SIAM J. NUMER. ANAL.
Vol. 29 No. 1, pp. 182-193, February 1992

012

IMAGE SELECTIVE SMOOTHING AND EDGE DETECTION BY NONLINEAR DIFFUSION*

FRANCINE CATTÉ,† PIERRE-LOUIS LIONS,† JEAN-MICHEL MOREL,† AND TOMEU COLL‡

Abstract. A new version of the Perona and Malik theory for edge detection and image restoration is proposed. This new version keeps all the improvements of the original model and avoids its drawbacks: it is proved to be stable in presence of noise, with existence and uniqueness results. Numerical experiments on natural images are presented.

Key words. multiscale image analysis, edge detection, parabolic equation, nonlinear diffusion, stability

AMS(MOS) subject classifications. 49F22, 53A10, 82A60, 76T05, 49A50, 80A15, 40F10

1. Introduction. The "low level" analysis of monodimensional and bidimensional signals presents two opposite requirements. We generally wish to extract the tendency of the signal. This corresponds to a rather nonlocal analysis of the signal. For instance, we may look for regions of an image where the signal is of constant mean.

On the other hand, we wish to determine accurately where the signal changes its tendency. In the case of image analysis, the curves where this happens are called "edges" [16]. In the classical theory, the tendency of the signal can be extracted by a low pass filtering [24]-[26]. This theory comes from Marr and Hildreth [16] and has been better formalized by Witkin, Koenderink, and improved by Canny [6]. The low pass filtering is made by convolution with Gaussians of increasing variance. Then the "edges" are defined as the set of points where the norm of the gradient of the resulting smooth signal has a local maximum. (The set of "edges" is therefore contained in the set of the points where the Laplacian of the smoothed signal changes sign.) The necessity of a previous low pass filtering is easily understood: 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, to the presence of "textures." According to the above-mentioned edge detection theory, the boundaries of different textured regions can be found only if the mean value of the signal is different for each texture. Witkin [25] noticed that the convolution of the signal with Gaussians at each scale was equivalent to the solving of the heat equation with the signal as initial datum. Denote by u_0 this datum. Then the "scale space" analysis associated with u_0 consists in solving the system

$$\partial u(x, t)/dt = \Delta u(x, t), \qquad u(x, 0) = u_0(x). \tag{1.1}$$

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

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

is the Gaussian function. Then x is an edge point for the "scale" $t^{1/2}$ at points where $\Delta u(x, t)$ changes sign and $|\nabla u(x, t)|$ is "large." Of course, this last condition introduces

* Received by the editors September 24, 1990; accepted for publication March 27, 1991. This work was supported by U.S. Army contract DAJA 45-88-C-0009.

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

‡ Department de Matematiques Informatica, Universitat de les Illes Balears, Carret. Valldemossa km 7.5, Palma de Mallorca, Spain.

an a priori defined threshold. Unfortunately, it is well known (and it is enough to look at the "edges" found by this method to know it well [13]) 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 [6], [23] which is generally used as the best linear filter for white noise elimination and edge detection. An important improvement of the edge detection theory has been introduced by Malik and Perona [23]. They propose to replace the heat equation by a nonlinear equation of the porous medium type:

$$\partial u/dt = \operatorname{div}\,(g(|\nabla u|)\nabla u), \qquad u(0)=u_0. \tag{1.2}$$

In this equation, g is a smooth nonincreasing function with $g(0)=1$, $g(x)\geqq 0$, and $g(x)$ tending to zero at infinity. The idea is that the smoothing process obtained by the equation is "conditional": if $\nabla u(x)$ is large, then the diffusion will be low and therefore the exact localization of the "edges" will be kept. If $\nabla u(x)$ is small, then the diffusion will tend to smooth still more around x. Thus the choice of g corresponds to a sort of thresholding which has to be compared to the thresholding of $|\nabla u|$ used in the final step of the classical theory. Since this thresholding introduces a nonlinear device anyway, it was natural to ask whether it could not be used 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. (This property has been sought by many researchers in the past two decades.)

However, the Malik and Perona model has several serious practical and theoretical difficulties which this paper aims to solve. The first is a straightforward objection that every researcher in signal analysis will surely raise (and that Malik and Perona themselves considered). Assume that the signal is noisy, with white noise, for instance. Then the noise introduces very large, in theory unbounded, oscillations of the gradient ∇u. Thus the conditional smoothing introduced by the model will not help, since all these noise edges will be kept!

Malik and Perona proposed to eliminate this difficulty by a first smoothing of the image using some low pass filter. It seems to work well in practice, but it is nonetheless a trick which should be avoided in a correct theory. An obvious drawback is the introduction in the method of a new parameter, the variance of the Gaussian with which this preliminary smoothing should be made. Moreover, *the theory seems to adopt again what it tried to avoid*: a nonadaptive filtering which makes lost the accuracy of the edges.

