

FEATURE-ORIENTED IMAGE ENHANCEMENT USING SHOCK FILTERS*

STANLEY OSHER† AND LEONID I. RUDIN†

Abstract. Shock filters for image enhancement are developed. The filters use new nonlinear time dependent partial differential equations and their discretizations. The evolution of the initial image $u_0(x, y)$ as $t \rightarrow \infty$ into a steady state solution $u_\infty(x, y)$ through $u(x, y, t)$, $t > 0$, is the filtering process. The partial differential equations have solutions which satisfy a maximum principle. Moreover the total variation of the solution for any fixed $t > 0$ is the same as that of the initial data. The processed image is piecewise smooth, nonoscillatory, and the jumps occur across zeros of an elliptic operator (edge detector). The algorithm is relatively fast and easy to program.

Key words. shock filters, total variation, image enhancement, deconvolution

AMS(MOS) subject classifications. primary 68U10; secondary 65M05, 65M10

1. Introduction. A basic step in the processing of signals (images, radar, acoustic signals) is enhancement. By this we mean the removal of blurring.

An elementary example of blurring comes from the degradation of a signal through some kind of convolution. More precisely, let

$$x = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

and let $u(x)$ be the original real valued function which is blurred through convolution with a kernel: $j(x)$

$$(1.1) \quad w_0(x) = j \star u = u \star j = \int_{\mathbb{R}^n} j(x-y)u(y) dy.$$

Typically, j has the following properties:

$$(1.2) \quad \begin{aligned} j(x) &\geq 0, \\ j(x) &\rightarrow 0 \text{ rapidly as } |x| \rightarrow \infty, \text{ where } |x| = (x_1^2 + x_2^2 + \dots + x_n^2)^{1/2}, \\ \int_{\mathbb{R}^n} j(x) ds &= 1. \end{aligned}$$

Examples include for any $t > 0$:

$$(1.3a) \quad j^{(1)}(x) = \frac{1}{(4\pi t)^{n/2}} e^{-(|x|^2/4t)} \quad (\text{the heat kernel}),$$

$$(1.3b) \quad \begin{aligned} j^{(2)}(x) &= j^{(2)}(|x|), \\ j^{(2)}(|x|) &\text{ is strictly decreasing with } |x| \text{ for } |x| < 1, \\ j^{(2)}(|x|) &\equiv 0 \text{ if } |x| \geq 1, \\ j &\in C^\infty. \end{aligned}$$

Then consider for any $\delta > 0$

$$j_\delta^{(2)} = \frac{1}{\delta^n} j^{(2)}\left(\frac{|x|}{\delta}\right).$$

* Received by the editors February 13, 1989; accepted for publication (in revised form) November 2, 1989.

† Cognitech Inc., 2020 Broadway, Santa Monica, California 90404.

We call $j_\delta^{(2)}$ a “mollifying” kernel:

$$(1.3c) \quad j_\delta^{(3)}(x) \equiv \frac{1}{\delta^n} \left\{ x / -\frac{\delta}{2} \leq x_i \leq \frac{\delta}{2}, i = 1, \dots, n \right\},$$

$$j_\delta^{(3)}(x) \equiv 0 \quad \text{otherwise,}$$

$j_\delta^{(3)}$ is the “box” kernel.

If we take the Fourier transform of (1.1) we arrive at

$$(1.4) \quad \hat{w}_0(\xi) = \hat{j}(\xi) \hat{u}(\xi).$$

To recover $u(x)$, we need to deconvolve, i.e., to reverse the procedure in (1.1). From (1.4), this amounts to dividing by $\hat{j}(\xi)$ and applying the inverse Fourier transform. The problem with this procedure, of course, is that this is generally very ill-posed. Since j is usually relatively smooth, $\hat{j}(\xi) \rightarrow 0$ rapidly as $|\xi| \rightarrow \infty$, and large frequencies in $\hat{w}_0(\xi)$ get amplified considerably.

The function $w(x)$ is often taken to be band limited, i.e., $\hat{w}(\xi) \equiv 0$ for $|\xi|$ large enough, say for $|\xi_i| > N$ for each $i = 1, \dots, n$. An important example of this comes through discretization. Let $w_h(x)$ be a grid function defined at grid points

$$(1.5) \quad \begin{aligned} (x_i)_l &= l_i h, \\ i &= 1, \dots, n; \quad l = (l_1, \dots, l_n), \\ l_i &= 0, \pm 1, \dots, \pm N; \quad (2N+1)h = 1, \\ x_l &= lh, \end{aligned}$$

and suppose w_h is extended to be periodic on the grid with

$$w_h(x + e_i) \equiv w_h(x)$$

for each $e_i = (0, 0, \dots, 1, 0, \dots, 0)$, where the 1 occurs in the i th component, $i = 1, \dots, N$.

The grid function has a unique trigonometric interpolant

$$(1.6) \quad Iw_h(x) = \sum (\tilde{w}_h)_l e^{il \cdot x},$$

where the sum is taken over the cube: $-N \leq l_i \leq N$, $i = 1, \dots, N$ and

$$(1.7) \quad Iw_h(x_l) = w_h(x_l)$$

for each grid point x_l .

See, e.g., [7] for a description of this interpolant and its properties.

Thus deconvolution on a grid amounts to multiplying each discrete Fourier coefficient $(\tilde{w}_h)_l$ by $(\hat{j}(l))^{-1}$. This number typically grows like $|l|$ to some positive power, or even like the exponential of a positive power of $|l|$. If w_h is the discretization of a very smooth function, then the coefficients $(\tilde{w}_h)_l$ decay rapidly—perhaps more rapidly than the growth in $(\hat{j}(l))^{-1}$. However, there is generally high frequency noise in all this, i.e., the discretization itself can well introduce nonsmooth, but low amplitude, roundoff errors. Deconvolution in this simple fashion will amplify this noise in a very unstable manner.

The situation is far worse if the underlying function w , which is sampled on grid points, is only piecewise continuous—if it has jumps or jumps in derivatives. Then there exist two problems.

The first is global. It was shown in [9] that there is a *global* error between the interpolant Iw_h and w which is, e.g., $O(h)$ for functions that have jumps. This is true

globally, i.e., at any finite distance from the discontinuity. More seriously, any attempt to approximate w by either a Fourier interpolant I_{w_h} , or a truncated Fourier transform of w , e.g., by considering

$$(1.8) \quad \hat{\rho}(\xi) \hat{w}(\xi)$$

for

$$\begin{aligned} \hat{\rho}(\xi) &\equiv 1 && \text{if } |\xi| \leq m_1, \\ \hat{\rho}(\xi) &\equiv 0 && \text{if } |\xi| > m_2, \end{aligned}$$

will lead to Gibbs' phenomena. These are $O(1)$ errors near the discontinuity in $w(x)$, which do not disappear as $m_1, m_2 \rightarrow \infty$ (or $h \downarrow 0$). These errors cause the well-known phenomenon of "ringing" in image processing. This is particularly problematic for the machine processing of images. Thus, an attempt to deconvolve without amplifying high frequency errors due to noise, by cutting off the high frequencies, will lead to severe oscillations near the discontinuities of the original function w .

The procedures discussed so far are all linear and/or all involve the Fourier transform of w . This way of thinking is inherently problematic for the processing of images that are only piecewise smooth.

We remark that feature-dependent image processing was tried in [18]. There a filter was used which is the sum of two components—a linear low pass and a linear high pass filter. This leads to a generally oscillatory procedure of limited practical value. It does, however, represent a first attempt to perform a context sensitive enhancement.

It was pointed out in [14], [15] that images are dominated by the geometry of their features—edges, corners, lines, etc. In fact, the space of functions of bounded variation appears to be the correct class for image analysis.

In [14] the concepts and techniques developed in the numerical solution of nonlinear hyperbolic equations were applied for the first time to feature-oriented image enhancement. There the first experimental shock filter, based on a modification of the nonlinear Burgers' equation was used. However, this first shock filter did not incorporate the crucial feature-detector switch, the total variation preserving (TVP) computational approach, and the theoretical basis developed herein.

