Sādhanā, Vol. 22, Part 4. August 1997, pp. 553–574. © Printed in India.

A variational formulation-based edge focussing algorithm

T J RICHARDSON[1] and S K MITTER[2]

[1] AT & T Bell Laboratories, 600-700 Mountain Av., Murray Hill, NJ 07974, USA
[2] Laboratory for Informations and Decision Systems, Massachussetts Institute of Technology, Cambridge, MA 02139, USA
e-mail: mitter@lids.mit.edu; tjr@bell-labs.com

Abstract. Many edge detection techniques exhibit scale-dependent distortion of edges. We develop two ideas, which may also be of independent interest, to produce sharp edge localization at all scales. The first is approximation of the functional associated with the variational formulation via epi-convergence, replacing the edge set with a function. We provide a fast algorithm for minimizing the approximate functional. The second is to scale parameters and data to focus the edges. The resulting edge detector is a singular perturbation of a coupled pair of partial differential equations, yielding an elegant structure, suitable for digital or analog parallel implementation on mesh-connected arrays.

Keywords. Edge detection; parallel implementation; mesh-connected arrays.

1. Introduction

The notion of 'scale', scale of features and scale of representation, is widely held to be of fundamental importance in vision. One reason for this is that hierarchical descriptions offer potential reductions in complexity of various visual processing problems. Coarse scale segmentation of an image, for example, can be used to identify regions of interest for further processing, thereby reducing the computational load. It is important, therefore, that coarse scale descriptions retain those features of the data that are required for effective decision making. In the case of edge detection, T-junctions and corners play important roles in estimating the depth and shape of objects in a scene (Gamble & Poggio 1987). It is desirable, therefore, to accurately represent these features even at coarse scales.

Coarse scale edges have the advantage of reduced complexity; fine scale edges are more convoluted but yield more accurate and detailed information. A problem many edge-detection algorithms exhibit is systematic distortion of high curvature edges, such as T-junctions and corners; this distortion is aggravated by operating on coarse scales. This is especially true of algorithms which smooth the data in a scale-dependent way prior to locating edges in the smoothed image. The Marr–Hildreth edge detector (Marr

& Hildreth 1980) and the Canny (1986) edge detector are examples of this type. More recent approaches to edge detection combat the distortion of high curvature edges by introducing interaction between image smoothing and edge placement. Examples are the Markov random field formulations (Geman & Geman 1984; Marroquin 1985; Bilbro *et al* 1992) and the related variational formulation of edge detection (Mumford & Shah 1985, 1989; Blake & Zisserman 1987). There is evidence (Blake & Zisserman 1987) that edge-detection techniques of this type do exhibit smaller localisation distortion than those of the first class. Problems remain, however; distortions still occur, and, as before, the degree of distortion depends on the scale of the edge detector. Coarse scale edges exhibit greater distortion.

Edge focussing aims at improving the localization of coarse scale edges without introducing fine scale edges. In this paper, we have derived an algorithm for performing edge focussing by starting with a variational formulation of an edge-detection problem. The resulting algorithm is described by a coupled set of nonlinear second order parabolic partial differential equations ((5)–(9) below) with explicit parameters β and c which are appropriate-adjusted (see (10)–(12)). The adjustment induces focussing of the edges. The global coarse scale nature of the edges is retained by introducing scale-stabilizing feedback mechanisms. The adjustment process commences after the nonlinear parabolic equations have nearly converged to their equilibrium. The set of equations (5)–(9) and (10)–(12) should really be viewed as an adaptive nonlinear filter which performs edge detection via focussing. Indeed, the equations are the fundamental objects in this theory and, apparently, are far more well behaved (for example, convergence to global minima) than the original variational problem.

The foundation and motivation of the adjustment of the parameters lies in certain limit theorems for the variational formulation proved by one of us (Richardson 1990, 1992). These theorems are discussed in § 3. An outline of the variational formulation appears in § 2. The algorithm is developed in a continuum setting in § 4, and refined and discretized in § 5. Simulation results can be found in § 6.