The second difficulty arises from the equation itself; among the functions g which Malik and Perona consider admissible are found functions of the type $g(s)=e^{-s}$ or $g(s)=(1+s^2)^{-1}$ for which no correct theory of (1.2) 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 in theory can be the initial condition of solutions divergent in time (see [11] for simple and explicit examples, and also [8]). *In practice, that means that very close pictures can produce divergent solutions and therefore different edges.* This is certainly a drawback for most applications; think, for instance, of stereo vision, where this situation arises. However, it would be reasonable to try to define a model where the function g, which is a sort of thresholding, will be quickly decreasing, as e^{-s}, for instance. Before arriving at an improved model, let us just sketch why, if $sg(s)$ is nonincreasing, Perona and Malik's equation will be an ill-posed problem. Without loss of generality, take the case where the signal is one-dimensional. The equation becomes $du/dt - (g(u')u')' = 0$, that is, $du/dt - (g'(u')u' + g(u'))u'' = 0$. If $sg(s)$ is decreasing at some s and therefore with a negative derivative $-a$ at s, and if it happens that at some point x, $u'(x) = s$, then the equation looks near x like $du/dt + au'' = 0$. This is the inverse heat equation, which actually is known to be ill posed; starting with a smooth initial datum, it can develop singularities of any order in arbitrarily small time. (This equation is well known in signal processing, since inverting it would provide a perfect deblurring algorithm [22].)

The model which we now propose is a synthesis of Malik and Perona's ideas which avoids the above-mentioned difficulties; it will be robust in the presence of noise and will be consistent from the formal viewpoint described above. Moreover, it will be based, as in the space scale theory of Witkin, on a single parameter: the scale.

We shall define the "selective smoothing" of u_0 at scale $t^{1/2}$ as the function $u(x, t)$ verifying

$$\partial u/\partial t - \operatorname{div}(g(|DG_\sigma * u|)\nabla u) = 0 \quad \text{in }]0, T[\times \Omega, \qquad u(0) = u_0, \tag{1.3}$$

where $G_\sigma(x) = C\sigma^{-1/2} \exp(-|x|^2/4\sigma)$. It is easily seen that $G(x, t) = G_t(x)$ is nothing but the fundamental solution of the heat equation. Therefore, the term $(DG_\sigma * u)(x, t)$ which appears inside the divergence term of (1.3) is simply the gradient of the solution at time σ of the heat equation with $u(x, t)$ as initial datum. Thus it appears to be an estimate of the gradient of u at point x, obtained by the classical Marr–Hildreth–Witkin theory, recalled above. The modification of the model of Malik and Perona, therefore, is only to replace the gradient $|\nabla u|$ by its estimate $|DG_\sigma * u|$. As we shall prove, this slight change of the model is enough to avoid both inconsistencies of the Malik and Perona model. Indeed, the equation, as announced by those authors, will now diffuse if and only if the gradient is estimated to be small, the job of this estimate being done by the new term which we introduce. This does not alter the scope of the "anisotropic diffusion" model. Indeed, the initial datum in the new equation is u_0, and the necessity of smoothing the initial datum u_0 by a Gaussian in order to eliminate the noisy estimates of $|\nabla u|$ is now removed. Moreover, (1.3) can be proved to have a unique smooth solution. The main part of this paper is devoted to the proof of existence, uniqueness, and regularity properties for (1.3). We shall work with two-dimensional data-like pictures, and therefore defined, without loss of generality, on a square. (With the adequate changes in the definition of the convolution (see later), the domain where the signal is defined may be a rectangle, a disk, or any open set with smooth enough boundary.)

The function G to be considered can be any "low pass filter," or, to use the terminology of calculus, any smoothing kernel. However, in order 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 (1.3) is what time is best for "stopping" the evolution of the signal $u(x, t)$. This choice is quite important, since it is clear that all of the above-mentioned models, including ours, diffuse completely at $t \to +\infty$ and give therefore a constant function. 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 $u(x, t)$ at time t corresponds to a scale $t^{1/2}$. Indeed, roughly speaking, $u(x, t)$ appears in the Witkin model as a smooth version of u_0 obtained by convolving it with a filter of spatial width $t^{1/2}$.) Thus in the model (1.3) it is coherent to correlate the stopping time t and the time introduced via the estimator G_σ. We should therefore choose a stopping time t of the order of σ. Then the spatial scale under 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 scale at which edge information is lost will depend on the shape of the thresholding function g. Therefore, even if we wrote above that we should stop the equation at a time of the order of σ, this is rather a lower estimate and there is no inconsistency in looking at what happens to the signal $u(x, t)$ for times greater than σ. In experimentation, it might be convenient to play with two parameters that are anyway at hand in any edge detection model: the scale parameter (spatial width of the filtering) on the one side, $\sigma^{1/2}$, and the threshold 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 threshold for edges. Where $|\nabla u|$ 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, we might use the same threshold b on the gradient for keeping the edges as the one implicit in g.

Before beginning with the proofs of the consistency of our model, let us give a brief account of several related works which came to our knowledge as this work was in the final experimental phase. Osher and Rudin's theory [22] tries to get as close as possible to the inverse heat equation by defining some conservative scheme like $du/dt = -|u_x|F(u_{xx})$ with initial datum u_0. F is a function such that $sF(s) \geqq 0$. 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 $u(x, t)$ becomes discontinuous in x.