Both subjects (image enhancement, shock calculations) deal with the discrete representation of discontinuous functions. The relevant concepts include: characteristic speed, variation diminishing, or essentially nonoscillatory approximations, the need for nonlinear approximations to linear problems, and compressive methods. See, e.g., [1], [13] for overviews of this subject.

In this paper, the first of a series, we develop shock filters for image enhancement. (Our next paper will consider deconvolution in the presence of noise.) The filters use nonlinear time-dependent partial differential equations. The evolution of the initial data $u_0(x)$ into a steady state solution $u_\infty(x)$ as $t \rightarrow \infty$ through $u(x, t)$, $t > 0$, is the filtering process. The partial differential equations have solutions that satisfy a maximum principle and more. In fact, the total variation of the solution for any fixed positive time is the same as that of the initial data, i.e., the operator is total variation preserving. The steady state solution is achieved relatively quickly in most cases, and this is the processed image. The initial data is, of course, the discretization of the original image.

The processed image is piecewise smooth. In fact, it is a solution of any one of a class of second-order elliptic partial differential equations in regions of smoothness. The jumps occur across the zeros of the elliptic operator applied to the initial data.

The essential features of the true image are recovered in many cases—these include: number of jumps, relative size of jumps, and location of jumps. In some more special cases, given the knowledge of the method of blurring, our results can be made exact if the blurring (e.g., δ in (1.3b)) is not too large.

We will draw from experience in the numerical solution of hyperbolic problems that have discontinuous solutions. Such a problem is exemplified by a scalar conservation law

$$(1.9a) \quad u_t + f(u)_x = 0$$

to be solved for $-\infty < x < \infty (x \in R^1)$, $t > 0$ with initial data:

$$(1.9b) \quad u(x, 0) = u_0(x).$$

If $f'' \neq 0$, then the solution generally develops discontinuities even for very smooth $u_0(x)$.

For a small time the solution is constant along characteristics; i.e.,

$$(1.10) \quad u(x, t) = u_0(x - f'(u)t)$$

for t sufficiently small. This means that u is constant along characteristics: $x - f'(u)t = \text{constant}$. In finite time these characteristics generally intersect and shocks develop, i.e., the solution becomes a weak solution. This solution is typically piecewise continuous with jumps across curves satisfying

$$\frac{dx}{dt} = \frac{[f]}{[u]},$$

whose $[]$ denotes the jump across the curve.

The modern way to solve this problem numerically uses a shock-capturing method. The solution, shocks and all, is obtained through a single, globally defined algorithm.

We approximate (1.8) by setting up a grid: $x_i = ih$, $t^n = n\Delta t$

$$i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, \dots$$

A shock-capturing approximation is, by definition, in conservation form:

$$(1.11) \quad u_i^{n+1} = u_i^n - \frac{\Delta t}{h} (g_{i+1/2}^n - g_{i-1/2}^n),$$

where

$$(1.12) \quad g_{i+1/2}^n = g(u_{i-k}^n, \dots, u_{i+k+1}^n)$$

is the numerical flux approximating $f(u)$; g is Lipschitz continuous with:

$$g(u, u, \dots, u) = f(u).$$

This guarantees that bounded almost everywhere convergent sequences of solutions to (1.11) will yield weak solutions of (1.9) [8].

Of course, much more than conservation form is needed for a good scheme. Obvious issues involve stability and convergence. A more subtle issue involves accuracy in the presence of discontinuities. Schemes which are simple (e.g., linear if $f(u)$ is linear, $f(u) = au$) and which do not generate oscillations near discontinuities of the solution must, of necessity, be only first-order accurate (see, e.g., [2], [6]). Thus they smear discontinuities badly.

A great deal of successful work has been done to overcome this limitation. See, e.g., [5], [16], [17]. The goal is to get highly accurate (in regions of smoothness) methods that resolve discontinuities sharply and accurately in a nonoscillatory way.

In the next sections we will borrow ideas from shock calculations, modify them appropriately, and apply them to image enhancement.

The format of this work is as follows.

In § 2, we will discuss a one-dimensional enhancement procedure. We set up a new partial differential equation (PDE) which acts as our enhancement filter and describe its (surprising) properties. We then set up a total variation preserving (TVP) approximation to this PDE, describe its properties, and perform numerical experiments.

In § 3, we extend the procedure of § 2 to two space dimensions, i.e., to real images. The rigorous mathematical theory of this problem in two dimensions is not as extensive. However, we do have a local discrete maximum principle for this algorithm. The numerical schemes seem to yield the described enhancement features. This will be demonstrated again by numerical experiments on real pictures in § 4.

A big difference between one and two dimensions comes in the freedom of choice regarding which second-order elliptic operator to use in the enhancement PDE. This is equivalent to the choice of edge detector used. The zero level curves of this operator (generalized inflection points) applied to the initial data will give us the location of the edges of our enhanced image. The final enhanced image will be a piecewise smooth function that satisfies the homogeneous elliptic equation in regions of smoothness.

The numerical method we use below to enhance images is fast. It is an $O(kN)$ method. Here N is the number of points (pixels) on the screen and k is the number of time iterations. In practice, significant improvement occurs very quickly, and therefore k is quite small. The method also involves only very local operations—communication between pixels located more than two or three points apart is nil. Thus the code parallelizes quite easily.

The method results in a piecewise constant enhanced image in one dimension; however, as we stated above, the result is a piecewise smooth function which locally solves a homogeneous elliptic equation in two dimensions. In one dimension we show below that our enhancement is an exact deconvolution for data that was originally piecewise constant, then blurred, as long as the edges were not originally too close. This is true quite independently of the details of the convolution kernel used to blur the image. We believe a similar result to be true in two dimensions for piecewise smooth functions that are locally solutions to the above-mentioned elliptic equation.

We mention that any noise in our blurry image will also (unfortunately) be enhanced by our procedure. We are currently using related nonlinear partial differential equation type ideas to obtain an edge-preserving noise removal algorithm. We will report on this in the near future.

Finally, we note that the usefulness of our procedure is limited when a significant amount of information is lost through the blurring process, i.e., features may merge or change topologically. In this case we expect to develop and use more sophisticated shock filters together with information about the blurring degradation parameters.

2. One-dimensional enhancement. We begin by considering the equation

$$(2.1a) \quad u_t = -|u_x|F(u_{xx})$$

to be solved for all x , $t \geq 0$ with initial data:

$$(2.1b) \quad u(x, 0) = u_0(x).$$

Here the Lipschitz continuous function F satisfies

$$(2.1c) \quad \begin{aligned} & \text{(i) } F(0) = 0, \\ & \text{(ii) } X(u)F(u) > 0, \quad u \neq 0, \end{aligned}$$

where $X(u) = 1$ if $u > 0$, $X(u) = -1$ if $u < 0$, $X(0) = 0$.

An example is:

$$(2.2) \quad u_t = -|u_x|u_{xx}.$$

This looks at first like an extremely ill-posed problem. The coefficient of u_{xx} is never positive. In fact it is negative except at extrema of u where it vanishes. This initial value problem turns out to be well posed, or at least to satisfy the following a priori estimates (which are true for (2.1a-c) with general F satisfying (2.2))

$$(2.3) \quad \begin{aligned} (i) \quad & \text{TV } u(\cdot, t) = \text{TV } u_0(\cdot), \\ (ii) \quad & \max u(\cdot, t) = \max u_0(\cdot), \\ (iii) \quad & \min u(\cdot, t) = \min u_0(\cdot). \end{aligned}$$

The intuitive reason for the validity of these estimates is as follows. The solution to (2.2) stays smooth initially at smooth extrema. Thus, since u_x vanishes at these points, u_t also vanishes and maxima and minima stay invariant in time.

