooth the homogeneous regions of the picture with two scopes: noise elimination and image interpretation.

On the other side, we wish to keep the accurate location of the boundaries of these regions. Those boundaries are called “step edges” [17]. In the classical theory, these aims are achieved by a previous low pass filtering [27], [29], [30]. Then the edges are defined as the curves where the gradient of the smoothed picture has a maximum. (The set of “edges” is therefore contained in the set of the points where the Laplacian of the smoothed signal changes sign.)

This theory comes from Marr and Hildreth [18] and has been improved by Witkin [29], Koenderink [13], and Canny [3]. The low pass filtering is generally made by convolution with Gaussians of increasing variance. It is easy to understand the necessity of a previous low pass filtering: if the signal is noisy, the gradient will have a lot of irrelevant maxima which must be eliminated. Of course, strong oscillations can be due to different causes, for instance, the presence of “textures.” Koenderink [13] noticed that the convolution of the signal with Gaussians at each scale was equivalent to the solution of the heat equation with the signal as initial datum. Denote this datum by  $u_0$ ; the “scale space” analysis associated with  $u_0$  consists in solving the system

$$(2) \quad \frac{\partial u(x, y, t)}{\partial t} = \Delta u(x, y, t), \quad u(x, y, 0) = u_0(x, y).$$

The solution of this equation for an initial datum with bounded quadratic norm is  $u(x, y, t) = G_t * u_0$ , where

$$G_\sigma(x, y) = C\sigma^{-1} \exp(-(x^2 + y^2)/4\sigma)$$

is the Gauss function.

Then  $(x, y)$  is an edge point for the “scale”  $t$  at points where  $\Delta u(x, y, t)$  changes sign and  $|Du(x, y, t)|$  is “big.” Of course, this last condition introduces some a priori defined threshold. Unfortunately, it is well known (and it is enough to look at the “edges” found by this method to observe it [17]) that the edges at low scales give an inexact account of the boundaries which, according to our perception, should be considered correct. This is still true for the low pass filtering of Canny [3], [25], which is generally used as the best linear filter for white noise elimination and edge detection.

On the opposite side, if we make a sharp low filtering, with small variance, all the edges will keep their correct location. Now, the “main” edges will be embedded in a crowd of “spurious” edges due to noise, texture, etc. The theory of Witkin [29] proposes therefore to identify the main edges at a low scale, and then to “follow them backward” by making the scale decreasing again. This method could theoretically give the exact location of all main edges. However, its implementation is rather heavy from the computational viewpoint and unstable, because of the following of edges across scales and the multiple thresholdings involved in the edge detection at each scale.

The intuitive idea of “edges” is that they are generally piecewise smooth. Therefore, it seems natural to modify the diffusion operator so that it diffuses more in the direction parallel to the edge and less in the perpendicular one. In the extreme case, we could think of a diffusion which is made *only* in the direction of the edge: such a diffusion would apparently keep *exactly the location and sharpness of the edge, while smoothing the picture on both sides on this edge*. Let us first consider this limit case in a linear framework. It is not difficult to see that the diffusion equation, which does not diffuse at all in the direction of the gradient  $Du$ , can be written as

$$(3) \quad \frac{\partial u}{\partial t} = \Delta u - \frac{1}{|Du|^2} D^2 u(Du, Du).$$

The first term, the Laplacian, is the same as in scale space theory, and the second is an “inhibition” of the diffusion in the direction of the gradient.

Let us denote by  $\xi$  the coordinate associated with the direction orthogonal to  $Du$ . Therefore a formulation of the preceding equation with respect to this new coordinate is

$$\frac{\partial u}{\partial t} = u_{\xi\xi},$$

where, of course,  $\xi$  depends on  $Du$ .

In a quasi “divergence form,” the equation can also be written as

$$(4) \quad \frac{\partial u}{\partial t} = |Du| \operatorname{div} \frac{Du}{|Du|},$$

and in more literal formulation as

$$(5) \quad \frac{\partial u}{\partial t} = \frac{1}{u_x^2 + u_y^2} (u_y^2 u_{xx} - 2u_x u_y u_{xy} + u_x^2 u_{yy}).$$

This equation has recently received a lot of attention because of its geometrical interpretation; indeed, at least formally (see Osher and Sethian [24] and Evans and Spruck [8]) the level sets of the solution move in the normal direction with a speed proportional to their mean curvature. (This “mean curvature motion” effect will be shown in the experimental results presented below.)

A theory of weak solutions based upon the so-called viscosity solution theory has been proposed by Chen, Giga, and Goto [5], Evans and Spruck [8], Giga, Goto, Ishii, and Sato [9], and Soner [28].

**1.2. Relation with the Malik and Perona theory.** The preceding idea is, as we shall see, quite close to an important improvement of the edge detection theory proposed by Malik and Perona [25]. Their main idea is to introduce a part of the edge detection step in the filtering itself, allowing an interaction between scales from the beginning of the algorithm.

They propose to replace the heat equation by a nonlinear equation

$$(6) \quad \frac{\partial u}{\partial t} = \operatorname{div} (g(|Du|)Du), \quad u(0) = u_0.$$

In this equation,  $g$  is a smooth nonincreasing function with  $g(0) = 1$ ,  $g(s) \geq 0$ , and  $g(s) \rightarrow 0$  at infinity. The idea is that the smoothing process obtained by the equation is “conditional”: if  $Du(x, y)$  is big, then the diffusion will be low and therefore the exact localization of the “edges” will be kept. If  $Du(x, y)$  is small, then the diffusion will tend to smooth still more around  $(x, y)$ . Thus the choice of  $g$  corresponds to a sort of thresholding which has to be compared to the thresholding of  $|Du|$  used in the final step of the classical theory explained above. Since this thresholding introduced a nonlinear device anyway, it was natural to use it earlier in the method, in the smoothing process itself. The experimental results obtained by Malik and Perona are perceptually impressive and show that an “edge detector” based on this theory gives edges which remain much more stable across the scales, therefore making the backward following of edges across scales unnecessary. (This property has been sought by many researchers in the last two decades.)

However, the Malik and Perona model had several serious, practical, and theoretical difficulties which have been improved in a recent work by some of the authors of this paper [4]. The first difficulty was a straightforward objection that Malik and Perona acknowledged themselves. Assume that the signal is noisy, with white noise for instance. Then the noise introduces very large, in theory unbounded, oscillations of the gradient  $Du$ . Thus, the conditional smoothing introduced by the model will not give good results, since all these noise edges will be kept!

The second difficulty arose from the equation itself; among the functions  $g$  which Malik and Perona consider advisable, we find functions of the type  $g(s) = e^{-s}$  or  $g(s) = (1 + s^2)^{-1}$  for which no correct theory of equation (2) is available. Indeed, in order to obtain both existence and uniqueness of the solutions,  $g$  must verify that  $sg(s)$  is nondecreasing. If this condition is not verified, we can observe, for some functions  $g$  with  $sg(s)$  nonincreasing, a nondeterministic and therefore unstable process; the same picture can in theory be the initial condition of solutions divergent in time (see [11] for simple and explicit examples, and also [7]). *In practice, that means that very close pictures could produce divergent solutions and therefore different edges.* However, it was reasonable to try to define a model where the function  $g$ , which is a sort of thresholding, could be quickly decreasing, as  $e^{-s}$  for instance.

The model which has been proposed in [4] is a synthesis of Malik and Perona’s ideas which avoids the above-mentioned difficulties; it is robust in the presence of noise and consistent from the formal viewpoint mentioned above. Since we shall use this improvement in our final model, let us first explain it.

We shall define the “selective smoothing of  $u_0$  at scale  $t^{1/2}$  based on estimates at the scale  $\sigma$ ” as the function  $u(x, y, t)$ , verifying

$$\begin{aligned} \frac{\partial u}{\partial t} - \operatorname{div} (g(|DG_{\sigma^*}u|)Du) &= 0 \quad \text{in } ]0, T[ \times \Omega, \\ u(0) &= u_0, \end{aligned}$$

where

$$G_{\sigma}(x, y) = C\sigma^{-1} \exp(-(x^2 + y^2)/4\sigma).$$

It is easily seen that  $G(x, y, t) = G_t(x, y)$  is nothing but the fundamental solution of the heat equation. Therefore the term  $(DG_{\sigma^*}u)(x, y, t)$  which appears inside the

divergence term of (3) is simply the gradient of the solution at time  $\sigma$  of the heat equation with  $u(x, y, 0)$  as initial datum. Therefore, it appears to be an estimate of the gradient of  $u$  at point  $(x, y)$ , obtained by the classical Marr-Hildreth-Witkin theory recalled above. Thus, the modification of the model of Malik and Perona is only to replace the gradient  $|Du|$  by its estimate  $|DG_{\sigma} * u|$ . As proved in [4], this slight change of the model is enough to avoid both inconsistencies of the Malik and Perona model. The equation diffuses at a point with more or less strength, according to the estimate of the gradient. This estimate is achieved by the new term. This term, as in Witkin's theory, serves to recognize the location of the main edges. This information is then used in the equation to avoid too much diffusion at these locations.

However, this last model keeps some of the drawbacks of the previous models. First it has no clear geometric interpretation, because the term inside the divergence is hybrid and combines the estimate of the gradient and the gradient. Moreover, even if existence and uniqueness are proved, *the stability of the model as the scale parameter  $\sigma \rightarrow 0$  is generally not true, because the limit model can "invert" the heat equation. As we shall see, these drawbacks are related to the excessive generality of the model.*

Let us now see which equation is given by the combination of the degenerate diffusion and the above-mentioned "estimate method":

$$(7) \quad \frac{\partial u}{\partial t} - g(|DG_{\sigma} * u|)|Du| \operatorname{div} \frac{Du}{|Du|} = 0 \quad \text{in } ]0, T[ \times \Omega,$$

$$u(0) = u_0.$$

Note that if  $g(s) = 1/s$  and  $\sigma = 0$ , we then obtain as a limit case of this new model the equation  $\partial u / \partial t = \operatorname{div} (Du / |Du|)$  which is a particular case of the Malik and Perona model. This limit case is particularly important because it explains the geometric behaviour of the solution, which might be hidden in our general formulation. Indeed, the equation  $\partial u / \partial t = \operatorname{div} (Du / |Du|)$  corresponds to the descent method ( $\partial u / \partial t = -\nabla E(u)$ ), associated with the energy functional:

$$E(u) = \int |Du(x, y)| \, dx \, dy.$$

How can we interpret this energy? A particular case for  $u$ , which is also the ideal case for the detection of "step edges," is when  $u_0 = X_A$  is the characteristic function of some set  $A$  with smooth boundary. Then  $E(u_0)$  is nothing but the perimeter of  $A$ , and the evolution equation can therefore be interpreted as a smoothing of the boundary of this characteristic function which tends to preserve the "edges." This effect is quite visible in the experiments in § 4.