Nordström [21] proposes a new presentation of Perona and Malik's theory which relates it to the well-known variational global edge detection methods of Terzopoulos, Blake and Zisserman, and Mumford and Shah [18]. Nordström introduces a new term in (1.2) which forces $u(x, t)$ to remain close to u_0. Because of the forcing term $u - u_0$, the new equation

$$du/dt - \operatorname{div}(g(\nabla u)\nabla u) = u_0 - u$$

has a priori the advantage of having a nontrivial steady state, eliminating therefore 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.) However, there is the same theoretical difficulty as in Malik and Perona's model for this equation, at least for the g proposed by Nordström; $tg(t)$ decreases at infinity and the author insists in wanting this phenomenon to happen. The maximum γ of $tg(t)$ should play the role of edge thresholding; if the gradient is locally greater than γ, then there would be enhancement (and as a matter of fact, the equation behaves locally as an inverse heat equation). Conversely, if the gradient happens to have a norm smaller than γ, then diffusion will continue. This fact eliminates every hope of getting a uniqueness and regularity theorem for the solution $u(x, t)$. Therefore, the maximum principle stated in the same paper is empty; the C^2 assumption for the solution $u(x, t)$ generally is not verified. Thus the observed numerical stability of the attained solution is not likely to be theoretically justified and should depend strongly on the kind of discretization that was considered. However, the model that we propose could explain this stability. Indeed, discrete schemes are likely to introduce an implicit diffusion and therefore to "blurr" the computed solution.

Our scheme contains this blurring explicitly and, as a matter of fact, we "see" no significant difference between the experimental results obtained by Perona and Malik, by Nordström, and by Nitzberg and Shiota [20], whose work we shall comment on now. The approach of these authors is related to adaptive filtering methods which were introduced for TV images denoising by Graham [10]. The idea is to blur selectively and anisotropically the signal with "oblong" Gaussians. The Gaussian used for blurring at a point x 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 will be conserved. Since a corner is the crossing point of two edges, there will be two directions of "nondiffusion" instead of one, and therefore corners will be well conserved by this method. Nitzberg and Shiota's method is truly "anisotropic" at each point, in contrast to Perona and Malik's diffusion equation, which is quite isotropic. 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 Malik and Perona's. Their method appears, therefore, to be an alternative solution to ours for avoiding ill posedness, and it provides true anisotropic analysis, which our model does not take into account. In our opinion, its only drawback is the increased number of parameters. In the same paper is proposed another scheme for making (1.2) well posed. The idea is quite similar to that proposed here. One just replaces the gradient ∇u by some smoothed version v of it by setting (we omit the scale parameters)

$$du/dt = \operatorname{div}((1+v)^{-1}\nabla u), \qquad dv/dt = G * \|\nabla u\|^2 - v.$$

Since the second equation is intended to be a regularization of the steady state $v = G * |\nabla u|^2$, it is clear that this system boils down to our proposed equation.

However, the case where G has zero variance, that is, no convolution is actually done, might still lead to a well-posed system. The same paper states a conjecture of Mumford according to which the preceding system (without convolution) should have global solutions and approach a step function as time tends to infinity. We therefore would have a model quite close to Osher and Rudin's mentioned above, for which this last property is ensured.

Finally, let us mention an approach very different in technique but quite close in spirit and experimental results, namely, that due to Mallat and Zhong [14]. Mallat and Zhong use wavelet transform as a multiscale edge detection device, and are able to reconstruct an image from its edges. They use this new reconstruction technique to build a selective edge removing. The idea is to keep only edges present at several scales, to remove the other ones, and then apply the reconstruction algorithm to this simplified edge map. The result is visually almost equivalent to our version of anisotropic diffusion. Indeed, we also introduced a conditional edge removing by plugging the convolution by a Gaussian into our equation. In essence, that means that an edge is kept only if present at a coarser scale.

In the next section we shall state and prove the main existence and uniqueness results announced above. In § 3, we describe briefly the numerical scheme which was used in experimentations and present some experimental results on pictures. As we said, they are not much different from those obtained by the authors mentioned above. Now our scope is to obtain a correct and simple theory of what can be "nonlinear scale space," and to define mathematically experimental devices which should be stable and reliable.

2. Consistency of the model: Existence and uniqueness results. In the following, Ω denotes the open set $]0,1[\times]0,1[$ of $\mathbf{R}^2$, with boundary Γ. We denote by $H^k(\Omega)$, k a positive integer, the set of all function $u(x)$ defined in Ω such that u and its distributional derivatives $D^s u$ of order $|s| = \sum_{j=1}^n s_j \leqq k$ all belong to $L^2(\Omega)$. $H^k(\Omega)$ is a Hilbert space for the norm

$$\|u\|_{H^k(\Omega)} = \left(\sum_{|s|\leqq k} \int_\Omega |D^s u(x)|^2 \, dx \right)^{1/2}.$$

We denote by $L^p(0, T, H^k(\Omega))$ the set of all functions u, such that, for almost every t in $(0, T)$, $u(+)$ belong to $H^k(\Omega)$. $L^p(0, T, H^k(\Omega))$ is a normed space for the norm

$$\|u\|_{L^p(0,T,H^k(\Omega))} = \left(\int_0^T \|u(t)\|^p_{H^k(\Omega)} \right)^{1/p},$$

$p > 1$ and k a positive integer.

We denote by $L^\infty(0, T, \mathscr{C}^\infty(\Omega))$ the set of all functions such that, for almost every t in $(0, T)$, $u(t)$ belong to $\mathscr{C}^\infty(\Omega)$. $L^\infty(0, T, \mathscr{C}^\infty(\Omega))$ is a normed space for the norm

$$\|u\|_{L^\infty(0,T,\mathscr{C}^\infty(\Omega))} = \inf\{C; \|u(t)\|_{\mathscr{C}^\infty(\Omega)} \leqq C, \text{ a.e. on } (0, T)\}.$$

We denote by $(H^1(\Omega))$ the dual of $H^1(\Omega)$.