This same argument works for

$$(2.4) \quad u_t = |u_x|u_{xx}.$$

Also the time-reversed version of our approximation to (2.2) (see (2.10a) with the $-$ sign made $+$) satisfies a discrete version of (2.3). However, P. L. Lions has pointed out to us (private communication) that viscosity solutions (see e.g., [19]) to (2.4) written in conservation form develop singularities immediately at extrema in such a way that $(\partial/\partial x)(|u_x|u_x/2)$ stays away from zero; the resulting value of u_t is then such that the extrema diminish in magnitude as time increases. The same is true for monotone conservation form approximations to (2.4), and we have found that the time-reversed version of such schemes which approximate (2.2) apparently become unstable very quickly. One such monotone conservation form scheme is described below in (2.21).

Thus, we are not seeking solutions to (2.2) in the class of time-reversed viscosity solutions. Instead we are committed to variation preservation as a constraint on solutions of (2.2) in particular, and (2.1a) in general.

We can best exemplify the behavior of this enhancement procedure by taking a case which is outside our definition, in that F is not Lipschitz continuous. Let $F(u) = X(u)$.

Then u satisfies:

$$(2.5a) \quad u_t = \pm u_x \quad \text{if } u_{xx} \neq 0,$$

$$(2.5b) \quad u_t = 0 \quad \text{if } u_{xx} = 0.$$

We thus have a simple linear advection equation in which the direction of propagation changes sign or becomes zero at extrema and at inflection points of u . An interesting example comes from taking

$$(2.5c) \quad u_0(x) = \cos x.$$

Then

$$(2.6) \quad \begin{aligned} u(x, t) &= \cos(x - t) \quad \text{for } t < x < \frac{\pi}{2}, \\ u(x, t) &= 1 \quad \text{for } -t < x < t, \\ u(x, t) &= \cos(x + t) \quad \text{for } -\frac{\pi}{2} < x < -t. \end{aligned}$$

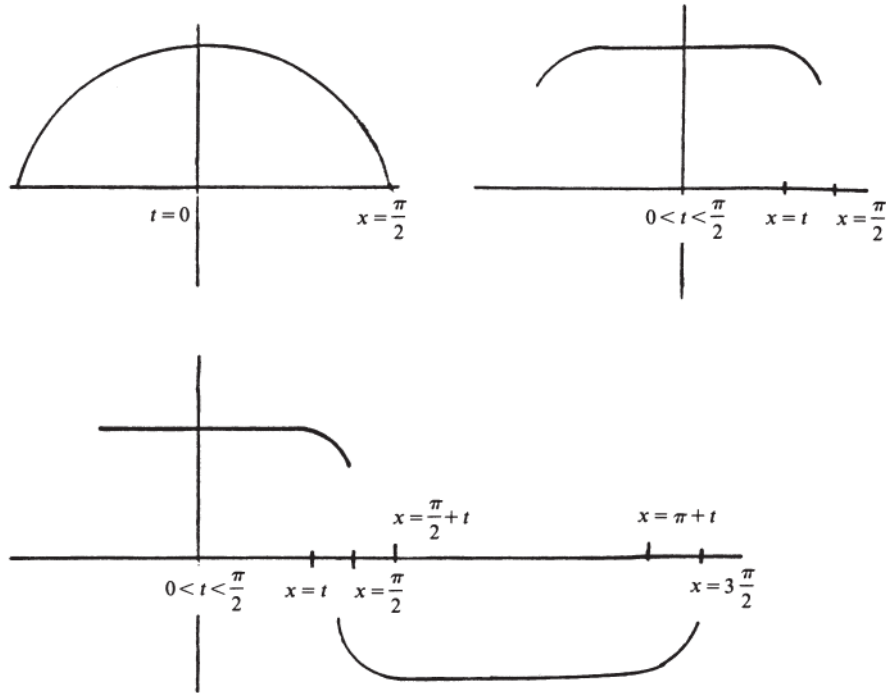


FIG. 2.1

By symmetry, discontinuities develop at $x = (2k + 1)\frac{\pi}{2}$, $k = 0, \pm 1, \pm 2, \dots$. Finally, at $t = \frac{\pi}{2}$ and for all $t > \frac{\pi}{2}$, the enhanced image is:

$$u_\infty = (-1)^k \quad \text{for } (2k - 1)\frac{\pi}{2} < x < (2k + 1)\frac{\pi}{2}.$$

To show that the transformation

$$u_0 \rightarrow u_\infty$$

is an approximate deconvolution, we consider what it does to convolution of a step function by $j_\delta^{(2)}(x)$ as defined in (1.3b).

Let $w(x)$ be a step function with values

$$(2.7) \quad w(x) \equiv w_\nu \quad \text{in } I_\nu = \{x/x_{\nu-1/2} \leq x < x_{\nu+1/2}\},$$

where the real line is the union of all these I_ν 's. Let $0 < d = \inf_\nu [x_{\nu+1/2} - x_{\nu-1/2}]$. Finally, take $0 < \delta < \frac{d}{2}$ be given.

Define

$$(2.8) \quad u_0 = j_\delta^{(2)} \star w.$$

It is easy to see that

$$u_0(x) \equiv w_\nu \quad \text{for } x_{\nu-1/2} + \delta < x < x_{\nu+1/2} - \delta,$$

$$u_0'(x) = (w_{\nu+1} - w_\nu)j_\delta^{(2)}(x - x_{\nu+1/2}),$$

for

$$x_{\nu+1/2} - \delta < x < x_{\nu+1/2} + \delta,$$

$$u_0''(x) = (w_{\nu+1} - w_\nu)j_\delta^{(2)'}(x - x_{\nu+1/2}).$$

Thus, for x between $x_{\nu+1/2} - \delta$ and $x_{\nu+1/2} + \delta$, u_0 moves strictly monotonically from w_ν to $w_{\nu+1}$ after which it becomes constant until x reaches $x_{\nu+3/2} - \delta$. Also u_0 has only one inflection point in this interval, and that is at $x = x_{\nu+1/2}$. We have the following theorem.

THEOREM 2.1. *If $F = X$ in (2.1a), then the transformation $u_0 \rightarrow u_\infty$ with $u_0 = j_\delta^{(2)} \star w$ for $\delta < \frac{d}{2}$ yields $u_\infty = w$ where w is the step function defined in (2.7). Thus the shock filter is an exact deconvolution in this case.*

Remark (2.1). We are tacitly assuming that the discontinuities developing in (2.1) occur exactly where we want them—at inflection points of the initial data, where the characteristics intersect. This is borne out by all our numerical experiments.

Our theory for solutions to (2.1) comes from the following. If we differentiate (2.1) with respect to x , multiply by $X(u_x)$ and integrate the result with respect to x , we arrive at

$$(2.9) \quad \frac{\partial}{\partial t} \int |u_x| = - \int \frac{\partial}{\partial x} (u_x F(u_{xx})) = 0$$

if u is eventually constant for $|x|$ large. This proof is rigorously true as long as u stays smooth. Moreover, our discrete approximation has this property for general initial data. Thus we have the following conjecture.

CONJECTURE 2.1. The evolution equation, (2.1), with $u_0(x)$ continuous, has a unique solution $u(x, t)$ which has jumps only at inflection points of $u_0(x)$ and for which the total variation in x of $u(x, t)$ is invariant in time, as are the location and value of local extrema.