This model is still not quite optimal. Indeed, it is not necessary to diffuse anisotropically at points where the gradient is low. We do not want to enhance, or even to preserve, the edges without contrast. Therefore, rather than (7), we shall prefer the following formulation, which separates the behaviour for large gradients from the behaviour small gradients:

$$(8) \quad \frac{\partial u}{\partial t} - g(|DG_{\sigma} * u|) \left( (1 - h(|Du|)) \Delta u + h(|Du|)|Du| \operatorname{div} \frac{Du}{|Du|} \right) = 0,$$

where  $h(s)$  is a smooth nondecreasing function such that  $h(s) = 0$  if  $s \leq e$ ,  $h(s) = 1$  if  $s \geq 2e$ . The parameter  $e$  is *not* an additional parameter. It is only a refinement of our contrast model. We know that if  $|Du|$  is large,  $g(|Du|)$  is small. Thus  $e$  must depend on the same contrast parameter as  $g$ : it is the upper bound of the interval where  $u$  is allowed to diffuse freely.

In order to understand this further refinement well, let us take the example where  $u_0(x, y) = u_0(x)$  is a cylindrical function. Then the gradient has constant orientation and therefore the pure anisotropic diffusion (7) will simply have no effect on  $u_0$ . It yields  $u(x, y, t) = u_0(x)$  for all  $t$ . The model (8), instead, will act as a Perona and Malik model in dimension 1: it will diffuse at points where the gradient is too low.

**1.3. The role of the scale parameter and its relation with “time.”** As we commented in [4], the function  $G$  to be considered can be any “low pass filter,” or, to use the calculus terminology, any smoothing kernel. However, to preserve the notion of scale in the gradient estimate, it is convenient that this kernel depends on a scale parameter. A good and classical example is, as mentioned above, the Gaussian. It is important to keep this particular case in mind. Indeed, a question which arises immediately in the consideration of model (3) is what time is best for “stopping” the evolution of the signal  $u(x, y, t)$ . Now, we may appeal to the Witkin model to answer this question: according to this model, time is interpreted as a “scale factor.” (More precisely, the solution  $u(x, y, t)$  at time  $t$  corresponds to a scale  $t^{1/2}$ . Indeed, roughly speaking,  $u(x, y, t)$  appears in the Witkin model as a smoothed version of  $u_0$ , obtained by convolving it with a filter of spatial width  $t^{1/2}$ .) Thus in model (3) it is coherent to correlate the stopping time  $t$  and the time introduced via the estimator  $G_\sigma$ . One should therefore choose a stopping time  $t$  of the order of  $\sigma$ . Then the scale above which the signal is smoothed in regular zones of the image will be of the order of  $t^{1/2}$ .

On the parts of the picture where edges are present, the situation is different. Since the scope of the equation is to delay diffusion in these zones, the time at which edge information is lost will depend on the sharpness of the edge and on the shape of the enhancement function  $g$ . Therefore, there is no inconsistency in looking at what happens to the signal  $u(x, y, t)$  for times greater than  $\sigma$ . In experiments, it might be convenient to play with two parameters which are already at hand in any edge detection model: the scale parameter (spatial width of the filtering) on the one side,  $\sigma^{1/2}$ , and the enhancement parameter for edges which is implicit in the shape of  $g$ : if  $g$  is near 1 on some interval containing zero and decreases briskly at the end  $b$  of this interval, then  $b$  is the enhancement parameter for edges: where  $|Du|$  is greater than  $b$ , the edges will remain and where it is smaller, they will disappear. Thus it is clear that if our model is used as a preliminary step for an edge detection device, it would be possible to use the same enhancement parameter  $b$  on the gradient for keeping the edges as the one implicit in  $g$ .

**1.4. Related models.** Before beginning with the proofs of the mathematical validity of our model, let us give a brief account of several related works. Osher and Rudin’s theory [23] tries to get as close as possible to the inverse heat equation by defining some conservative scheme like

$$\frac{\partial u}{\partial t} = -|u_x|F(u_{xx}) \quad \text{with initial value } u_0.$$

$F$  is a function such that  $sF(s) \geq 0$ . The big advantage of this new method is to have a scheme which lets the image develop true edges, that is, shock lines along which  $u(x, y, t)$  becomes discontinuous in  $(x, y)$ .

Nordström [22] proposes a new presentation of Perona and Malik’s theory which relates it to the well-known variational global edge detection methods of Kass, Witkin, and Terzopoulos [12] and Mumford and Shah [20]. Nordström introduces a new term in equation (6) which forces  $u(x, y, t)$  to remain close to  $u_0$ . Because of the forcing

term  $u - u_0$ , the new equation

$$\frac{\partial u}{\partial t} - \operatorname{div} (g(Du) Du) = u_0 - u$$

has the advantage of having a nontrivial steady state, therefore eliminating the problem of choosing a stopping time. (Of course, we can do the same in our model, and the proof below is not altered by this modification.)

The approach of Nitzberg and Shiota [21] is related to adaptive filtering methods which were introduced by Graham [10] for TV images denoising. The idea is to blur, selectively and anisotropically, the signal with “oblong” Gaussians. The Gaussian used for blurring at a point  $(x, y)$  depends on the intensity and direction of the gradient in the neighborhood. Roughly speaking, the blurring will be faster in the direction orthogonal to the gradient. Therefore, the signal will be smoothed on both sides of an edge, but the edge is conserved. Since a corner is the crossing point of two edges, there will be two directions of “nondiffusion” instead of one, and therefore corners are well conserved by this method. Moreover, Nitzberg and Shiota prove by a scaling argument that as the size of the Gaussians tends to zero, their diffusion method tends to some ill-posed partial differential equation analogous to the one by Malik and Perona.

In a recent paper, Mallat and Zhong [16] follow Marr’s suggestion that the edge representation at several dyadic scales should be a complete representation of a picture. They propose an algorithm, based on some stability properties of the wavelet transform, which indeed reconstructs the picture from the edges at five scales. An interesting and suggestive fact of this algorithm is that if the “spurious edges” (e.g., the edges which have low gradient or cannot be prolonged) are removed, then the reconstruction is still almost perfect. This justifies a posteriori the importance given to edge detection and enhancement in the literature. An important application of this algorithm is, as in this paper, selective smoothing. Indeed, these authors are able to clean up the picture by eliminating the “spurious” edges (i.e., the edges which are not stable across several scales) and thereafter by reconstructing the picture. The visual result looks like ours, with some artifact near the main edges, however.

**2. Approximated models.** In order to define a numerical scheme associated with the equation, we must take into account that the image is sampled on a grid. Of course, the discretizations of the differential operators at a point  $(i, j)$  of the grid, for obvious fastness and simplicity reasons, must involve a few other points around it. Typically, we would consider four other points for the discretization of the Laplacian, namely  $(i \pm 1, j)$  and  $(i, j \pm 1)$ . Now, in our case, the differential operator is strongly anisotropic and can hardly be represented by two directions. Denote by  $\xi = -x \sin \theta + y \cos \theta$ , where  $(\cos \theta, \sin \theta) = Du/|Du|$ , the coordinate in the diffusion direction (which is orthogonal to the gradient). The anisotropic term of the equation is therefore  $\partial^2 u / \partial^2 \xi$  and can easily be discretized only if, in the direction  $(-\sin \theta, \cos \theta)$ , one can find points of the grid near  $(i, j)$ . Therefore, the accuracy of the scheme depends strongly on the cardinality of the neighborhood. Anyway, we are led to a new formulation of the equation, which will take into account the discrete number of diffusion directions.

Let  $0 \leq \theta_1^n < \theta_2^n < \dots < \theta_n^n < \pi$  be  $n$  angles and  $x_1^n, \dots, x_n^n$  the coordinates defined by  $x_j^n = -x \sin \theta_j^n + y \cos \theta_j^n$ . In other terms,  $x_j^n$  is the coordinate orthogonal to the direction given by the angle  $\theta_j^n$ . We shall decompose the anisotropic diffusion operator

$$\frac{\partial^2 u}{\partial^2 \xi} = (\sin^2 \theta) \frac{\partial^2 u}{\partial^2 x} - 2(\sin \theta \cos \theta) \frac{\partial^2 u}{\partial x \partial y} + (\cos^2 \theta) \frac{\partial^2 u}{\partial^2 y}$$

into a linear nonnegative combination of the fixed directional diffusion operators  $\partial^2 u / \partial^2 x_j$ .

Consider the degenerate elliptic operator  $A^n$  defined by

$$(9) \quad A^n v = \sum_{j=1, \dots, n} f_j^n \left( \frac{Du}{|Du|} \right) \frac{\partial^2 v}{\partial^2 x_j^n},$$

where the  $f_j^n(\cos \theta, \sin \theta) \geq 0$  are smooth functions designed to be “active” (that is, nonzero) only when  $\theta$  is close to  $\theta_j^n$ .

By rearranging  $A^n$  as a differential operator, with respect to the space variables  $x$  and  $y$ , we can set

$$A^n v = a_n \frac{\partial^2 v}{\partial^2 x} + b_n \frac{\partial^2 v}{\partial x \partial y} + c_n \frac{\partial^2 v}{\partial^2 y}.$$

We shall now state and prove two results which indicate to which extent this kind of operator can approximate our degenerate diffusion.