Let $g: \mathbf{R}^+ \to \mathbf{R}^+$ be a decreasing function with $g(0) = 1$, $\lim_{t\to+\infty} g(t) = 0$ and $t \to g(\sqrt{t})$ is smooth. For instance, we can take

$$g(t) = \frac{1}{1+t^2}.$$

Let $\sigma > 0$, we suppose that G_σ is a Gaussian filter

$$G_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{|x|^2}{4\sigma} \right), \qquad x = (x_1, x_2) \in \mathbf{R}^2.$$

Let us now study the existence of a solution for the nonlinear partial differential equation

$$\frac{\partial u}{\partial t} - \operatorname{div}[g(|\nabla G_\sigma * u|)\nabla u] = 0 \quad \text{on }]0, T[\times\Omega,$$

$$u(0) = u_0.$$

The initial data u_0 is in $L^2(\Omega)$ and

$$|\nabla G_\sigma * u| = \left[\sum_{i=1}^2 \left(\frac{\partial G_\sigma}{\partial_{x_i}} * \tilde{u} \right)^2 \right]^{1/2},$$

where $\tilde{u}$ is a linear and continuous extension of u to $\mathbf{R}^2$.

The definition of $\tilde{u}$ depends on the boundary condition on Γ imposed on $u(x, t)$.

In the case of a Neumann boundary condition,

$$\tilde{u}(x, y) = u(-x, y), \quad -1 \leqq x \leqq 0, \quad 0 \leqq y \leqq 1,$$
$$\tilde{u}(x, y) = u(x, -y), \quad 0 \leqq x \leqq 1, \quad -1 \leqq y \leqq 0 \cdots.$$

THEOREM 2.1. *Let $u_0 \in L^2(\Omega)$; then we have a unique function $u(x, t)$ such that $u \in \mathscr{C}([0, T]; L^2(\Omega)) \cap L^2(0, T; H^1(\Omega))$, and verifying*

$$\frac{\partial u}{\partial t} - \operatorname{div}[g(|\nabla G_\sigma * u|)\nabla u] = 0 \quad \text{on }]0, T]\times\Omega,$$

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \Gamma\times]0, T], \tag{2.1}$$

$$u(0) = u_0,$$

where this system is verified in the distributional sense. Moreover, this unique solution is in $\mathscr{C}^\infty(]0, T[\times\bar{\Omega})$.

Proof. (a) *Existence of a solution.* In this first section, we show the existence of a weak solution of (2.1) by a classical fixed point theorem of Schauder [9], [19]. We introduce the space

$$W(0, T)=\left\{w\in L^2(0, T; H^1(\Omega)), \frac{dw}{dt}\in L^2(0, T; (H^1(\Omega))')\right\}.$$

This space is a Hilbert space for the graph norm [13]. Let $w\in W(0, T)\cap L^\infty(0, T, L^2(\Omega))$ such that

$$\text{(C)} \qquad \|w\|_{L^\infty(0,T,L^2(\Omega))}\leqq \|u_0\|_{L^2(\Omega)}$$

and (E_w) the problem

$$\left\langle\frac{\partial u}{\partial t}(t), v\right\rangle+\int_\Omega g(|(\nabla G_\sigma * w)(t)|)\nabla u(t)\nabla v=0 \quad \forall v\in H^1(\Omega) \quad \text{a.e. in } [0, T],$$

$$u(0)=u_0.$$

Since $w\in L^\infty(0, T; L^2(\Omega))$ and g, G are C^∞, one easily deduces that $g(|\nabla G_\sigma * w|)\in L^\infty(0, T; \mathscr{C}^\infty(\Omega))$. Thus, since g is decreasing, there exists a constant $\nu>0$ such that

$$g(|\nabla G_\sigma * w|)\geqq \nu \quad \text{a.e. in }]0, T[\times\Omega,$$

where ν depends only on g, G and $\|u_0\|_{L^2(\Omega)}$.

By classical results on parabolic equations [3], [4], we prove that the problem (E_w) as a unique solution $U(w)$ in $W(0, T)$ [1], [4]. Then we can deduce

$$\text{(2.2)} \qquad \|U(w)\|_{L^2(0,T;H^1(\Omega))}\leqq C_1,$$

$$\text{(2.3)} \qquad \|U(w)\|_{L^\infty(0,T;L^2(\Omega))}\leqq \|u_0\|_{L^2(\Omega)},$$

$$\text{(2.4)} \qquad \|U(w)\|_{L^2(0,T;(H^1(\Omega))')}\leqq C^3,$$

where C_1, C_2, and C_3 are constants which only depends on G, g, and u_0. These estimates lead us to introduce the subset W_0 of $W(0, T)$ defined by

$$W_0=\left\{w\in W(0, T);\ \|w\|_{L^2(0,T;H^1(\Omega))}\leqq C_1,\ \|w\|_{L^\infty(0,T,L^2(\Omega))}\leqq\|u_0\|_{L^2(\Omega)},\right.$$
$$\left.\left\|\frac{dw}{dt}\right\|_{L^2(0,T;(H^1(\Omega)'))}\leqq C_3,\ w(0)=u_0\right\}.$$

By (2.2)–(2.4), U is a mapping from W_0 into W_0. Moreover, W_0 is a nonempty, convex, and weakly compact subset of $W(0, T)$.