We now set up a discrete approximation to (2.1) which preserves the variation and the size and location of extrema. We approximate (2.1) by

$$(2.10a) \quad u_i^{n+1} = u_i^n - \frac{\Delta t}{h} |m(\Delta_+ u_i^n, \Delta_- u_i^n)| F\left(\frac{\Delta_+ \Delta_- u_i^n}{h^2}\right).$$

Here $m(x, y)$ is the minmod function defined by

$$(2.10b) \quad m(x, y) = \begin{cases} (\text{sign } x) \min(|x|, |y|) & \text{if } xy > 0, \\ 0 & \text{if } xy \leq 0, \end{cases}$$

and

$$(2.10c) \quad \Delta_\pm u_i = \pm(u_{i\pm 1} - u_i).$$

Call $F(\Delta_+ \Delta_- u_i^n / h^2) = F_i^n$. Then

$$(2.11) \quad \Delta_+ u_i^{n+1} = \Delta_+ u_i^n - \frac{\Delta t}{h} |m(\Delta_+ u_{i+1}^n, \Delta_+ u_i^n)| F_{i+1}^n + \frac{\Delta t}{h} |m(\Delta_+ u_i^n, \Delta_- u_i^n)| F_i^n.$$

We require (CFL restriction)

$$\sup_i \frac{\Delta t}{h} F_i^n \leq \frac{1}{2}.$$

It is now easy to see that the right side of (2.6) has the same sign as $\Delta_+ u_i^n$ (and vanishes if $\Delta_+ u_i^n = 0$). Thus we have

$$(2.12) \quad \begin{aligned} |\Delta_+ u_i^{n+1}| &= |\Delta_+ u_i^n| - \frac{\Delta t}{h} \chi_{i+1/2}^n |m(\Delta_+ u_{i+1}^n, \Delta_+ u_i^n)| F_{i+1}^n \\ &\quad + \frac{\Delta t}{h} \chi_{i+1/2}^n |m(\Delta_+ u_i^n, \Delta_- u_i^n)| F_i^n, \end{aligned}$$

where

$$\chi_{i+1/2}^n = \text{sign } \Delta_+ u_i^n.$$

By the definition of the minmod function, it follows that

$$(2.13) \quad \chi_{i+1/2}^n |m(\Delta_+ u_{i+1}^n, \Delta_+ u_i^n)| = \chi_{i+3/2}^n |m(\Delta_+ u_{i+1}^n, \Delta_+ u_i^n)|.$$

Thus we arrive at:

$$(2.14) \quad |\Delta_+ u_i^{n+1}| = |\Delta_+ u_i^n| - \frac{\Delta t}{h} \Delta_+ (\chi_{i+1/2}^n m)(\Delta_+ u_i^n, \Delta_- u_i^n) |F_i^n.$$

Finally

$$(2.15) \quad \text{TV}(u^{n+1}) = \sum |\Delta_+ u_i^{n+1}| = \sum |\Delta_+ u_i^n| = \text{TV}(u^n).$$

This is a total variation preserving (TVP) method. Moreover, if $u_{i_0}^{n+1}$ is a local maximum for u_i^{n+1} , then since $\text{sign } \Delta_+ u_{i_0}^{n+1} = \text{sign } \Delta_+ u_{i_0}^n$ and $\text{sign } \Delta_- u_{i_0}^{n+1} = \text{sign } \Delta_- u_{i_0}^n$, then $u_{i_0}^n$ is a local maxima for u_i^n . By the definition of minmod it follows that

$$u_{i_0}^{n+1} = u_{i_0}^n.$$

The same is true for $u_{i_0}^{n+1}$, a local minima. We now have the following theorem.

THEOREM 2.2. *The scheme (2.10) enforces a local maximum and minimum principle—in fact such local extrema remain unchanged in time. Moreover the scheme is TVP.*

Thus, for fixed $\Delta t, h$, as $n \rightarrow \infty$ the sequence of discrete solutions has a convergent subsequence; we call the limit $(u^\infty)_h$ the discrete processed image. (Uniqueness of the limit is unclear at this time.)

As $\Delta t \rightarrow 0, h \rightarrow 0$, the sequence of discrete solutions also has a convergent subsequence (the continuum processed image), and again uniqueness of the limit has not been proven.

Let

$$\max(x, 0) = x^+,$$

$$\min(x, 0) = x^-.$$

Then it turns out that we may rewrite (2.10a) equivalently as

$$(2.16) \quad \begin{aligned} u_i^{n+1} = u_i^n &- \frac{\Delta t}{h} \sqrt{((\Delta_+ u_i)^+)^2 + ((\Delta_- u_i)^-)^2} (F_i^n)^- \\ &- \frac{\Delta t}{h} \sqrt{((\Delta_+ u_i)^-)^2 + ((\Delta_- u_i)^+)^2} (F_i^n)^+. \end{aligned}$$

If we fix the value of F_i^n to be constant c , then (2.16) is Godunov's method approximating

$$(2.17) \quad u_i = -|u_x|c;$$

see, e.g., [2], [11]. We chose this method because, among three point approximations to (2.17) that are monotone, this is the most compressive, i.e., the least dissipative [11]. This scheme (2.16) is almost monotone in the sense that if $(F_i^n)^-$ is a fixed constant less than or equal to zero or $(F_i^n)^+$ is a fixed constant greater than or equal to zero, then the right-hand side is a nondecreasing function of the u_i^n .

The fact that (2.16) is of the form:

$$(2.18) \quad u_i^{n+1} = u_i^n + g^+(u_{i+1}^n, u_i^n, u_{i-1}^n)(F_i^n)^- + g^-(u_{i+1}^n, u_i^n, u_{i-1}^n)(F_i^n)^+,$$

where the right-hand side is a nondecreasing function of its arguments for “frozen” values of $(F_i^n)^-, (F_i^n)^+$, and where $g^+(c, c, c) = g^-(c, c, c) \equiv 0$, for any constant c , means that

$$(2.19) \quad \min(u_{i-1}^n, u_i^n, u_{i+1}^n) \leq u_i^{n+1} \leq \max(u_{i-1}^n, u_i^n, u_{i+1}^n).$$

This provides us with a different local maximum and minimum principle than those referred to in Theorem 2.2 above.

To approximate (2.4) written in conservation form, i.e., as:

$$(2.20) \quad u_t = \frac{\partial}{\partial x} \left(\frac{|u_x| u_x}{2} \right),$$

we might try a conservative scheme

$$(2.21a) \quad u_i^{n+1} = u_i^n + \frac{\Delta t}{2h} \Delta_- \left(\frac{|\Delta_+ u_i^n|}{h} \frac{(\Delta_+ u_i^n)}{h} \right),$$

which is monotone if

$$(2.21b) \quad \max_i \frac{\Delta t}{h^3} [|\Delta_+ u_i^n| + |\Delta_- u_i^n|] \leq 1.$$

This is guaranteed if

$$(2.21c) \quad 4 \left(\frac{\Delta t}{h^3} \max_i |u_i^0| \right) < 1$$

is true initially.

As $\Delta t, h \rightarrow 0$, this scheme will converge to the viscosity solution of this initial value problem. If the + is replaced by - in (2.21a), the resulting scheme is extremely unstable.

We repeat, therefore, that the scheme (2.16) is constructed in order to preserve variation. This immediately takes us far away from the concept of viscosity solutions for parabolic equations.

In a sense our difference operator is too compressive. Isolated extrema are unchanged. The procedure, for example, does not remove “salt and pepper noise effects.” In succeeding papers, we shall apply related ideas to the development of noise removing shock filters.

It is easy to see that discrete piecewise constants are left invariant by (2.10). So are piecewise linear, continuous profiles with extrema coinciding with the “kinks,” i.e., jumps in derivative. In order to accept general piecewise linear functions we may modify F_i^n .