**PROPOSITION 1.** *Let  $(\theta_1^n, \dots, \theta_n^n)$  and  $(f_1^n, \dots, f_n^n)$  be such that*

- (i)  $\max_k |\theta_k^n - \theta_{k-1}^n| \rightarrow 0$ ,  $\theta_1^n \rightarrow 0$ ,  $\theta_n^n \rightarrow \pi$  as  $n \rightarrow \infty$ ,
- (ii)  $\sum_{j=1, \dots, n} f_j^n \rightarrow 1$  as  $n \rightarrow \infty$ ,
- (iii) *On  $[\theta_{k-1}^n, \theta_k^n]$ , only  $f_{k-1}^n$  and  $f_k^n$  are nonzero.*

*Then for any coordinate  $\xi$ , the coefficients  $a_n$ ,  $b_n$ ,  $c_n$  of the operator  $A^n$  tend to the coefficients of  $\partial^2 u / \partial^2 \xi$ .*

*Proof.* It is obvious. Note that the approximation is stable in the sense that the approximate operator is also elliptic. Thus the meaning of the proposition is that when the number of points of the grid involved in the discretization of the differential operator tends to infinity, the operator tends to represent perfectly all diffusion directions.

**PROPOSITION 2.** *If  $\xi$  is not one of the coordinates  $x_j$ , then no nonnegative coefficients  $\lambda_n$  can be found ensuring that*

$$\frac{\partial^2 u}{\partial^2 \xi} = \sum_{j=1, \dots, n} \lambda_j \frac{\partial^2 u}{\partial^2 x_j}.$$

This proposition explains why the preceding approximation cannot be perfect.

*Proof.* The preceding equality is equivalent to

$$\begin{aligned} \sin^2 \theta &= \sum_{j=1, \dots, n} \lambda_j \sin^2 \theta_j, \\ \sin \theta \cos \theta &= \sum_{j=1, \dots, n} \lambda_j \sin \theta_j \cos \theta_j, \\ \cos^2 \theta &= \sum_{j=1, \dots, n} \lambda_j \cos^2 \theta_j. \end{aligned}$$

By combining these equations we obtain

$$\left( \sum_j \lambda_j \sin^2 \theta_j \right) \left( \sum_j \lambda_j \cos^2 \theta_j \right) = \left( \sum_j \lambda_j \sin \theta_j \cos \theta_j \right)^2.$$

Since the  $\lambda_j$  are nonnegative, equality in this Cauchy-Schwarz relation is only achieved if the vectors  $(\lambda_j^{1/2} \sin \theta_j)_j$  and  $(\lambda_j^{1/2} \cos \theta_j)_j$  are collinear. Since  $\theta$  is different from each  $\theta_j$  by hypothesis, we know that at least two of the  $\lambda_j$  are nonzero. Thus for two different indices  $i$  and  $j$  and some real number  $a$ , we get  $\cos \theta_i = a \sin \theta_i$  and  $\cos \theta_j = a \sin \theta_j$ . Since these angles are different and contained in  $[0, \pi]$  we get a contradiction.



We are finally in a position to state the exact model which will contain all the successive improvements and be easy to discretize. This is obtained by plugging into model (8) the approximated operator  $Au$  defined by (9), instead of “ $|Du| \operatorname{div} (Du/|Du|)$ .”

### 2.1. Final model.

$$(10) \quad \frac{\partial u}{\partial t} = g(|DG_{\sigma} * u|) \left( (1 - h(|Du|)) \Delta u + h(|Du|) \sum_{j=1, \dots, n} f_j \left( \frac{Du}{|Du|} \right) \frac{\partial^2 u}{\partial^2 x_j} \right).$$

The only assumption to be used in the mathematical analysis is that the  $f_j$  are smooth and nonnegative. However, they must be chosen for the applications according to the assumptions of Proposition 1. We shall give some examples in § 4.

**3. Mathematical validity of the models.** In this section, we wish to show that (1) and (10) (or the approximated equations presented in § 2) are well posed. In fact, we shall show uniqueness and existence of Lipschitz solutions (for Lipschitz initial data) for a general class of equations which contain our models. Indeed, we first observe that these equations all take the form

$$(11) \quad \frac{\partial u}{\partial t} - g(u * DG) a_{ij}(Du) \partial_{ij} u = 0 \quad \text{in } [0, +\infty] \times \mathbb{R}^n,$$

where we denote by  $\partial_i u = \partial u / \partial x_i$  and use the convention on repeated indices. Of course, in our models,  $n = 2$  or  $3$ , but mathematically we can take any  $n \geq 2$ . Next, we assume that

$$(12) \quad g \in C^{1,1}(\mathbb{R}^n, \mathbb{R}), \quad g(p) > 0 \quad \text{for all } p \text{ in } \mathbb{R}^n,$$

$$(13) \quad D^\alpha G \in L^1(\mathbb{R}^n) \quad \text{for all } |\alpha| \leq 2,$$

$$(14) \quad a_{ij}(p) \xi_i \xi_j \geq 0 \quad \text{for all } p \in \mathbb{R}^n \setminus \{0\}, \quad \xi \in \mathbb{R}^n,$$

$$(15) \quad a_{ij} \text{ is continuous and bounded on } \mathbb{R}^n \setminus \{0\}.$$

It is a trivial fact to check that (1) and (10) are indeed of the above form. Of course, the application of such models to image processing requires solving (11) only in a domain  $R$  of  $\mathbb{R}^n$  (in fact, a rectangle in  $\mathbb{R}^2$ ), that we can assume to be convex and piecewise smooth to simplify the presentation. To fix ideas we should think of  $R = [0, 1]^n$ . In that case, we have to prescribe boundary conditions for  $u$  on  $\partial R$ , and the most natural choice for image processing is Neumann boundary conditions, i.e.,

$$(16) \quad \frac{\partial u}{\partial \nu} = 0 \quad \text{on } \partial R,$$

where  $\nu$  denotes the unit exterior normal. Indeed, the Neumann condition corresponds to the reflection of the picture across the boundary and has the advantage of not imposing any value on the boundary and not creating “edges” on it. It is therefore very natural if we assume that the boundary of the picture is an arbitrary cutoff of a larger scene in view.

To simplify the presentation, we shall work with periodic boundary conditions or, in other words, solve (11) with solutions satisfying  $u(x + h) = u(x)$  for all  $x$  in  $\mathbb{R}^n$ ,  $h$  in  $\mathbb{Z}^n$ . In fact, everything we show below could be adapted directly to the case of (16) but this adaptation requires some tedious and unpleasant technicalities at the corners of  $R$ . Anyway, the natural extension of a picture adopted in image processing is, as we mentioned, to extend  $u$  by reflection across the boundary of the rectangle. Thus

we set  $u(-x, y) = u(x, y)$  if  $-1 \leq x \leq 0$ , and  $0 \leq y \leq 1$ , etc. It is easily seen that with this extension,  $u$  can be assumed to be periodic on  $(2\mathbb{Z})^n$ . (This extension preserves the Lipschitz assumption on  $u$ .)

Of course, we complement (11) with an initial condition

$$(17) \quad u(0, x) = u_0(x) \quad \text{in } \mathbb{R}^n,$$

where  $u_0$  is continuous on  $\mathbb{R}^n$  and periodic, as above. In all that follows, we shall no longer mention periodicity, and the reader should remember that all initial data and all solutions are periodic, as described above.

Next, we have to explain the meaning of (11) which is a second-order parabolic equation with possible high degeneracy and two types of nonlinear terms, namely, a quasilinear term  $(a_{ij}(Du)\partial_{ij}u)$  and a nonlocal term  $g(u * DG)$ .

This is why it is important to work here with viscosity solutions (see the survey of Crandall, Ishii, and Lions [6]) as was done in [5], [8], [9], [28] for the case when  $g = 1$ . This is not the place to explain the use of viscosity solutions; let us only point out that classical solutions are automatically viscosity solutions and that general existence, uniqueness, and approximation results are available for viscosity solutions [6].

As we shall see, the additional nonlinear term  $g(\cdot \cdot \cdot)$  induces some tricky modifications of the uniqueness proofs made in these papers. Now, except for these modifications, the strategy of the uniqueness proof is the by now classical one (see [6]).

We begin by a brief recall of the definition of viscosity solutions of (11) periodic on  $R$ . Let  $u$  in  $C([0, T] \times \mathbb{R}^n)$  for some  $T$  in  $]0, +\infty[$ . Then  $u$  is a viscosity subsolution of (11) if, for all  $\phi$  in  $C^2(\mathbb{R} \times \mathbb{R}^n)$ , the following condition holds at any point  $(t_0, x_0)$  in  $]0, T] \times \mathbb{R}^n$  which is a local maximum point of  $(u - \phi)$

$$(18) \quad \begin{aligned} & \frac{\partial \phi}{\partial t}(t_0, x_0) - g(u * DG(t_0, x_0))a_{ij}(D\phi(t_0, x_0))\partial_{ij}\phi(t_0, x_0) \leq 0 \\ & \qquad \qquad \qquad \text{if } D\phi(t_0, x_0) \neq 0, \\ & \frac{\partial \phi}{\partial t}(t_0, x_0) - g(u * DG(t_0, x_0)) \limsup_{p \rightarrow 0} a_{ij}(p)\partial_{ij}\phi(t_0, x_0) \leq 0 \\ & \qquad \qquad \qquad \text{if } D\phi(t_0, x_0) = 0. \end{aligned}$$

Notice that the above inequality (in the case when  $D\phi(t_0, x_0) \neq 0$ ) is the inequality expected from the classical maximum principle, the other case being a technical variant.

We define a viscosity supersolution in a similar manner, replacing “local maximum point” by “local minimum point,” “ $\leq 0$ ” by “ $\geq 0$ ” and “lim sup” by “lim inf.” Finally, a viscosity solution is a function which is both a subsolution and a supersolution. Notice that  $u * DG$  is in  $C([0, T] \times \mathbb{R}^n)$ ; hence, its value at  $(t_0, x_0)$  is meaningful. We may now state our main result.