The work presented in this paper is similar in spirit to that of Bergholm (1987) who focussed edges produced by the Marr–Hildreth edge detector. Since the variational formulation has better localization than the Marr–Hildreth edge detector, our approach requires less drastic adjustment of the edges.

The edge-focussing algorithm is not the only point of interest in this paper. The variational formulation is a mathematical model, not an algorithm. The primary difficulty in constructing an algorithm is appropriately representing the edges. One approach is to absorb the edges into the interaction between neighbouring pixels. This idea appears in the anisotropic diffusion approach (Nordström 1990; Perona & Malik 1990), GNC (Graduated non-convexity) type algorithms (Blake & Zisserman 1987), and in mean field annealing (Geiger & Yuille 1989; Bilbro *et al* 1992). There is another approach, which we adopt in this paper, that has been developed within the framework of approximation (of variational principles) via Γ-convergence or epi-convergence (De Giorgi & Franzoni 1979; Attouch 1984). This theory has been successfully applied to the variational formulation of edge detection by Ambrosio & Tortorelli (1990, 1992). Some computational work based on these results has appeared (March 1988, 1989, 1992). These approximations, and some further variations, are presented in § 2. (The variations we indicate allow great flexibility

in the approximation, which may be tailored to particular needs or taken as an indication of the robustness of the approach.)

The approximation is achieved by reformulating the variational problem. The edge set is replaced with a function that modulates the smoothing of the image. The variational principle forces the function to have the appearance of a smoothed neighbourhood of the corresponding edge set. The degree of smoothing, i.e. the width of the effective edges, is controlled by a parameter. The 'convergence' of the approximation occurs by taking the parameter to the appropriate limit where the width of the effective edges tends to zero. One significant advantage of this reformulation is that the 'edge' function can be discretized in a straightforward way and minimization can proceed via the solution of discretized partial differential equations. In particular, the parabolic partial differential equations (5)–(9) comprise a 'gradient' descent on the approximate functional. Actually, in our implementation, we vary the step size of the 'gradient' descent and approximate a Newton-type descent (see (16)). By doing this we achieve very fast rates of convergence. Thus, we obtain a fast and elegant algorithm.

Having a separate function to represent the edges, as opposed to absorbing them into the interactions between image pixels, has some advantages. Machine Vision researchers are interested in combining various low-level vision processes, intensity edge detection and stereo depth edge detection for example, into a single operation. Having an edge representation such as the one we employ may greatly facilitate this.

It turns out that the form and the properties of the epi-convergent approximation mesh well with the parameter adjustment proposed for the edge-focussing algorithm. In particular, one can argue heuristically, and demonstrate computationally, that with 'wide' edges some edge distortions are relaxed. The price paid for this is a drop in resolution of the edges. The adjustment attempts to produce the best of both worlds, relaxing the edges initially and then sharpening them as the parameters associated with the variational formulation are adjusted for finer scales.

2. The variational formulation

Mumford & Shah (1985, 1989) suggested performing edge detection by minimizing functionals of the form

$$E(f, \Gamma) = \beta \int_{\Omega} (f - g)^2 \mathrm{d}\mu + \int_{\Omega - \Gamma} |\nabla f|^2 d\mu + \alpha |\Gamma|,$$

where Ω is the image domain (a rectangle), μ is Lebesgue measure, g is the observed grey level image, i.e. a real valued function, Γ denotes the set of edges, $|\Gamma|$ is the length of Γ, and β and α are real positive scalars. This approach is a modification of one due to Geman & Geman (1984) using Markov random fields, which was developed by Marroquin (1985) and by Blake & Zisserman (1987).