Let

$$(2.22) \quad \tilde{F}_i^n = F \left(m \left[\frac{\Delta_- \Delta_- u_i^n}{h^2}, \frac{\Delta_- \Delta_+ u_i^n}{h^2}, \frac{\Delta_+ \Delta_+ u_i^n}{h^2} \right] \right),$$

where minmod of n numbers is defined by induction:

$$(2.23) \quad m[(x^1, x^2, \dots, x^n)] = m[(x^1, \dots, x^{n-1}), x^n].$$

Our modified scheme is:

$$(2.24) \quad u_i^{n+1} = u_i^n - \frac{\Delta t}{h} |m[\Delta_- u_i^n, \Delta_+ u_i^n]| \tilde{F}_i^n.$$

2.1. Results of one-dimensional implementation of the shock filter. We now present the computed time evolution of one-dimensional wave forms acted upon by the shock filter described by the scheme (2.10a). In the following plots the edge switch F in (2.10a) is taken to be:

$$(2.25) \quad F = \text{sign}(\Delta_+ \Delta_- u^n).$$

Plots 1(a)–1(e) demonstrate the enhancement procedure applied to the function of 1(b) which is a slightly diffused version of the sinusoid in Plot 1(a). The evolution procedure, with stationary inflection points, which correspond here to zeros of the function, eventually produces a steady state 1(e) which is a “square wave.” Observe that the transformation is a continuous process, for otherwise, by knowing the TVP nature of (2.1a) we could have done the same just by finding inflection points and extrema and then performing thresholding. However, such an algorithm would result in a futile exercise if we tried to extend it to a two-dimensional calculation. Furthermore, (2.1a) will be generalized in a subsequent paper to include reconstruction of higher derivatives, i.e., to allow growth of the extrema (losing the TVP property); hence no thresholding could give the desired enhancement (which could be enhancements of higher derivatives).

The sequence Plot 2(a)–2(f) demonstrates that the edge development process is not necessarily producing smooth wave forms culminating in a piecewise constant function, but also that breaks in derivatives (“kinks”) do develop (see 2(b)).

Again the significance of the one-dimensional filter is not just necessarily the steady state result, but the continuous process it produces. Incidentally, we intend to experiment with this class of filters on one-dimensional speech-wave forms to determine if any perceptible speech enhancement results. In this case the transitory solution is of interest.

3. Two-dimensional enhancement. We now consider the equation

$$(3.1a) \quad u_t = -\sqrt{u_x^2 + u_y^2} F(\mathcal{L}(u))$$

to be solved for all x, y , and for $t \geq 0$, with initial data

$$(3.1b) \quad u(x, y, 0) = u_0(x, y).$$

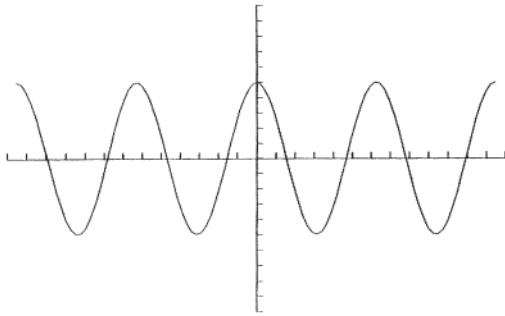
Here $F(u)$ satisfies (2.1c). Also $\mathcal{L}(u)$ is a second-order, (generally) nonlinear elliptic operator.

The general idea for the construction of the multidimensional shock filter should be evident by now.

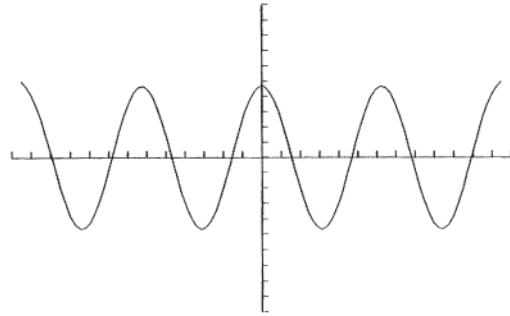
In both of the equations (2.1a) and (3.1a) we have nonlinear combination of the propagation term: $|\nabla u|$ -magnitude of the gradient and an edge-detection term $F(\mathcal{L}(u))$ whose desired behavior involves changing sign across any essential singular feature so that the local flow field is directed towards the features. Thus the edge formation and sharpening process will occur at the places where

$$(3.2a) \quad \mathcal{L}(u) = 0.$$

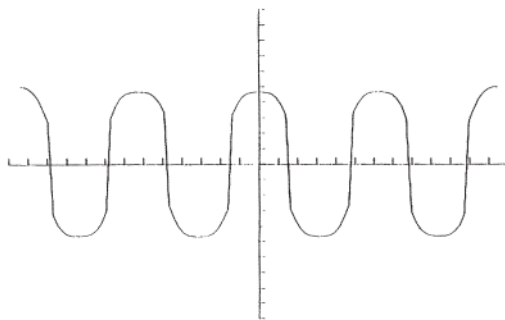
Thus the choice of $\mathcal{L}(u)$ is governed by how faithfully the zero crossings of this differential operator define edges of the processed image.



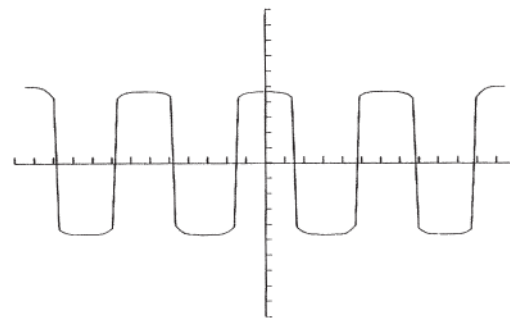
PLOT 1(a)



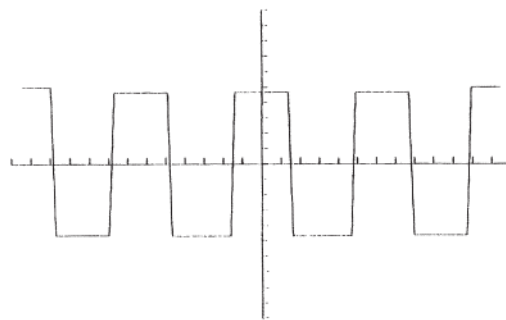
PLOT 1(b)



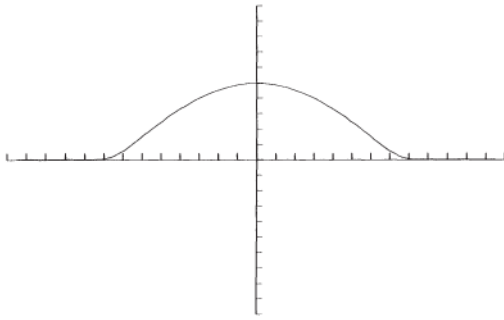
PLOT 1(c)



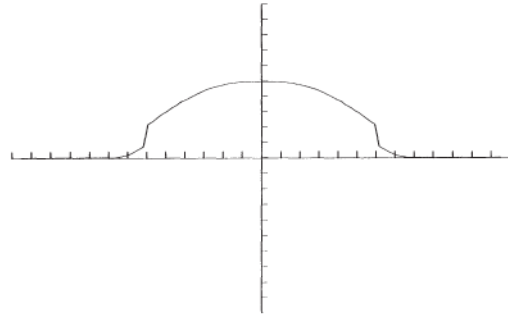
PLOT 1(d)



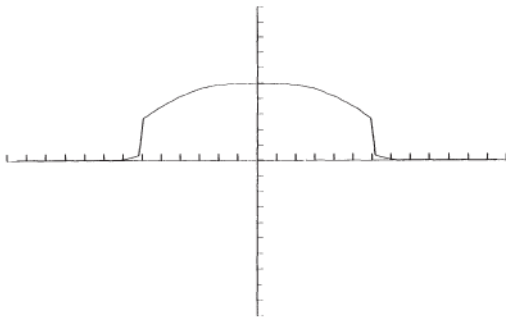
PLOT 1(e)



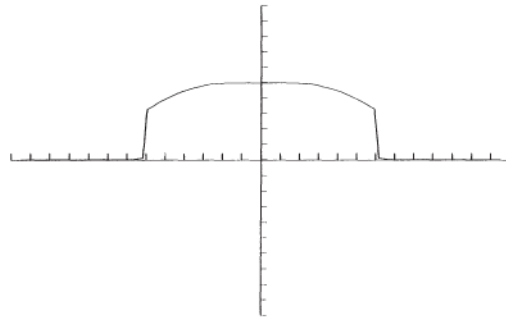
PLOT 2(a)