**THEOREM.** *Let  $u_0, v_0$  be, respectively, Lipschitz continuous and continuous on  $R$ .*

(1) *Then the system (11)–(17) has a unique viscosity solution  $u$  in  $C([0, \infty[ \times \mathbb{R}^n) \cap L^\infty(0, T; W^{1,\infty}(\mathbb{R}^n))$  for any  $T < \infty$ . Moreover,  $\inf_{\mathbb{R}^n} u_0 \leq u(x, t) \leq \sup_{\mathbb{R}^n} u_0$ .*

(2) *Let  $v$  in  $C([0, \infty[ \times \mathbb{R}^n)$  be a viscosity solution of (11) satisfying (17) with  $u_0$  replaced by  $v_0$ . Then for all  $T$  in  $[0, +\infty[$ , there exists a constant  $K$  which depends only on  $\|u_0\|_{W^{1,\infty}}$  and  $\|v_0\|_{L^\infty}$  such that*

$$(19) \quad \sup_{0 \leq t \leq T} \|u(t, \cdot) - v(t, \cdot)\|_{L^\infty(\mathbb{R}^n)} \leq K \|u_0 - v_0\|_{L^\infty(\mathbb{R}^n)}.$$

*Remarks.*

(1) In fact, the proof of (19) will provide an estimate of  $K$  of the form  $\exp(MT)$ , where  $M$  depends on  $\|u_0\|_{W^{1,\infty}}$  and  $\|v_0\|_{L^\infty}$ .

(2) Analogous results hold if we add a right-hand side  $f(t, x)$  to (11), provided we assume, for instance, that  $f$  is in  $C([0, \infty[ \times \mathbb{R}^n) \cap L^\infty(0, T; W^{1,\infty}(\mathbb{R}^n))$  for any  $T < \infty$ . By adapting this remark, it is easy to see that  $f$  can be a term which forces  $u$  to remain close to  $u_0$ ; that is,  $f(x, t) = -u(x, t) + u_0(x)$ .

(3) The uniqueness of solutions of (11)–(17) is not clear if we do not assume  $u_0, u$ , or  $v$  to be Lipschitz continuous. It is, however, possible to prove the existence of a solution assuming only that  $u_0$  is continuous. Any such solution satisfies  $\inf_{\mathbb{R}^n} u_0 \leq u(x, t) \leq \sup_{\mathbb{R}^n} u_0$ .

(4) Above and below, we denote by  $W^{1,\infty}(\mathbb{R}^n)$  the space of bounded Lipschitz continuous functions in  $\mathbb{R}^n$ .

*Proof of the theorem.* We begin with the uniqueness part and the stability estimate claimed in Part (2) of the theorem. We follow the arguments given in Crandall, Ishii, and Lions [6], and we consider a maximum point  $(t_0, x_0, y_0)$  of

$$(20) \quad u(t, x) - v(t, y) - (4\varepsilon)^{-1}|x - y|^4 - \lambda t, \quad t \in [0, T], \quad x, y \in \mathbb{R}^n,$$

where  $T, \varepsilon, \lambda \in ]0, \infty[$  will be determined later.

We first assume that  $t_0 > 0$ . Then, as in [6], we find  $a$  and  $b$  in  $\mathbb{R}$ ,  $X, Y, (n \times n)$  symmetric matrices such that

$$(21) \quad a - b = \lambda, \quad \begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq \begin{pmatrix} A + \mu A^2 & -A - \mu A^2 \\ -A - \mu A^2 & A + \mu A^2 \end{pmatrix}$$

for each  $\mu > 0$ ,

$$(22) \quad a - g((u * DG)(t_0, x_0))a_{ij}(\varepsilon^{-1}|x_0 - y_0|^2(x_0 - y_0))X_{ij} \leq 0,$$

$$(23) \quad b - g((v * DG)(t_0, y_0))a_{ij}(\varepsilon^{-1}|x_0 - y_0|^2(x_0 - y_0))Y_{ij} \geq 0,$$

where

$$A = \varepsilon^{-1}|x_0 - y_0|^2 I_n + 2\varepsilon^{-1}(x_0 - y_0) \otimes (x_0 - y_0),$$

so that

$$A^2 = \varepsilon^{-2}|x_0 - y_0|^4 I_n + 8\varepsilon^{-2}|x_0 - y_0|^2(x_0 - y_0) \otimes (x_0 - y_0).$$

In fact, (22), (23) have to be interpreted if  $x_0 = y_0$ . In that case,  $A = 0$  so that by (21),  $X \leq 0$  and  $Y \geq 0$ . We then write

$$(22') \quad a - g((u * DG)(t_0, x_0)) \limsup_{p \rightarrow 0} a_{ij}(p)X_{ij} \leq 0,$$

$$(23') \quad b - g((v * DG)(t_0, y_0)) \liminf_{p \rightarrow 0} a_{ij}(p)Y_{ij} \geq 0.$$

Hence, in particular,  $a \leq 0$ ,  $b \geq 0$ : a contradiction with  $a - b = \lambda > 0$ .

Therefore,  $x_0 \neq y_0$  and we may write and use (22), (23). We next choose  $\mu = \varepsilon|x_0 - y_0|^{-2}$ , and we deduce

$$(24) \quad \begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq (2/\varepsilon) \begin{pmatrix} B & -B \\ -B & B \end{pmatrix},$$

where  $B = |x_0 - y_0|^2 I_n + 5(x_0 - y_0) \otimes (x_0 - y_0)$ .

We then set

$$g_1 = g((u * DG)(t_0, x_0)), \quad g_2 = g((v * DG)(t_0, x_0)), \\ a = (a_{ij}(\varepsilon^{-1}|x_0 - y_0|^2(x_0 - y_0)))_{1 \leq i, j \leq n},$$

and we consider the matrix

$$\Gamma = \begin{pmatrix} g_1 a & (g_1 g_2)^{1/2} a \\ (g_1 g_2)^{1/2} a & g_2 a \end{pmatrix}.$$

Obviously,  $\Gamma$  is a nonnegative symmetric matrix so that multiplying (24) to the left by  $\Gamma$  and taking the trace we find

$$(25) \quad \begin{aligned} g_1 a_{ij} X_{ij} - g_2 a_{ij} Y_{ij} &\leq 2\varepsilon^{-1} (g_1^{1/2} - g_2^{1/2})^2 \text{trace}(aB) \\ &\leq C_0 \varepsilon^{-1} (g_1^{1/2} - g_2^{1/2})^2 |x_0 - y_0|^2 \end{aligned}$$

for some  $C_0$  which depends only on  $(a_{ij}(p))_{1 \leq i, j \leq n}$ . Next, if we combine (21)–(23) and (25) we obtain

$$(26) \quad \lambda \leq C_0 \varepsilon^{-1} (g_1^{1/2} - g_2^{1/2})^2 |x_0 - y_0|^2.$$

We now estimate  $(g_1^{1/2} - g_2^{1/2})$ . First of all, we observe that (12) yields that  $g^{1/2}$  is Lipschitz on bounded sets, therefore

$$(g_1^{1/2} - g_2^{1/2}) \leq C_1 |(u * DG)(t_0, x_0) - (v * DG)(t_0, y_0)|$$

for some  $C_1$ , depending only on  $g$  and on  $\sup |u|, \sup |v|$ .

But this last quantity is estimated by  $C_2(\sup_{[0, T] \times \mathbb{R}^n} |u - v| + |x_0 - y_0|)$ , where  $C_2$  depends only on  $G$  and on  $\sup |u|, \sup |v|$ . This allows us to deduce from (26) that

$$(27) \quad \lambda \leq C \left\{ \left( \sup_{[0, T] \times \mathbb{R}^n} |u - v| \right)^2 \frac{|x_0 - y_0|^2}{\varepsilon} + \frac{|x_0 - y_0|^4}{\varepsilon} \right\},$$

where  $C = 2C_0 C_1^2 C_2^2$ .

Next, we estimate  $|x_0 - y_0|$ . To this aim, we observe that

$$u(t_0, x_0) - v(t_0, y_0) - \frac{|x_0 - y_0|^4}{4\varepsilon} - \lambda t_0 \geq u(t_0, y_0) - v(t_0, y_0) - \lambda t_0$$

and thus

$$\frac{|x_0 - y_0|^4}{4\varepsilon} \leq L |x_0 - y_0|,$$

where  $L$  is a Lipschitz constant (in  $x$ ) for  $u$  on  $[0, T] \times \mathbb{R}^n$ . Therefore,  $|x_0 - y_0| \leq (4\varepsilon L)^{1/3}$ . This bound and (27) finally yield

$$(28) \quad \lambda \leq M \left\{ \varepsilon^{1/2} + \varepsilon^{-1/3} \left( \sup_{[0, T] \times \mathbb{R}^n} |u - v| \right)^2 \right\},$$

where  $M = \max\{(4L)^{2/3}, (4L)^{4/3}\}C$ . Without loss of generality, we may assume  $\sup_{[0, T] \times \mathbb{R}^n} |u - v| > 0$  (otherwise, we conclude!) and we choose

$$\varepsilon^{1/3} = \delta \sup_{[0, T] \times \mathbb{R}^n} |u - v|, \quad \lambda = (1 + \delta + 1/\delta) M \sup_{[0, T] \times \mathbb{R}^n} |u - v|,$$

where  $\delta > 0$  will be determined later. These choices contradict (28). This contradiction proves, in fact, that  $t_0 = 0$ . Therefore,

$$(29) \quad u(t, x) - v(t, y) - \frac{|x - y|^4}{4\varepsilon} - \lambda t \leq \sup_{x, y \in \mathbb{R}^n} \left\{ u_0(x) - v_0(y) - \frac{|x - y|^4}{4\varepsilon} \right\}.$$

In particular, we may choose  $x = y$  in (29) while the right-hand side can be estimated by  $\sup_{\mathbb{R}^n} (u_0 - v_0) + \sup_{r \geq 0} (Lr - r^4/4\varepsilon)$ .