In order to use the Schauder theorem, we need to prove that the mapping u is weakly continuous from W_0 into W_0.

Since $W(0, T)$ is contained in $L^2(0, T; L^2(\Omega))$, with compact inclusion, this will provide u in W_0 such that $u=U(u)$.

Let (w_j) be a sequence in W_0 which converges weakly to some w in W_0 and $u_j=U(w_j)$.

By using (2.2), (2.4), and classical theorems of compact inclusion (the theorem of Rellich and Kondrachov [4]), the sequence (w_j) of W_0 contains a subsequence (w_j)

such that

$$u_j \to u \quad \text{weakly in } L^2(0, T; H^1(\Omega)),$$

$$\frac{du_j}{dt} \to \frac{du}{dt} \quad \text{weakly in } L^2(0, T; (H^1(\Omega))'),$$

$$u_j \to u \quad \text{in } L^2(0, T; L^2(\Omega)) \quad \text{and a.e. on } \Omega \times]0, T[,$$

$$\frac{\partial u_j}{\partial x_i} \to \frac{\partial u}{\partial x_i} \quad \text{weakly in } L^2(0, T; L^2(\Omega)), \qquad i = (1, 2),$$

$$w_j \to w \quad \text{in } L^2(0, T; L^2(\Omega)),$$

$$\frac{\partial G}{\partial x_i} * w_j \to \frac{\partial G}{\partial x_i} * w \quad \text{in } L^2(0, T; L^2(\Omega)) \quad \text{and a.e. on } \Omega \times]0, T[,$$

$$g(|\nabla G * w_j|) \to g(|\nabla G * w|) \quad \text{in } L^2(0, T; L^2(\Omega)),$$

$$u_j(0) \to u(0) \quad \text{in } (H^1(\Omega))'.$$

Then we can pass to the limit in the relation (Ew_j), which yields $u = U(w)$. Moreover, by the uniqueness of the solution of (E_w), the whole sequence $u_j = U(w_j)$ converges weakly in $W(0, T)$ to $u = U(w)$.

(b) *Regularity of the solution.* By a classical bootstrap argument, we know that $u(t) \in H^1(\Omega)$ for all $t > 0$; therefore $u(t) \in H^2(\Omega)$ for all $t > 0$. Then, by iterating, using the general theory of parabolic equations [2], we can deduce that u is a strong solution of (2.1) and $u \in \mathscr{C}^\infty(]0, T] \times \Omega)$.

(c) *Uniqueness of the solution.* Let $\bar{u}$ and $\hat{u}$ be two solutions of (2.1). We have, for almost every t in $[0, T]$,

$$\frac{d\bar{u}}{dt}(t) - \operatorname{div}(\bar{\alpha}(t)\nabla \bar{u}(t)) = 0, \quad \frac{\partial \bar{u}}{\partial n}(t) = 0, \quad \bar{u}(0) = u_0, \tag{2.5}$$

$$\frac{d\hat{u}}{dt}(t) - \operatorname{div}(\hat{\alpha}(t)\nabla \hat{u}(t)) = 0, \quad \frac{\partial \hat{u}}{\partial n}(t) = 0, \quad \hat{u}(0) = u_0, \tag{2.6}$$

where $\bar{\alpha}(t) = g(|(\nabla G_\sigma * \bar{u})(t)|)$ and $\hat{\alpha}(t) = g(|(\nabla G_\sigma * \hat{u})(t)|)$.

By using (2.5) and (2.6), we obtain

$$\frac{d}{dt}(\bar{u}(t) - \hat{u}(t)) - \operatorname{div}[\bar{\alpha}(t)(\nabla \bar{u}(t) - \nabla \hat{u}(t))] = \operatorname{div}[(\bar{\alpha}(t) - \hat{\alpha}(t))\nabla \hat{u}(t)]. \tag{2.7}$$

Now, multiplying (2.7) by $\bar{u}(t) - \hat{u}(t)$ and using (C),

$$\begin{aligned} &\frac{1}{2}\frac{d}{dt}(\|\bar{u}(t) - \hat{u}(t)\|^2_{L^2(\Omega)}) + \nu \|\nabla \bar{u}(t) - \nabla \hat{u}(t)\|^2_{L^2(\Omega)} \\ &\quad \leqq \|\bar{\alpha}(t) - \hat{\alpha}(t)\|_{L^\infty(\Omega)} \|\nabla \hat{u}(t)\|_{L^2(\Omega)} \|\nabla \bar{u}(t) - \nabla \hat{u}(t)\|_{L^2(\Omega)}. \end{aligned} \tag{2.8}$$

Moreover, since g and G are C^∞, we have

$$\|\bar{\alpha}(t) - \hat{\alpha}(t)\|_{L^\infty(\Omega)} \leqq C \|\bar{u}(t) - \hat{u}(t)\|_{L^2(\Omega)}, \tag{2.9}$$

where C is a constant which only depends on G, g, and u_0.