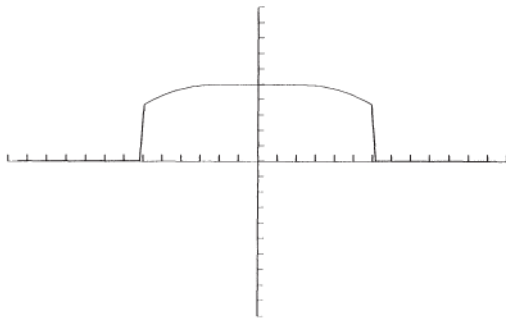
PLOT 2(b)



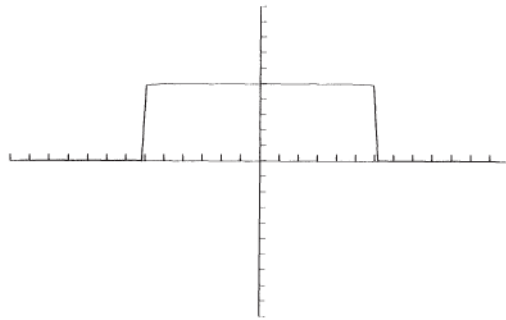
PLOT 2(c)



PLOT 2(d)



PLOT 2(e)



PLOT 2(f)

The scheme of D. Marr (well known in computer vision literature) works by finding zeros of

$$(3.2b) \quad \mathcal{L}(u) = \nabla^2(u(x, y) \star G(x, y)),$$

where $G(x, y)$ is a two-dimensional Gaussian. Since the function we are processing has already been blurred, i.e., convolved with a Gaussian, the choice for an edge operator would be

$$(3.2c) \quad \mathcal{L}(u) = u_{xx} + u_{yy} \quad (\text{the Laplacian}).$$

The reader is referred to [15] for the rigorous local analysis of this and a host of other edge-detectors. In [15] an analytical tool (called the numerical analysis of singularities) is developed in order to evaluate the behavior of “feature detectors” in the vicinity of singularities. There it is shown that the detector (3.2c) will entirely miss any geometrically nontrivial singular boundaries. Since the generalized Laplacian does not contain any curvature-dependent term, it is curvature insensitive. A somewhat better version of (3.2c) is a scheme in which edges are extracted from the zero crossings of second directional derivatives [4]. We shall skip the polynomial approximation of [4] and simply set

$$(3.2d) \quad \mathcal{L}(u) = u_{xx} \cdot u_x^2 + 2 \cdot u_{xy} u_x u_y + u_{yy} \cdot u_y^2,$$

which is simply an expression for the second derivative of u in the direction of the gradient.

We finally normalize $F(u)$ by

$$(3.2e) \quad \tilde{F}(u) = \frac{F(u)}{1 + |F(u)|}.$$

It should be noted here that the enhancement procedure is only as good as the quality of the feature detector it utilizes. The choice of (3.2d) is governed by a compromise between the quality of filtering and the computational complexity limitations, for the shock filter needs to recompute the “edge-switch” in the beginning of each iteration.

We believe that a superior quality of enhancement should result from the kind of local feature detectors proposed in [15]. A fast (perhaps wavelet transform based) “edge” transform is needed.

The estimate analogous to that in (2.9) is unfortunately lacking here. We rederive a result used in [3] here.

Let ∇ and $\nabla \cdot$ denote gradient and divergence, respectively. Let $|(x_1, x_2)| = \sqrt{x_1^2 + x_2^2}$.

We wish to compute

$$(3.3) \quad \begin{aligned} \frac{\partial}{\partial t} |\nabla u| &= \frac{\nabla u}{|\nabla u|} \cdot (u_{xt}, u_{yt}) = \left(\frac{u_x}{|\nabla u|} \frac{\partial}{\partial x} + \frac{u_y}{|\nabla u|} \frac{\partial}{\partial y} \right) (-|\nabla u| F(\mathcal{L}(u))) \\ &= -\nabla \cdot (F(\mathcal{L}(u)) \nabla u) + F(\mathcal{L}(u)) |\nabla u| \left(\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \right) \\ &= -\nabla \cdot (F(\mathcal{L}(u)) \nabla u) - |\nabla u| \mathcal{K}(u) F(\mathcal{L}(u)). \end{aligned}$$

Here $\mathcal{H}(u)$ is the curvature of level sets, $u = \text{constant}$

$$(3.4) \quad \mathcal{H}(u) = -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) = - \left(\frac{u_{xx}u_y^2 - 2u_{xy}u_xu_y + u_{yy}u_x^2}{|\nabla u|^3} \right).$$

If we integrate both sides of (3.3) over R^2 , we arrive at:

$$(3.5) \quad \frac{\partial}{\partial t} \int_{R^2} |\nabla u| = \frac{\partial}{\partial t} \text{TV}|u| = - \int_{R^2} |\nabla u| \mathcal{H}(u) F(\mathcal{L}(u)).$$

For this evolution procedure to be TV bounded at any positive time, it suffices that $-\mathcal{H}(u)F(\mathcal{L}(u))$ be bounded above. Unfortunately, such an estimate is not generally true.

Parenthetically, we note that if $\mathcal{L}(u) = \mathcal{H}(u)$ (which makes (3.1) anticompressive), then we are computing level surfaces of a curve being deformed under its mean curvature (see [3], [12]). We have rederived here in (3.5) a dissipative estimate—decay of variation in this case.

We will approximate (3.1) by setting up a grid

$$\begin{aligned} x_i &= ih_1, & y_j &= jh_2, & t^n &= n\Delta t, \\ i, j &= 0, \pm 1, \pm 2, \dots, & n &= 0, 1, 2, \dots \end{aligned}$$

For simplicity of exposition, we will take $h_1 = h_2 = h$. The approximate solution is to satisfy:

$$u_{ij}^n \approx u(ih, jh, n\Delta t).$$

Our first approximation is

$$(3.6a) \quad u_{ij}^{n+1} = u_{ij}^n - \frac{\Delta t}{h} \sqrt{(m(\Delta_+^x u_{ij}^n, \Delta_-^x u_{ij}^n))^2 + (m(\Delta_+^y u_{ij}^n, \Delta_-^y u_{ij}^n))^2} F_{ij}(\mathcal{L}(u^n)).$$

Here we have defined $\Delta_\pm^x, \Delta_\pm^y$ to be the forward and backward difference operators in the x and y directions. Also $F_{ij}(\mathcal{L}(u)) = F(\mathcal{L}_{ij}(u))$, where \mathcal{L}_{ij} is a discretization of the second-order edge detector, using central differencing for u_{xx} and u_{yy} , i.e., $u_{xx} \approx (1/h^2)\Delta_+^x \Delta_-^x u_{ij}$, etc. Symmetric differencing for $u_{xy} \approx (1/2h^2)(\Delta_-^x \Delta_-^y + \Delta_+^x \Delta_+^y u)$ and the u_x, u_y terms (if needed) are approximated using the minmod operator.

The CFL restriction is

$$(3.6b) \quad \sup \frac{\Delta t}{h} F_{ij}(\mathcal{L}(u^n)) \leq \frac{1}{4}.$$

A slight modification of this scheme comes from

$$(3.7) \quad \begin{aligned} u_{ij}^{n+1} &= u_{ij}^n - \frac{\Delta t}{h} \sqrt{((\Delta_+^x u_{ij}^n)^+)^2 + ((\Delta_-^x u_{ij}^n)^-)^2 + ((\Delta_+^y u_{ij}^n)^+)^2 + ((\Delta_-^y u_{ij}^n)^-)^2} F_{ij}^-(\mathcal{L}(u^n)) \\ &\quad - \frac{\Delta t}{h} \sqrt{((\Delta_+^x u_{ij}^n)^-)^2 + ((\Delta_-^x u_{ij}^n)^+)^2 + ((\Delta_+^y u_{ij}^n)^-)^2 + ((\Delta_-^y u_{ij}^n)^+)^2} F_{ij}^+(\mathcal{L}(u^n)) \end{aligned}$$

with the same CFL restriction (3.6b).