We finally obtain

$$(30) \quad \sup_{[0,T] \times \mathbb{R}^n} (u-v) \leq \sup_{\mathbb{R}^n} |u_0 - v_0| + \frac{3}{4} L^{4/3} \delta \sup_{[0,T] \times \mathbb{R}^n} |u-v| + M(1+\delta+\delta^{-1})T \sup_{[0,T] \times \mathbb{R}^n} |u-v|.$$

Exchanging the role of  $u$  and  $v$ , and choosing  $\delta = L^{-4/3}$ , we deduce

$$(31) \quad \sup_{[0,T] \times \mathbb{R}^n} |u-v| \leq 4 \sup_{\mathbb{R}^n} |u_0 - v_0| + KT \sup_{[0,T] \times \mathbb{R}^n} |u-v|,$$

where  $K = 4M(1+\delta+\delta^{-1})$ . In order to conclude, we choose  $T = t_1 = 1/2K$ , and we find

$$(32) \quad \sup_{[0,t_1] \times \mathbb{R}^n} |u-v| \leq 8 \sup_{\mathbb{R}^n} |u_0 - v_0|.$$

Therefore, if  $T$  is an arbitrary time in  $[0, +\infty[$  and  $N \geq 1$  is such that  $Nt_1 \geq T$ , we deduce easily by reiterating this argument that

$$(33) \quad \sup_{[0,T] \times \mathbb{R}^n} |u-v| \leq 8^N \sup_{\mathbb{R}^n} |u_0 - v_0|.$$

This proves part (2) of the theorem and the uniqueness claim in part (1).

We next prove the existence claim in part (1). We begin by remarking that definition (18) of viscosity solutions immediately implies that if  $u$  is a solution, then

$$\inf_{\mathbb{R}^n} u_0 - \delta t \leq u \leq \sup_{\mathbb{R}^n} u_0 + \delta t \quad \text{on } [0, +\infty] \times \mathbb{R}^n \quad \text{for all } \delta > 0.$$

Therefore, we have

$$(34) \quad \inf_{\mathbb{R}^n} u_0 \leq u \leq \sup_{\mathbb{R}^n} u_0 \quad \text{on } [0, +\infty] \times \mathbb{R}^n.$$

Indeed, set  $\phi(x, t) = \sup_{\mathbb{R}^n} u_0 + \delta t$  and assume that  $u - \phi$  has a local maximum at a point  $(t_0, x_0)$  with  $t_0 > 0$ . Then by the definition of subsolution, we get by the second relation of (18) that  $\partial \phi / \partial t(t_0, x_0) \leq 0$ . Thus  $\delta \leq 0$ , which yields a contradiction and therefore  $u - \phi$  attains its maximum, zero, for  $t_0 = 0$ .

Next, we prove an a priori estimate on  $Du$ . This estimate will be formal at that level and will be justified later. In fact, we consider a smooth solution  $u$  of

$$(35) \quad \frac{\partial u}{\partial t} - g(\omega * DG) a_{ij}(Du) \partial_{ij} u = 0 \quad \text{in } ]0, +\infty[ \times \mathbb{R}^n,$$

where  $a_{ij}$  is now supposed to be smooth on  $\mathbb{R}^n$ , and  $\omega \in L^\infty(]0, +\infty[ \times \mathbb{R}^n)$ . We are going to show that

$$(36) \quad \|Du(t, \cdot)\|_{L^\infty(\mathbb{R}^n)} \leq e^{Ct} \|Du_0\|_{L^\infty(\mathbb{R}^n)},$$

where  $C$  depends only on  $\sup_{|p| \leq R} |D^2 g(p)|$  and  $\sup_p |a_{ij}(p)|$  with  $R = \|w\|_{L^\infty(\mathbb{R}^n)} \|DG\|_{L^1(\mathbb{R}^n)}$ . Everywhere below,  $C$  will denote positive constants depending only on these quantities. To prove the a priori estimate (36), we use the “classical” Bernstein method and derive a parabolic inequality for  $|Du|^2$ . To this end, we differentiate (35) with respect to  $x_k$ , and we find

$$(37) \quad \begin{aligned} & \frac{\partial u_k}{\partial t} - g(\omega * DG) a_{ij}(Du) \partial_{ij} u_k - \frac{\partial g}{\partial l}(\omega * DG) \cdot (\omega * \partial_{lk} G) a_{ij}(Du) \partial_{ij} u \\ & - g(\omega * DG) \frac{\partial a_{ij}}{\partial l}(Du) \partial_{lk} u = 0 \quad \text{in } ]0, +\infty[ \times \mathbb{R}^n, \end{aligned}$$

where we denote by  $u_k = \partial_k u$ . Hence, we obtain by multiplying by  $u_k$

$$\begin{aligned}
 (38) \quad & \frac{\partial |Du|^2}{\partial t} - g(\omega * DG) a_{ij}(Du) \partial_{ij}(|Du|^2) - g(\omega * DG) \frac{\partial a_{ij}}{\partial l}(Du) \partial_l(|Du|^2) \\
 & = -2g(\omega * DG) a_{ij}(Du) u_{ki} u_{kj} + 2 \frac{\partial g}{\partial l}(\omega * DG) \\
 & \quad \cdot (\omega * \partial_{lk} G) a_{ij}(Du) u_k \partial_{ij} u \quad \text{in } ]0, +\infty[ \times \mathbb{R}^n.
 \end{aligned}$$

Next, we observe that in terms of constants  $C$  only depending on  $\sup |\omega|$  and  $g$ , we have

$$|\omega * \partial_{lk} G| \leq C, \quad \left| \frac{\partial g}{\partial l}(\omega * DG) \right| \leq C(g(\omega * DG))^{1/2},$$

and

$$|a_{ij}(Du) u_{ij}| \leq C(a_{ij}(Du) u_{ki} u_{kj})^{1/2}.$$

(This last inequality is purely algebraic and only uses that  $a_{ij} x_i x_j$  is nonnegative.)

Inserting these bounds in (38) and using the Cauchy-Schwarz inequality we get

$$\begin{aligned}
 (39) \quad & \frac{\partial |Du|^2}{\partial t} - g(\omega * DG) a_{ij}(Du) \partial_{ij}(|Du|^2) - g(\omega * DG) \frac{\partial a_{ij}(Du)}{\partial l} \partial_l(|Du|^2) \\
 & \leq C|Du|^2 \quad \text{in } ]0, +\infty[ \times \mathbb{R}^n.
 \end{aligned}$$

We then deduce easily (36) by applying the maximum principle [2]. In order to conclude, we only have to approximate (11) by a (slightly) simpler one of a similar form for which we will be able to produce smooth solutions. Then, we will conclude using the above a priori estimate (which will be valid on the approximated solutions). To this end, we consider  $u_0^\varepsilon$  in  $C^\infty(\mathbb{R}^n)$  (periodic) such that  $u_0^\varepsilon \rightarrow u_0$  uniformly,  $\|Du_0^\varepsilon\|_{L^\infty} \leq \|Du_0\|_{L^\infty}$ ,  $\|u_0^\varepsilon\|_{L^\infty} \leq \|u_0\|_{L^\infty}$ . We also introduce  $g_\varepsilon = g + \varepsilon$ ,  $a_{ij}^\varepsilon = \varepsilon \delta_{ij} + \alpha_{ij}^\varepsilon$ , where the  $\alpha_{ij}^\varepsilon$  tend monotonically to  $a_{ij}$ , satisfy (14), and have compact support in  $\mathbb{R}^n \setminus \{0\}$ .

Using the general theory of quasilinear uniformly parabolic equations (Ladyzhenskaya, Solonnikov and Ural'tseva [14]), it can easily be checked that there exists  $u^\varepsilon$  smooth on  $[0, +\infty[ \times \mathbb{R}^n$  solution of

$$(40) \quad \frac{\partial u^\varepsilon}{\partial t} - g_\varepsilon(u^\varepsilon * DG) a_{ij}^\varepsilon(Du^\varepsilon) \partial_{ij} u^\varepsilon = 0 \quad \text{in } ]0, +\infty[ \times \mathbb{R}^n.$$

In view of the general consistency-stability properties of viscosity solutions, there just remains to show that  $u^\varepsilon$  (or a subsequence) converges uniformly on  $[0, T] \times \mathbb{R}$  to some function  $u$  in  $C([0, T] \times \mathbb{R}^n) \cap L^\infty(0, T; W^{1,\infty}(\mathbb{R}^n))$  for any  $T < \infty$ . This will follow from the Ascoli-Arzelà theorem. Indeed, we may now apply the proof of estimate (36), and we find for all  $t$  in  $[0, T]$

$$(41) \quad \|Du^\varepsilon(t, \cdot)\|_{L^\infty(\mathbb{R}^n)} \leq e^{Ct} \|Du_0^\varepsilon\|_{L^\infty(\mathbb{R}^n)} \leq e^{Ct} \|Du_0\|_{L^\infty(\mathbb{R}^n)} \leq C_T.$$

In other words,

$$(42) \quad |u^\varepsilon(t, x) - u^\varepsilon(t, y)| \leq C_T |x - y| \quad \text{for any } x, y \text{ in } \mathbb{R}^n \text{ and } t \text{ in } [0, T],$$

where  $C_T$  denotes various constants independent of  $\varepsilon, t, x, y$ . In addition, this estimate combined with (40) yields, by a (somewhat) standard argument, that

$$(43) \quad |u^\varepsilon(s, x) - u^\varepsilon(t, x)| \leq C_T |t - s|^{1/2} \quad \text{for any } x \text{ in } \mathbb{R}^n \text{ and } s, t \text{ in } [0, T].$$

We conclude then by combining (42) and (43).