By using (2.8) and (2.9), we can write

$$\begin{aligned} &\frac{1}{2}\frac{d}{dt}(\|\bar{u}(t) - \hat{u}(t)\|^2_{L^2(\Omega)}) + \nu \|\nabla \bar{u}(t) - \nabla \hat{u}(t)\|^2_{L^2(\Omega)} \\ &\quad \leqq \frac{2}{\nu} C \|\bar{u}(t) - \hat{u}(t)\|^2_{L^2(\Omega)} \|\nabla \hat{u}(t)\|^2_{L^2(\Omega)} + \frac{\nu}{2} \|\nabla \bar{u}(t) - \nabla \hat{u}(t)\|^2_{L^2(\Omega)}, \end{aligned}$$

and so

$$\frac{d}{dt}(\|\bar{u}(t)-\hat{u}(t)\|^2_{L^2(\Omega)}) \leqq \frac{4C}{\nu}\|\bar{u}(t)-\hat{u}(t)\|^2_{L^2(\Omega)}\|\nabla\hat{u}(t)\|^2_{L^2(\Omega)}. \tag{2.10}$$

Since $\bar{u}(0)=\hat{u}(0)=u_0$, by using (2.10) and Gronwall's lemma [5] we obtain the announced uniqueness. □

3. An iterative scheme which converges to the solution of (2.1).

THEOREM 3.1. *Let $u_0 \in L^2(\Omega)$. The sequence $(u^n)_n$ defined by (E_n),*

$$\frac{du^{n+1}}{dt}(t)-\operatorname{div}(g(|\nabla G_\sigma * u^n(t)|\nabla u^{n+1}(t)))=0 \quad \text{a.e. on }]0,T[,$$

$$\frac{du^{n+1}}{dn}(t)=0 \quad \text{a.e. on }]0,T[,$$

$$u^{n+1}(0)=u_0$$

converges in $\mathscr{C}([0,T];L^2(\Omega))$ to the strong solution of (2.1).

Proof. We denote by $\alpha^n = g(|\nabla G * u^n|)$. By [4, Thm. X.10], the problem (E_n) has a unique solution u^n. It is clear that

$$\alpha^n \geqq g(\|\nabla G * u^0\|_{L^\infty(\Omega)}) \quad \text{a.e. on }]0,T[\times\Omega. \tag{3.1}$$

Now we prove that the sequence u^n converges in $\mathscr{C}([0,T];L^2(\Omega))$ to u, the strong solution of (2.1).

We can observe that (see the estimate (2.10))

$$\frac{d}{dt}(\|u^{n+1}(t)-u(t)\|^2_{L^2(\Omega)}) \leqq \frac{4C}{\nu}\|\nabla u(t)\|^2_{L^2(\Omega)}\|u^n-u(t)\|^2_{L^2(\Omega)}. \tag{3.2}$$

Moreover, by using Theorem 2.1, we have

$$\|u^0-u(t)\|^2_{L^2(\Omega)} \leqq C_0 \quad \forall t \in [0,T], \tag{3.3}$$

where C_0 is a constant which only depends on G, g, and u_0. Then for any $t \in [0,T]$, we have

$$\|u^1(t)-u(t)\|^2_{L^2(\Omega)} \leqq C_0 \int_0^t a(s)\,ds, \tag{3.4}$$

where

$$a(s)=\frac{4C}{\nu}\|\nabla u(s)\|^2_{L^2(\Omega)}.$$

By (3.2) and (3.4), we can deduce

$$\frac{d}{dt}(\|u^2(t)-u(t)\|^2_{L^2(\Omega)}) \leqq C_0 a(t)\int_0^t a(s)\,ds,$$

and thus

$$\|u^2(t)-u(t)\|^2_{L^2(\Omega)} \leqq C_0\frac{1}{2}\left(\int_0^t a(s)\,ds\right)^2.$$

Finally, we obtain by iterating

$$\|u^{n+1}(t)-u(t)\|^2_{L^2(\Omega)} \leqq C_0\frac{1}{(n+1)!}\left(\int_0^T a(s)\,ds\right)^{n+1}. \tag{3.5}$$