Our main rigorous theoretical results concern (3.7).

(A) For (3.7): Let

$$M_{ij}^n = \max \{u_{ij}^n, u_{i-1,j}^n, u_{i+1,j}^n, u_{i,j-1}^n, u_{i,j+1}^n\},$$

$$m_{ij}^n = \min \{u_{ij}^n, u_{i-1,j}^n, u_{i+1,j}^n, u_{i,j-1}^n, u_{i,j+1}^n\}.$$

Then, since for fixed constant values of $F_{ij} \leq 0$ and $F_{ij}^+ \geq 0$ the scheme is monotone, we have a local maximum principle (see (2.18), (2.19))

$$(3.8) \quad m_{ij}^n \leq u_{ij}^{n+1} \leq M_{ij}^n.$$

This means that for fixed $\Delta t, h$, as $n \rightarrow \infty$ there is a convergent subsequence whose limit is $\{u^\infty\}_h$, the discrete processed image. (Again uniqueness of limits is unknown.)

In addition, if $\mathcal{L}(u^n)$ is such that

$$(3.9a) \quad \mathcal{L}(u^n) \leq 0 \text{ at local discrete maxima,}$$

$$(3.9b) \quad \mathcal{L}(u^n) \geq 0 \text{ at local discrete minima,}$$

then these extrema are invariant in time, i.e., the values of u_{i_0, j_0}^n at local extrema are the same as the values of u_{i_0, j_0}^{n+1} .

We also have:

(B) For (3.6a): If u_{i_0, j_0}^n is a local maximum

$$u_{i_0, j_0}^n \geq \max (u_{i_0-1, j_0}^n, u_{i_0+1, j_0}^n, u_{i_0, j_0-1}^n, u_{i_0, j_0+1}^n),$$

then it is not increased—in fact

$$u_{i_0, j_0}^{n+1} = u_{i_0, j_0}^n.$$

The same is true for local minima.

We also consider an option which replaces $F_{ij}(\mathcal{L}(u^n))$ by $\tilde{F}_{ij}(\mathcal{L}(u^n))$

$$\tilde{F}_{ij} = F(m[\mathcal{L}_{i+\nu, j+\mu}(u^n)]),$$

$$\nu, \mu = -1, 0, 1.$$

Thus the processing will leave invariant piecewise continuous approximate solutions of

$$\mathcal{L}(u) = 0.$$

The algorithms described above have been incorporated [20] into U.S. patent application serial number 456,120.

4. Results of two-dimensional implementation of the shock filter. We demonstrate the shock-filtering enhancement scheme (3.7) on various standard images from the USC IPI Image Data Base. In this experiment we first blur a picture with number N_d of iterations of a standard approximation of the diffusion equation (4.1)

$$(4.1) \quad u_t = \Delta u$$

with $\Delta t/h^2 = .25$, the largest possible CFL number. Then we show a few iterations of the shock filter N_s with the final one at steady state. The CFL number equals .25 for the shock filter, the maximal possible under the stability restriction. There are 256 gray levels for each picture element (pixel). Hence the quantization error is $\sim 4 \times 10^{-3}$. If a simple reversal of (4.1) is attempted, the image blows up geometrically in just a few iterations (see e.g., [7]).

A black and white (B & W) (256×256) picture, Picture 1(a) was chosen for its visual simplicity and for the clear dependence of its information content on the presence of singularities. (It is easy to see the difficulties involved in writing a program to count the candies just from the image P1(b).) Picture 1(b) is the result of blurring by $N_d = 8$ iterations of the diffusion equation.

Pictures 1(c), (d) correspond to

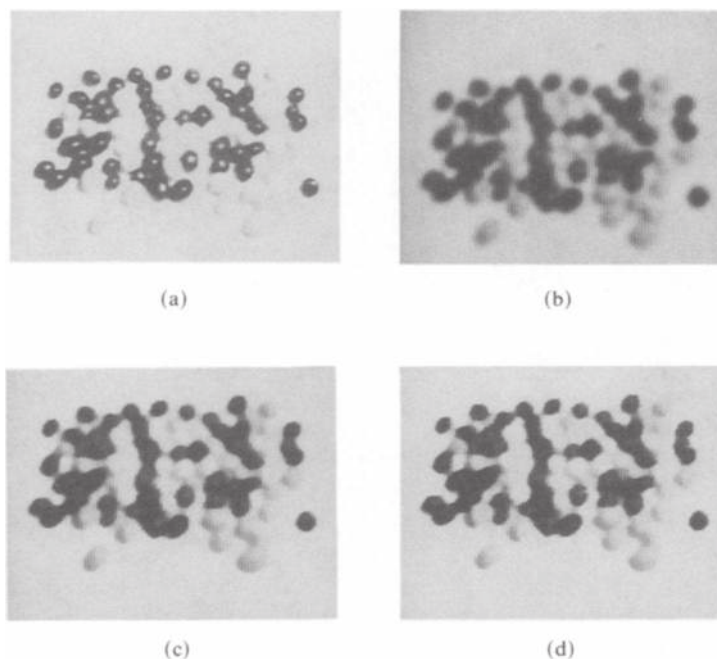
$$N_s = 9, 18.$$

An excellent, nonoscillatory piecewise linear reconstruction is evident.

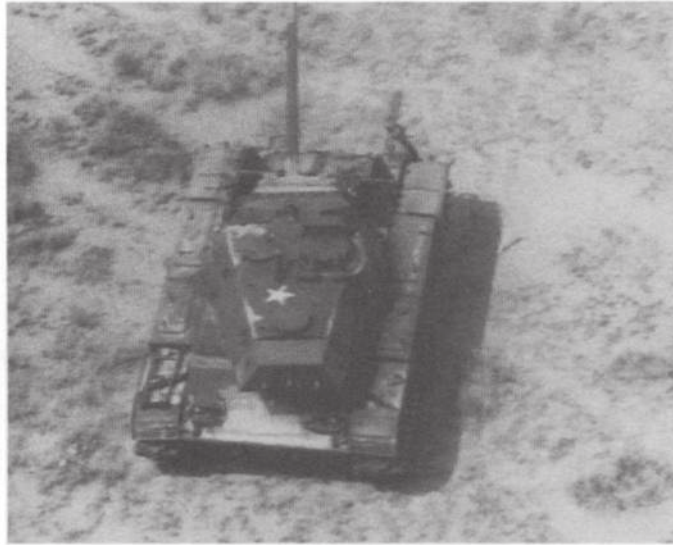
Pictures 2(a)–(c) are the enhancement procedures on a (512×512) B & W “Tank” image: $N_d = 8$, $N_s = 13$, i.e., Picture 2(b) is the result of eight blurring iterations, and Picture 2(c) is the enhanced tank after 13 enhancement steps.

The black and white 256×256 “clock” image (Pictures 3(a), (b)) is chosen because it has a great deal of small details, including easily perturbed numerals on the clock’s face. In this experiment the enhancement of the original image (which, we remind the reader, had 256 gray levels (as did all our examples 1–4)) is performed, i.e., $N_d = 0$, $N_s = 5$. The procedure seems to resolve the image beyond its original fidelity, revealing, for instance, small details on the background photograph. The appearance of “jagged” edges simply means that edges get too compressed for the initial resolution.