Let us sketch the proof of (43). It follows upon remarking that if  $s \leq t \leq T$ ,

$$(44) \quad \|u^\varepsilon(t, \cdot) - u_\delta^\varepsilon\|_{L^\infty(\mathbb{R}^n)} \leq \frac{C_T(t-s)}{\delta} + \|u^\varepsilon(s, \cdot) - u_\delta^\varepsilon\|_{L^\infty(\mathbb{R}^n)},$$

where  $\delta$  is arbitrary in  $]0, \infty[$ ,  $u_\delta^\varepsilon \in W^{2,\infty}(\mathbb{R}^n)$ ,  $\|u^\varepsilon(s, \cdot) - u_\delta^\varepsilon\|_{L^\infty(\mathbb{R}^n)} \leq C_T \delta$  and  $\|D^2 u_\delta^\varepsilon\|_{L^\infty(\mathbb{R}^n)} \leq C_T/\delta$ .

Indeed, we have for  $s \leq t \leq T$ ,  $x \in \mathbb{R}^n$ ,

$$(45) \quad \left| \frac{\partial u_\delta^\varepsilon}{\partial t} - g_\varepsilon(u^\varepsilon * DG) a_{ij}^\varepsilon(Du^\varepsilon) \partial_{ij} u_\delta^\varepsilon \right| \leq \frac{C_T}{\delta},$$

and (44) is deduced from the maximum principle. We then derive (43) by choosing  $\delta = (t-s)^{1/2}$ .

#### 4. Numerical scheme and experimental results

**4.1. Definition of the functions  $f_n$ .** In order to discretize the degenerate diffusion operator  $|Du| \operatorname{div}(Du/|Du|)$  we shall use the approximated operator  $Au$  defined by (9) in § 2, which corresponds to a slight preference given to a discrete set of directions  $\theta_n$ . There are several techniques to define the “partition of unity” given by the  $f_n$ . If the directions  $\theta_n$  are given by  $\theta_n = (n-1)\pi/2N$ ,  $n = 1, \dots, 2N$ , then define an even smooth function  $f$  with support in  $[-\pi/2N, \pi/2N]$  verifying  $f(\pi/2N - \theta) + f(\theta) = 1$ . Then we define  $f_n(\theta) = f(\theta - \theta_n)$ , where for simplicity we identify the bounds of the interval  $[0, \pi]$ .

In our experiments, we took a nine points square neighborhood for each point  $(i, j)$  of the grid. Thus four directions were present in the discretization: those of  $(i-1, j)$  to  $(i+1, j)$ ,  $(i, j-1)$  to  $(i, j+1)$ ,  $(i-1, j-1)$  to  $(i+1, j+1)$ , and  $(i-1, j+1)$  to  $(i+1, j-1)$ . We then defined

$$\begin{aligned} f(\theta) &= 1 \quad \text{if } |\theta| \leq \theta_0, \\ f(\theta) &= 0 \quad \text{if } |\theta| \geq \pi/4 - \theta_0, \\ f(\theta) &= (4/\pi)^4 (|\theta| - \pi/4 + \theta_0)^4 \quad \text{elsewhere.} \end{aligned}$$

The constant  $\theta_0$  is defined in order to penalize the diffusion as  $\theta$  is not close to a principal direction. We chose  $\theta_0$  so that the diffusion coefficient is nearly  $\frac{1}{16}$  for both closest principal directions in the worst case for our scheme: when  $\theta$  lies in the middle of the interval defined by two principal directions. This inhibition of the diffusion contradicts the assumption made in Proposition 1, that the  $f_n$  should be asymptotically a partition of unity. Now, our experience is that this asymptotic assumption cannot be respected with so few directions in the scheme.

#### 4.2. Discretization of the equation. We use the semi-implicit form

$$\begin{aligned} \frac{u^{k+1} - u^k}{\delta t} &= g(|G_\sigma * u^k|) \left( (1 - h(|Du^k|)) \Delta u^{k+1} + h(|Du^k|) \right. \\ &\quad \left. \cdot \sum_n f_n \left( \frac{Du^k}{|Du^k|} \right) \left( \frac{\partial^2 u^{k+1}}{\partial^2 x_n} \right) \right), \end{aligned}$$

which yields  $u^{k+1}$  as the solution of a linear system.

#### 4.3. The algorithm.

(1) Define the contrast and accuracy parameters yielding the functions  $g$ ,  $h$ , and the variance  $\sigma$ . Fix  $\delta t$ . (The spatial increment is assumed to be 1.)

(2) Computation of  $Du^k$ . We use the following approximate of the gradient:

$$\begin{aligned} u_x(i, j) &= (2 + 4C)^{-1} \{ u(i+1, j) - u(i-1, j) \\ &\quad + C[u(i+1, j+1) - u(i-1, j+1) \\ &\quad + u(i+1, j-1) - u(i-1, j-1)] \}, \\ u_y(i, j) &= (2 + 4C)^{-1} \{ u(i, j+1) - u(i, j-1) \\ &\quad + C[u(i+1, j+1) - u(i+1, j-1) \\ &\quad + u(i-1, j+1) - u(i-1, j-1)] \}. \end{aligned}$$

The constant  $C$  is chosen in order to get the right value of  $|Du|$  in the case where the direction of  $Du$  is one of the principal directions and  $u$  is a step function. Then we get  $C = (2^{1/2} - 1)/(2 - 2^{1/2})$ . This same discretization is used in Nitzberg and Shiota [21].

(3) Computation of  $g(G_\sigma * Du^k)(i, j)$ , and of  $h(|Du^k|)(i, j)$ .

(4) If  $h(|Du^k|)(i, j) > 0$ , computation of  $f_n(Du^k/|Du^k|)(i, j)$ .

(5)  $\Delta u^{k+1}(i, j) = 0.5[u^{k+1}(i+1, j) + u^{k+1}(i-1, j) + u^{k+1}(i, j+1) + u^{k+1}(i, j-1)]$

$$\cdot \begin{bmatrix} +0.25 \\ u^{k+1}(i+1, j+1) + u^{k+1}(i-1, j-1) + u^{k+1}(i-1, j+1) + u^{k+1}(i+1, j-1) \\ -3u^{k+1}(i, j) \end{bmatrix}.$$

This choice is made in order to use the nine points in the discretization.

$$\frac{\partial^2 u^{k+1}}{\partial^2 x_n(i, j)} = (m^2 + l^2)^{-1} (u^{k+1}(i+m, j+l) - 2u^{k+1}(i, j) + u^{k+1}(i-m, j-l)).$$

(6) With this discretization we obtain a linear system in  $u^{k+1}$  which can be solved by any iterative relaxation method and for which  $\sup |u^{k+1}| \leq \sup |u^k|$ . Of course, we have to fix natural boundary conditions to our problem. We impose that the normal derivative  $\partial u^{k+1}/\partial n$  is zero, which corresponds to a prolongation by reflection of the image across the boundary.

Let us explain briefly why the scheme does not increase the  $L^\infty$  norm of  $u^{k+1}$ . From the preceding discretization and the fact that the functions  $(1-h)$ ,  $h$ ,  $f_n$ ,  $g$  are nonnegative, it is easily seen that the iteration  $u^k \rightarrow u^{k+1}$  can be written under the generic form:

$$u^{k+1}(i, j) - u^k(i, j) = \sum_{m,l} \alpha_{m,l} [u^{k+1}(i+m, j+l) - 2u^{k+1}(i, j) + u^{k+1}(i-m, j-l)],$$

where the  $\alpha_i$  are all nonnegative. Let  $(i, j)$  be the point of the grid where  $u^{k+1}$  attains its maximum, then clearly  $u^{k+1}(i, j) \leq u^k(i, j)$ . This proves the announced property, which is important from the computational viewpoint. Indeed, the pictures are defined in the actual technology with entire values between zero and 255 and it is interesting to keep these bounds.

We shall not try to prove the convergence of this scheme. Let us only mention the general convergence results of Barles and Souganidis [1] in the context of viscosity solutions theory.



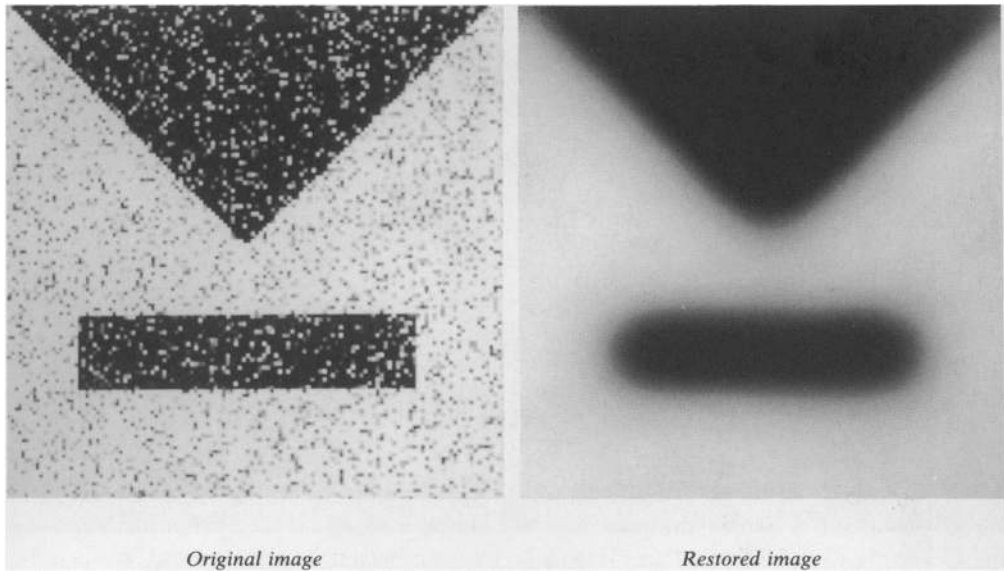


FIG. 1. *Isotropic diffusion of the original image ( $128 \times 128$ ) without selective direction. Parameters: Times = 2.5, number of iterations = 25, time to calculate on a SUN 4/110 = 75 seconds.*