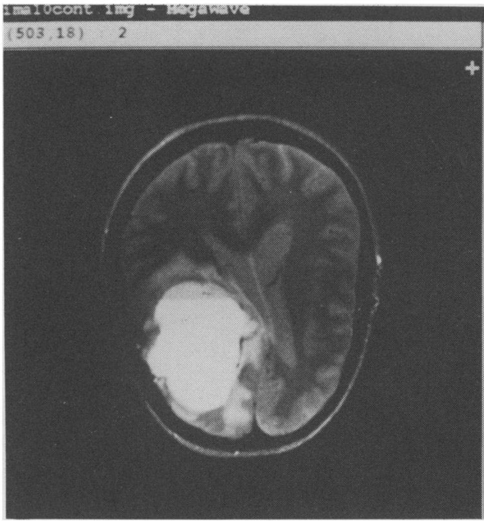


FIG. 1

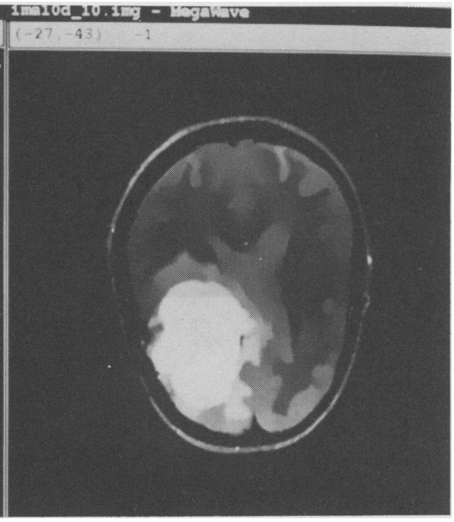


FIG. 2

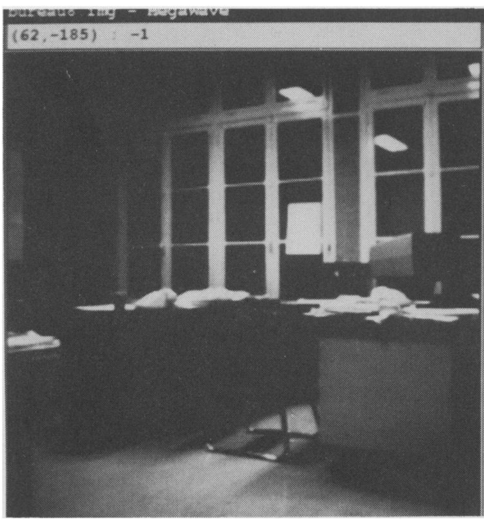


FIG. 3

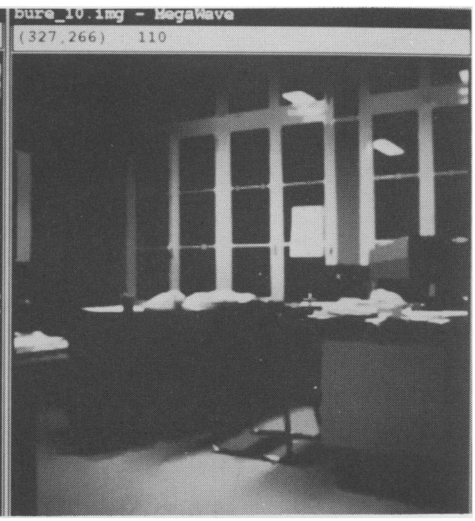


FIG. 4

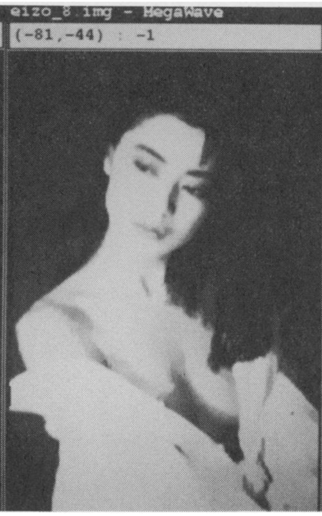
From (3.5) we conclude that the sequence $(u^n)_n$ converges in $\mathscr{C}([0, T]; L^2(\Omega))$ to the strong solution of (2.1). $\square$

4. Numerical results and experiments.

4.1. Discretisation. We introduce the lattice of coordinates $(ih, jh, n\Delta t)$ where $h = 1/N+1$, $0 \leqq i \leqq N+1$, $0 \leqq j \leqq N+1$; and we denote $u^n_{i,j}$ an approximation of $u(ih, jh; n\Delta t)$; $\alpha^n_{i,j}$ an approximation of $g(|\nabla G * u|)(ih, jh, n\Delta t)$. Then we discretise $[g(|\nabla G * u|)\partial u/\partial x]$ by $\alpha^n_{i,j}\partial u/\partial x(ih, jh, (n+1)\Delta t)$ and $\partial/\partial x[g[|(\nabla G * u)|]\partial u/\partial x]$ by

$$\frac{1}{2h^2}[(\alpha^n_{i-1,j} + \alpha^n_{i,j})u^{n+1}_{i-1,j} - (2\alpha^n_{i,j} + \alpha^n_{i-1,j} + \alpha^n_{i+1,j})u^{n+1}_{i,j} + (\alpha^n_{i,j} + \alpha^n_{i+1,j})u^{n+1}_{i+1,j}],$$

similarly for $\partial/\partial y[g|(\nabla G * u)|\partial u/\partial y]$ by exchanging the roles of i and j.



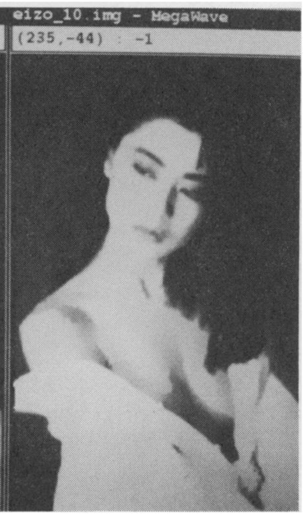


FIG. 5 FIG. 6 FIG. 7

Finally, we obtain the implicit scheme

$$\begin{aligned}\frac{u_{i,j}^{n+1}-u_{i,j}^{n}}{\Delta t}-\frac{1}{2h^2}&[(\alpha_{i-1,j}^n+\alpha_{i,j}^n)u_{i-1,j}^{n+1}+(\alpha_{i,j-1}^n+\alpha_{i,j}^n)u_{i,j-1}^{n+1}\\&+(\alpha_{i,j}^n+\alpha_{i+1,j}^n)u_{i+1,j}^{n+1}+(\alpha_{i,j}^n+\alpha_{i,j+1}^n)u_{i,j+1}^{n+1}\\&-(4\alpha_{i,j}^n+\alpha_{i-1,j}^n+\alpha_{i,j-1}^n+\alpha_{i+1,j}^n+\alpha_{i,j+1}^n)u_{i,j}^{n+1}]=0,\\ u_{i,j}^0=u_0(ih,jh),&\quad 1\leqq i\leqq N,\quad 1\leqq j\leqq N,\\ u_{i,0}^{n+1}=u_{i,1}^{n+1},\quad u_{N,j}^{n+1}=u_{N+1,j}^{n+1},&\quad 0\leqq i\leqq N+1,\quad 0\leqq j\leqq N+1,\\ u_{0,j}^{n+1}=u_{1,j}^{n+1},\quad u_{i,N}^{n+1}=u_{i,N+1}^{n+1},&\quad 0\leqq i\leqq N+1,\quad 0\leqq j\leqq N+1.\end{aligned}$$

Now the discrete problem can be written

$$\frac{u^{n+1}-u^n}{\Delta t}+A_h(u^n)u^{n+1}=0,\qquad n\geqq 0,$$

where the matrix A_h is tridiagonal by blocks and positive defined. By classical arguments [7], we know that the matrix $I+\Delta t A_h(u^n)$ is invertible.