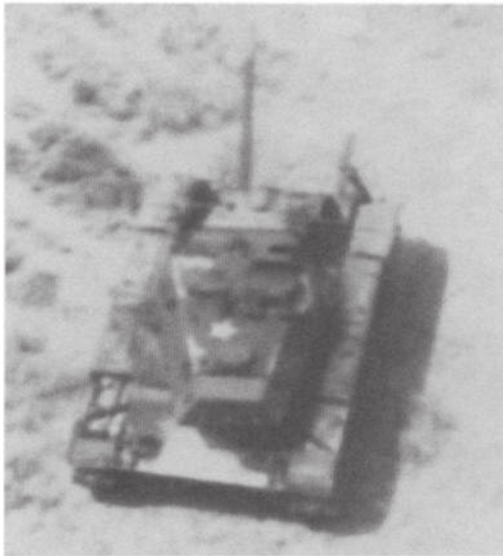
The black and white (256×256) Pictures 4(a)–(c) with $N_d = 0$, $N_s = 5, 11$ suggest resolution beyond the original optical limit. In particular a small “gate” appears in the white building in the background which was not initially apparent there. Again, it is possible to overenhance part of the image using our methods—the roadside becomes more jagged as N_s increases.



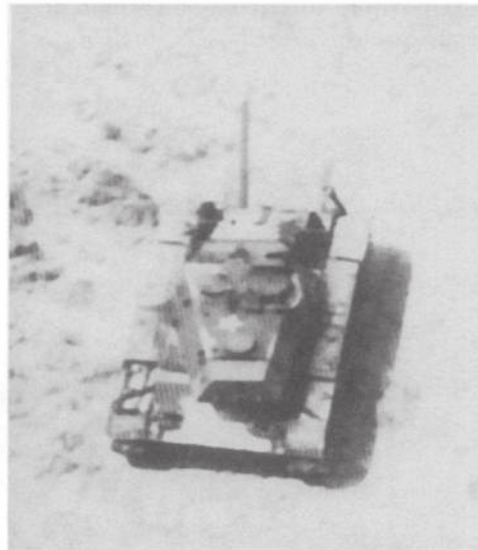
PICTURES 1a-d



(a)

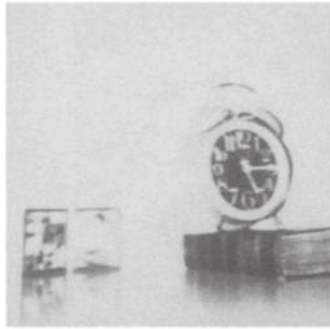


(b)

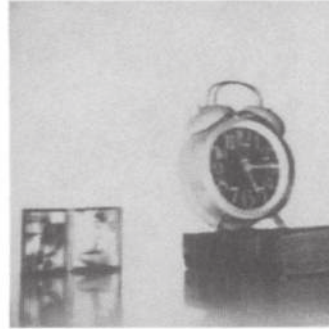


(c)

PICTURES 2a-c



(a)



(b)

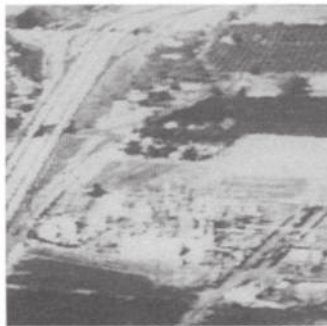
PICTURES 3a, b



(a)



(b)



(c)

PICTURES 4a-c



(a)



(b)

PICTURES 5a, b



(c)

PICTURE 5c

Finally, Pictures 5(a)–(c) demonstrate color enhancement on a (512×512) 24 B/pixel image of a lake. Here $N_d = 15$ and $N_s = 5$. Each separate color plane was shock filtered. This is less clear in the black-and-white reproductions shown here.

An interesting observation here is that Picture 5(c) appears to be an “impressionistic” version of the original. This “painting-like” quality comes from the fact that the shock filter (3.7) does not restore details beyond the scales lost in the diffusion process (i.e., features whose edges are not detected by the $\tilde{F}(u)$), yet all the higher scales get perfectly enhanced. Thus the result looks like a painting—not a blur!

Acknowledgment. The authors thank Professors Theo Pavlidis and Chris Anderson for helpful discussions about this subject.

REFERENCES

- [1] S. R. CHAKRAVARTHY AND S. OSHER, *Computing with high-resolution upwind schemes for hyperbolic equations*, in Lectures in Appl. Math., 22, B. E. Engquist, S. Osher, and R. C. J. Somerville, eds., American Mathematical Society, Providence, RI (1985), pp. 57–86.
- [2] S. K. GODUNOV, *A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics*, Mat. Sb., 047 (1959), pp. 271–290.
- [3] M. GRAYSON, *A short note on the evolution of surfaces via mean curvature*, Stanford University Math. Dept., preprint, 1987.
- [4] R. HARALICH, *Digital step edges from zero crossings of second directional derivatives*, IEEE Trans. Pattern Analysis and Machine Intelligence, 1, PAMI-6 (1984) pp. 58–68.
- [5] A. HARTEN, B. ENGQUIST, S. OSHER, AND S. R. CHAKRAVARTHY, *Uniformly high order accurate essentially non-oscillatory schemes, III*, J. Comput. Phys., 71 (1987), pp. 231–303.
- [6] A. HARTEN, J. M. HYMAN, AND P. D. LAX, *On finite difference approximations and entropy conditions for shocks*, Comm. Pure. Appl. Math., 29 (1976), pp. 297–322.

- [7] H.-O. KREISS AND J. OLIGER, *Methods for the approximation solution of time dependent problems*, GARP Publication Series #10, 1973.
- [8] P. D. LAX AND B. WENDROFF, *Systems of conservation laws*, *Comm. Pure Appl. Math.*, 12 (1960), pp. 217-237.
- [9] A. MAJDA, J. McDONOUGH, AND S. OSHER, *The Fourier method for nonsmooth initial data*, *Math. Comp.*, 22 (1978), pp. 1041-1081.
- [10] D. MARR AND E. HILDRETH, *Theory of edge detection*, *Proc. Roy. Soc. London Ser. B*, 207 (1980), pp. 187-217.
- [11] S. OSHER, *Riemann solvers, the entropy condition and difference approximations*, *SIAM J. Numer. Anal.*, 27 (1984), pp. 217-235.
- [12] S. OSHER AND J. A. SETHIAN, *Fronts propagating with curvature dependent speed. Algorithms based on a Hamilton-Jacobi formulation*, *J. Comput. Phys.*, 79 (1988), pp. 12-49.
- [13] S. OSHER AND P. K. SWEBY, *Recent developments in the numerical solution of nonlinear conservation laws*, in *The State of the Art in Numerical Analysis*, A. Iserles and M. J. D. Powell, eds., Clarendon Press, Oxford (1987), pp. 681-701.
- [14] L. RUDIN, *Shock filters*, DARPA Report, Rockwell International Science Center (1984).
- [15] ———, *Images, numerical analysis of singularities, and shock filters*, Ph.D. thesis, Computer Science Dept., Caltech, Pasadena, CA, Tech. Report 5250:TR:87, 1987.
- [16] C. W. SHU AND S. OSHER, *Efficient implementation of essentially non-oscillatory shock capturing schemes*, *J. Comput. Phys.*, 77 (1988), pp. 439-471.
- [17] B. VAN LEER, *Toward the ultimate finite difference scheme, II. Monotonicity and conservation combined in a second order scheme*, *J. Comput. Phys.*, 14 (1974), pp. 361-370.
- [18] G. H. GRANLUND, H. KNUTSSON, AND R. WILSON, *Image enhancement*, in *Computer Vision*, North-Holland, Amsterdam, 1983, pp. 56-67.
- [19] P. L. LIONS AND P. E. SOUGANIDES, *Viscosity solutions of second-order equations, stochastic control, and stochastic differential games*, in *Stochastic Differential Systems, Stochastic Control Theory, and Applications*, W. H. Fleming and P. L. Lions, eds., IMA Vol. Math. Appl., 10, Springer-Verlag, Berlin, New York, 1988.
- [20] S. OSHER AND L. RUDIN, *A system incorporating feature oriented signal enhancement using shock filters*, U.S. Patent Application serial no. 456,120 (filed December 22, 1989), U.S. Patent and Trademark Office.