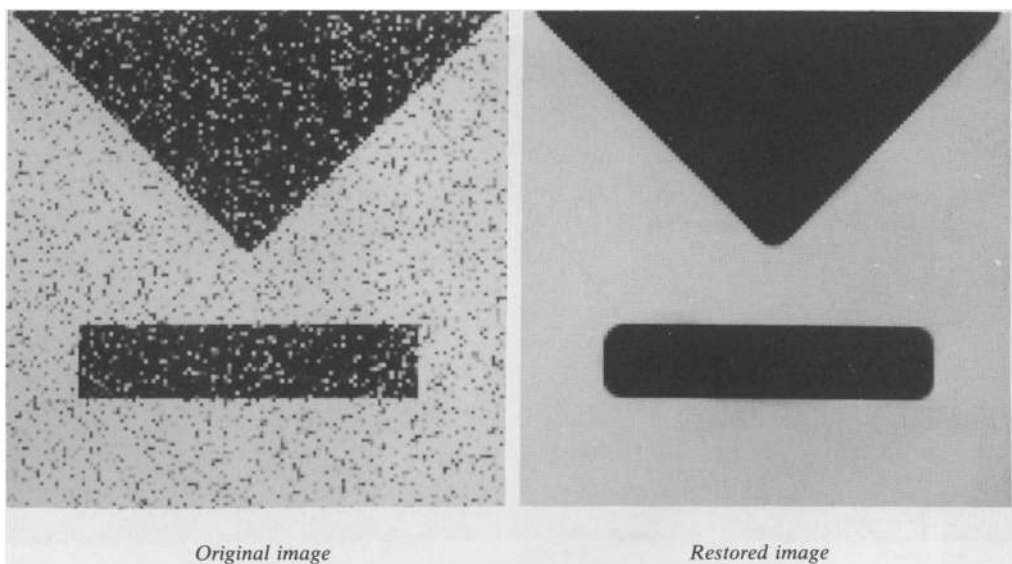


FIG. 2. *Anisotropic diffusion of the original image ( $128 \times 128$ ) with selective direction. Parameters: Times = 5, threshold gradient norm = 40, scale = 2, number of iterations = 25, time to calculate on a SUN 4/110 = 75 seconds.*

FIGS. 1 AND 2. *The left part is a synthetic picture made of the characteristic functions of rectangle and a triangle. It is then degraded by giving a random value to a randomly distributed subset of the pixels. In the first figure, 20 percent of the pixels are degraded. The right part of Fig. 1 is the restored image by the classical linear theory (isotropic linear diffusion). The right part of Fig. 2 is the restored image by our algorithm.*

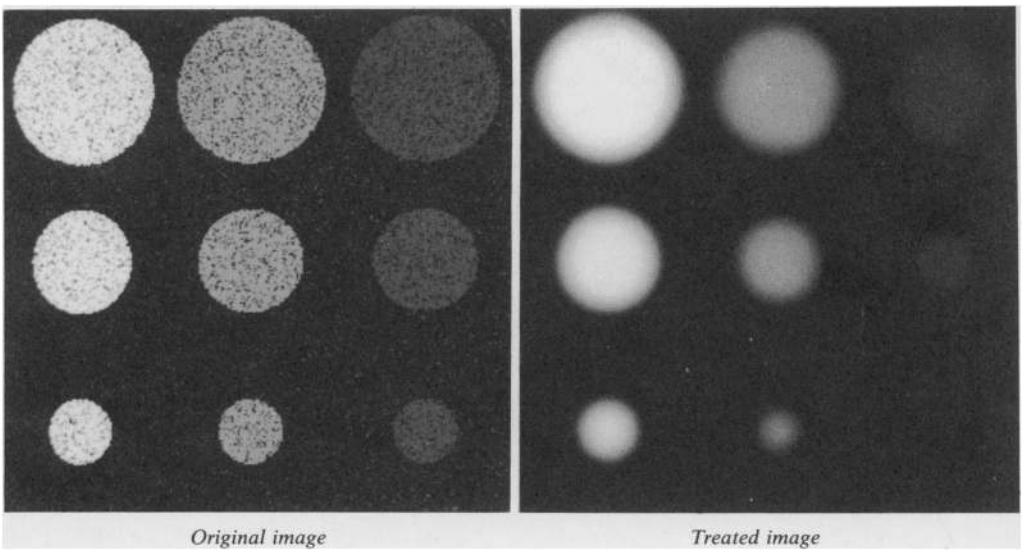


FIG. 3. Isotropic diffusion of the original image ( $256 \times 256$ ) without selective direction. Parameters: Times = 2.5, number of iterations = 20, time to calculate on a SUN 4/110 = 240 seconds.

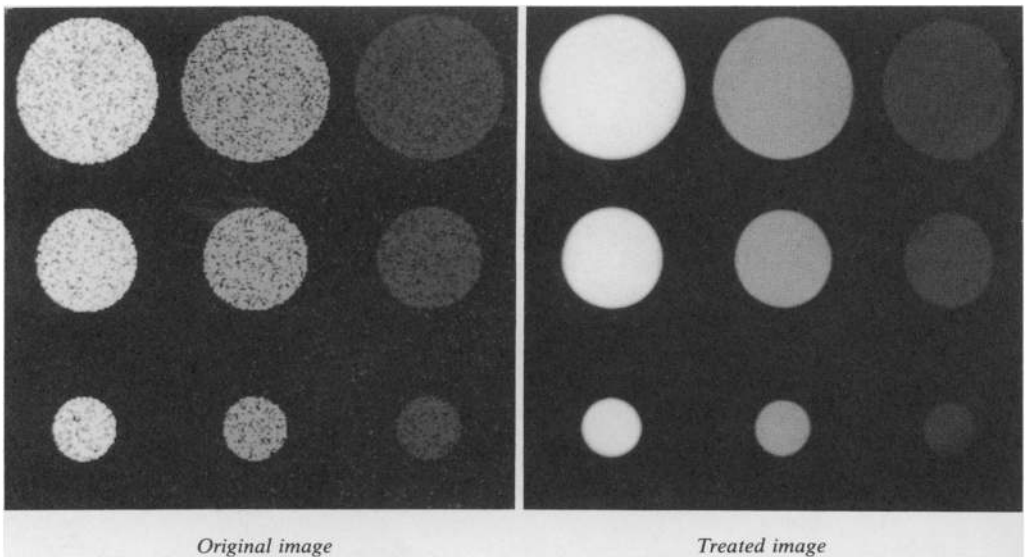


FIG. 4. Anisotropic diffusion of the original image ( $256 \times 256$ ) with selective direction. Parameters: Times = 10, threshold gradient norm = 40, scale = 2, number of iterations = 40, time to calculate on a SUN 4/110 = 480 seconds.

FIGS. 3 AND 4. They present exactly the same experiments for an initial synthetic image made of several disks with several sizes and contrasts. Note the "mean curvature motion" effect which makes the smaller disks decrease faster.

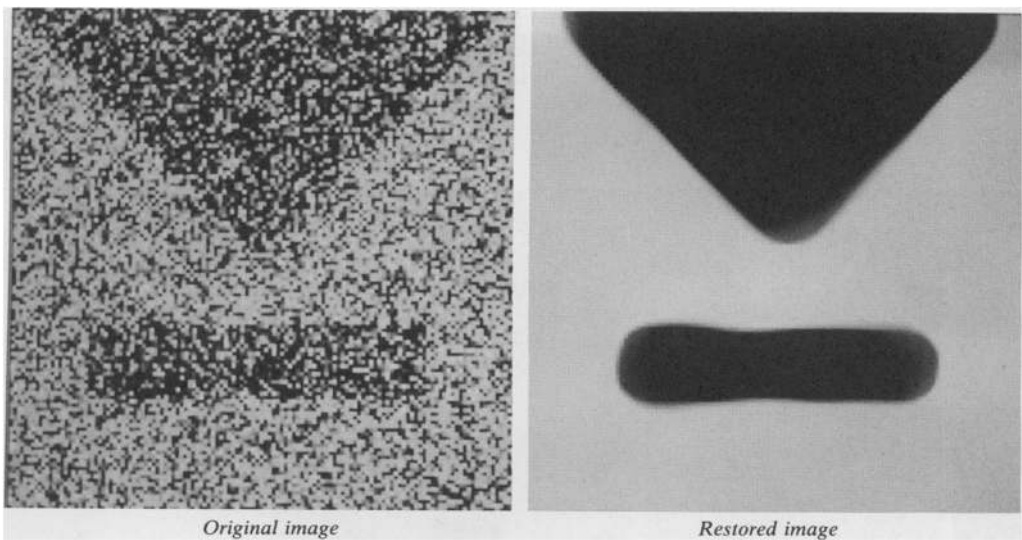


FIG. 5. Anisotropic diffusion of the original image ( $128 \times 128$ ) with selective direction. Parameters: Times = 5, threshold gradient norm = 80, scale = 2, number of iterations = 100, time to calculate on a SUN 4/100 = 300 seconds. This is the same experiment with a highly degraded version of the triangle-rectangle image. Indeed, 70 percent of the pixels are degraded. (If 100 percent were degraded, the image would be completely lost.) A large number of iterations is needed for restoration, with the subsequent smoothing effect of the corners. The restored image is renormalized for visual presentation, in order that the brightest pixel has value 255 and the darkest zero.



FIG. 6. Anisotropic diffusion of the original image ( $256 \times 256$ ) with selective direction. Parameters: Times = 5, threshold gradient norm = 80, scale = 1, number of iterations = 1, time to calculate on a SUN 4/110 = 13.5 seconds. Our algorithm applied to a "real picture" (without artificial degradation).

**4.4. Experimental results.** Figures 1-8 have been made on SUN 4/110 with a picture processing environment called MEGAWAVE, by Jacques Froment.