4.2. Some comments on the results of applying the schema to several images. Figures 1 and 2 show a tomography picture from the University of Joseph Fourier at Grenoble and its "cleaned" version by anisotropic diffusion. As indicated in the introduction, the experiments depend on two parameters: the time t (which yields the "scale" of diffusion) and the threshold Θ under which diffusion is strong and above which diffusion is low. In Fig. 2, $t=2$ and $\Theta=10$. Figure 3 is an original picture by the Institut National de Recherche en Informatique et en Automatique, and Fig. 4 its diffusion with $t=2$ and $\Theta=20$. Notice how texture present in many parts of the picture is removed while the edges are kept. Figure 5 is an original Eizo picture. Figure 6 shows its diffusion with $t=2$ and $\Theta=20$.

Acknowledgment. J. M. Morel thanks J. I. Diaz for valuable conversations.

REFERENCES

[1] R. A. ADAMS, *Sobolev Spaces*, Academic Press, New York, 1975.

[2] O. A. LADYSHEUSKAYA, V. A. SOLOMIKOV, AND N. N. URAL'TSEVA, *Linear and Quasi linear Equation of Parabolic Type*, American Mathematical Society, Providence, RI, 1968.

[3] PH. BENILAN AND H. BREZIS, *Solutions faibles d'équations d'évolution dans les espaces de Hilbert*, Ann. Inst. Fourier, 22 (1972), pp. 311-329.

[4] H. BREZIS, *Analyse Fonctionnelle, Théorie et Applications*, Masson, Paris, 1987.

[5] ———, *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, Math. Stud. North-Holland, Amsterdam, 1973.

[6] J. CANNY, *Finding edges and lines in images*, Tech. Report 720, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Boston, MA, 1983.

[7] P. CIARLET, *Introduction à l'Analyse Numérique Matricielle et à l'Optimisation*, Masson, Paris, 1982.

[8] J. I. DIAZ, *A nonlinear parabolic equation arising in image processing*, Extracta Matematicae, Universidad de Extremadura, 1990.

[9] G. GAGNEUX, *Sur des problèmes unilatéraux dégénérés de la théorie des écoulements diphasiques en milieu poreux*, Thèse d'Etat, Université de Besançon, 1982.

[10] R. E. GRAHAM, *Snow removal: A noise-stripping process for TV signals*, IEEE Trans. Inform. Theory, 9 (1962) pp. 129-144.

[11] K. HOLLIG AND J. A. NOHEL, *A diffusion equation with a nonmonotone constitutive function*, in Proceedings on Systems of Nonlinear Partial Differential Equation, J. Ball, ed., D. 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. L. LIONS, *Contrôle Optimal de Systèmes Gouvernés par des Équations aux Dérivées Partielles*, Dunod, Paris, 1968.

[14] S. MALLAT AND S. ZHONG, *Complete signal representation with multiscale edges*, Tech. report 483, Robotics Report 219, Computer Science Division, Courant Institute, New York.

[15] D. MARR, *Vision*, W. H. Freeman, San Francisco, CA, 1982.

[16] D. MARR AND E. HILDRETH, *Theory of edge detection*, Proc. Roy. Soc. London Ser. B, (1980), pp. 187-217.

[17] J. M. MOREL AND S. SOLIMINI, *Segmentation of images by variational methods: a constructive approach*, Rev. Mat. Universidad Complutense Madrid, 1-3 (1988), pp. 169-182.

[18] D. MUMFORD AND J. SHAH, *Boundary detection by minimizing functionals*, IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 1985.

[19] L. NIRENBERG, *Topics in nonlinear functional analysis*, Lecture Notes, New York University, New York, 1974.

[20] 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.

[21] K. N. NORDSTRÖM, *Biased anisotropic diffusion—A unified approach to edge detection*, preprint, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, 1989.

[22] S. OSHER AND L. RUDIN, *Feature-oriented image enhancement using shock filters*, SIAM J. Numer. Anal., to appear.

[23] P. PERONA AND J. MALIK, *Scale space and edge detection using anisotropic diffusion*, in Proc. IEEE Computer Society Workshop on Computer Vision, 1987.

[24] A. ROSENFELD AND M. THURSTON, *Edge and curve detection for visual scene analysis*, IEEE Trans. Comput., C-20 (1971), pp. 562-569.

[25] A. P. WITKIN, *Scale-space filtering*, in Proceedings of IJCAI, Karlsruhe, 1983, pp. 1019-1021.

[26] A. YUILLE AND T. POGGIO, *Scaling theorems for zero crossings*, IEEE Trans. Pattern Analysis Machine Intelligence, 8 (1986), pp. 15-25.