Each of the figures presents a pair of images, on the left the image to be processed and on the right the image processed, either by the preceding scheme, or, for comparison, by the linear heat equation. In the upper part of the figure are indicated the

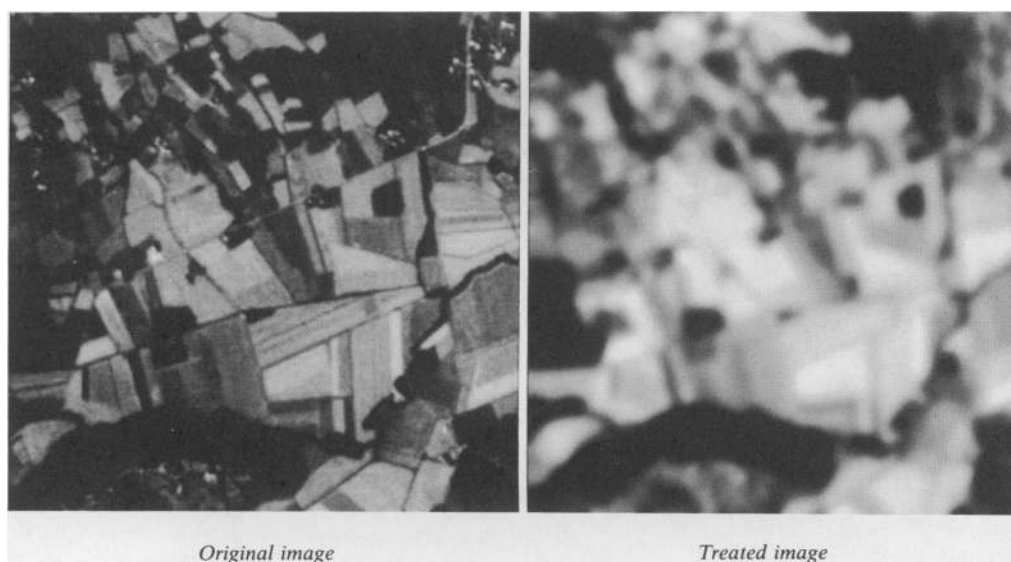


FIG. 7. Isotropic diffusion of the original image ( $256 \times 256$ ) without selective direction. Parameters: Times = 10.5, number of iterations = 2, time to calculate on a SUN 4/110 = 27 seconds. The isotropic diffusion algorithm is applied to a satellite image and then our algorithm.

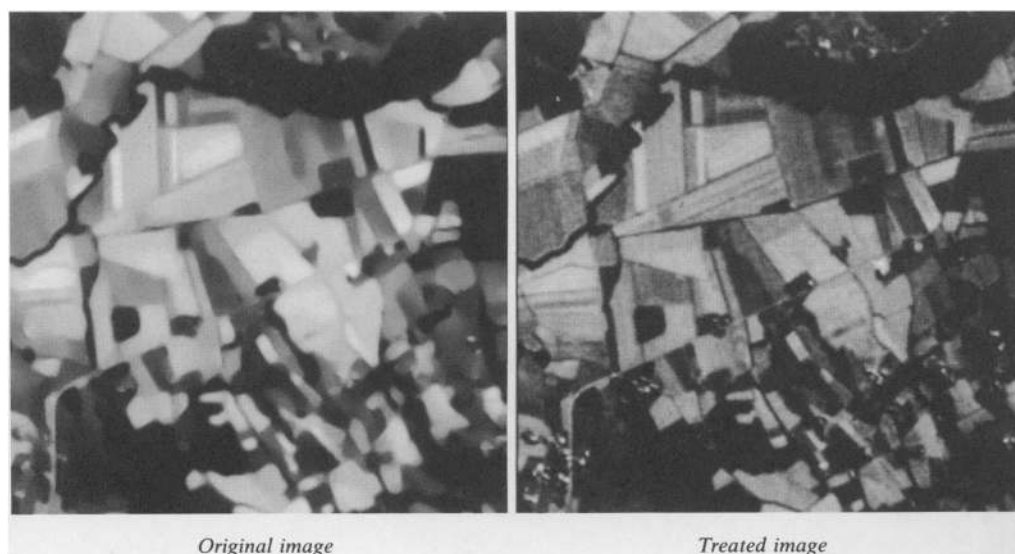


FIG. 8. Anisotropic diffusion of the original image ( $256 \times 256$ ) with selective direction. Parameters: Times = 20, threshold gradient norm = 40, scale = 1, number of iterations = 2, time to calculate on a SUN 4/110 = 27 seconds. The isotropic diffusion algorithm is applied to a satellite image and then our algorithm.

name of the method and the size of the picture (e.g.,  $256 \times 256$  for a square picture with  $256^2$  pixels). All of the figures were programmed by Luis Alvarez at CEREMADE, University of Paris IX (Dauphine).

Then the values of the parameters of the method are given in the following order:

1. The time  $t$ .
2. The enhancement parameter, which is in terms of  $|Du|^2$ . For instance, if this parameter is 49, that means that the diffusion is inhibited at any point where  $|G * Du|$  is greater than 7. Note that since the spatial unit is 1 (the side of one pixel), and since  $0 \leq u \leq 255$ , we have, by the kind of discretization explained above,  $0 \leq |Du| \leq 2^{-1/2} 255$ . Thus a meaningful threshold must lie between these values.
3. The scale parameter, which fixes the spatial radius of the smoothing kernel  $G$ . Thus "Scale = 1" means that this kernel involves  $(2 + 1)^2 = 9$  pixels. "Scale = 2" means that 25 pixels are involved.
4. The number of iterations of the scheme.
5. The real computing time on the SUN4/110, which is a machine with approximately 7 MIPS and 1.2 MegaFlops. It is the time that it is necessary to wait to see the result on the screen.

#### REFERENCES

- [1] G. BARLES AND P. E. SOUGANIDIS, *Convergence of approximation schemes for fully nonlinear second order equation*, Asymptotic Anal., pp. 182–193.
- [2] H. BREZIS, *Analyse Fonctionnelle, Théorie et Applications*, Masson, Paris, 1987.
- [3] J. CANNY, *Finding edges and lines in images*, Tech. Report 720, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1983.
- [4] F. CATTÉ, T. COLL, P. L. LIONS, AND J. M. MOREL, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM J. Numer. Anal., 29 (1992), pp. 182–193.
- [5] Y.-G. CHEN, Y. GIGA, AND S. GOTO, *Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations*, Hokkaido University, 1989, preprint.
- [6] M. G. CRANDALL, H. ISHII, AND P. L. LIONS, *User's guide to viscosity solution of second order partial differential equation*. CEREMADE, 1990, preprint.
- [7] J. I. DIAZ, *A nonlinear parabolic equation arising in image processing*, Extracta Math., Universidad de Extremadura, 1990.
- [8] L. C. EVANS AND J. SPRUCK, *Motion of level sets by mean curvature*, I, preprint.
- [9] Y. GIGA, S. GOTO, H. ISHII, AND M.-H. SATO, *Comparison principle and convexity preserving properties for singular degenerate parabolic equation on unbounded domains*, Hokkaido University, 1990, preprint.
- [10] R. E. GRAHAM, *Snow removal: A noise-stripping process for TV signals*, IRE Trans. Information Theory IT-9 (1962), pp. 129–144.
- [11] K. HOLLIG AND J. A. NOHEL, *A diffusion equation with a nonmonotone constitutive function*, Proc. on Systems of Nonlinear Partial Differential Equations, John Ball, ed., Reidel, Boston, MA, 1983, pp. 409–422.
- [12] M. KASS, A. WITKIN, AND D. TERZOPOULOS, *Snakes: active contour models*, ICCV 1987, IEEE 777 (1987).
- [13] J. J. KOENDERINK, *The structure of images*, Biol. Cybernet., 50 (1984), pp. 363–370.
- [14] O. A. LADYZHENSKAYA, V. A. SOLONNIKOV, AND N. N. URALTSEVA, *Linear and Quasilinear Equations of Parabolic Type*, American Mathematical Society, Providence, RI, 1968.
- [15] P. L. LIONS, *Generalized solutions of Hamilton–Jacobi equations*, Res. Notes Math., 69, Pitman, Boston, MA, 1982.
- [16] S. MALLAT AND S. ZHONG, *Complete signal representation with multiscale edges*, Tech. Rep. 483, Robotics Report 219, Computer Science Division, Courant Institute, New York.
- [17] D. MARR, *Vision*, W. H. Freeman, San Francisco, CA, 1982.
- [18] D. MARR AND E. HILDRETH, *Theory of edge detection*, Proc. Roy. Soc. London Ser. B, 207 (1980), pp. 187–217.
- [19] J. M. MOREL AND S. SOLIMINI, *Segmentation of images by variational methods: a constructive approach*, Rev. Mat. Univ. Complut. Madrid, Vol. 1 (1988), pp. 169–182.

- [20] D. MUMFORD AND J. SHAH, *Boundary detection by minimizing functionals*, IEEE Conf. Comput. Vision and Pattern Recognition, San Francisco, CA, 1985.
- [21] M. NITZBERG AND T. SHIOTA, *Nonlinear image smoothing with edge and corner enhancement*, Tech. Report 90-2, Division of Applied Sciences, Harvard University, Cambridge, MA, 1990.
- [22] K. N. NORDSTRÖM, *Biased anisotropic diffusion—A unified approach to edge detection*, Dept. of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, 1989, preprint.
- [23] S. OSHER AND L. RUDIN, *Feature-oriented image enhancement using shock filters*, SIAM J. Numer. Anal., 27 (1990), pp. 919–940.
- [24] S. OSHER AND J. SETHIAN, *Fronts propagating with curvature dependent speed: algorithms based on the Hamilton–Jacobi formulation*, J. Comput. Phys., 79 (1988), pp. 12–49.
- [25] P. PERONA AND J. MALIK, *Scale space and edge detection using anisotropic diffusion*, Proc. IEEE Comput. Soc. Workshop on Comput. Vision, 1987.
- [26] T. RICHARDSON, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1990.
- [27] A. ROSENFELD AND M. THURSTON, *Edge and curve detection for visual scene analysis*, IEEE Trans. on Comput., C-20, 1971, pp. 562–569.
- [28] M. SONER, *Motion of a set by the curvature of its mean boundary*, preprint.
- [29] A. P. WITKIN, *Scale-space filtering*, Proc. IJCAI, Karlsruhe, 1983, pp. 1019–1021.
- [30] A. YUILLE AND T. POGGIO, *Scaling theorems for zero crossings*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 8 (1986).