The three terms of E 'compete' to determine the set Γ and the function f. The first term penalizes infidelity of f to the data, while the second term forces smoothness of the approximation f, except on the edge set Γ. Thus, f is a piecewise smooth approximation to g. The third term forces some conservativeness in the use of edges by penalizing their total length. Roughly speaking, if g has a step discontinuity, approximating g with a smooth f

causes the first term to be large, tracking g more closely with a less smooth f increases the second term, and allowing Γ to approximate the support of the discontinuity reduces both the first and the second terms at some cost to the third.

This formulation was motivated in part by the desire to combine the processes of edge placement and image smoothing. Earlier edge-detection techniques such as the Marr–Hildreth edge detector, the Canny edge detector, and their variants separated these processes; the image is first smoothed, to suppress noise and control the scale, and edges are detected subsequently, as gradient maxima, for example. A consequence of this two-step approach is pronounced distortion of the edges, especially at high curvature locations. Corners tend to retract and be smoothed out; the connectedness of the edges at T-junctions is lost. The 'finger-print' images of gradient maxima of one dimensional images in scale space (Witkin 1983) are well known; the localization of edges degrades badly as scale increases. Many two-dimensional examples can be found in the literature. By introducing interaction between the edge placement and the smoothing it was expected that this effect could be abated. There is evidence, both theoretical, in one dimension (see, Blake & Zisserman 1987), and experimental, in two dimensions, that this is indeed the case.

2.1 *Approximation and computation*

To compute minimizers of E the critical question is how to represent the set Γ. A natural approach is to discretize Γ into "edge elements" and treat them combinatorially, adding or removing elements in an attempt to minimize E. Appending a stochastic component leads to the simulated annealing approach first suggested by Geman & Geman (1984). This tends to produce computationally impractical algorithms. Modifications which incorporate the edge elements into the interaction between image pixels have been proposed. One of these is based on mean field approximations of the Markov random field (Geiger & Girosi 1989; Bilbro *et al* 1992) and another, GNC (Blake & Zisserman 1987), is based on a homotopy of the interactions. Both these approaches have their strong points, and are in fact quite similar (Geiger & Yuille 1989; Bilbro *et al* 1992). A novel approach has appeared from the mathematical theory of approximation of functionals via Γ-convergence, also known as epi-convergence (De Georgi & Franzoni 1979; Attouch 1984). We will use the later terminology to avoid confusion. We will not give a general definition of epi-convergence and refer interested readers to the references cited and also to Ambrosio & Tortorelli (1990, 1992). The definition of epi-convergence is designed to allow approximation of one variational principle by another. We consider functionals of the form

$$E_c = \int_\Omega [\beta(f-g)^2 + \Phi(v)|\nabla f|^2 + \alpha(c\Psi(v)|\nabla v|^2 + (1-v)^2/4c)]\mathrm{d}\mu. \tag{1}$$

Here $\Phi(v)$ takes the role of the Γ in E, i.e., it modulates the smoothness constraint on f. The other terms involving v force $\Phi(v)$ to simulate the effect that Γ has in E. Implicitly we have $0 \leq v \leq 1$. The algorithmic intention is to minimize E_c with respect to f and v. An obvious advantage the approximation offers over the original formulation is that v, since it is a function on Ω, can be discretized in a straightforward way and (local) minimizers of

E_c can be computed by gradient descent. Ambrosio & Tortorelli (1992) recently showed that if one sets

$$\Phi(v) = v^2, \quad \text{and} \quad \Psi(v) = 1, \tag{2}$$

then E_c epi-converges to E as $c \to 0$. Some computational results for this functional have already appeared (March 1989) (see also Shah 1991), and a scheme similar to the one presented here has been applied to the stereo-matching problem by March (1988).

The choice for Φ and Ψ given above may be one of the simplest possible but it is not unique. The first functional which was shown to be epi-convergent to E (by Ambrosio & Tortorelli 1990) has the form of (1) with the formal substitutions $\Phi(v) = c\Psi(v) = (1-(1-v)^2)^{2c^{-(1/2)}}$. When one considers algorithms based on these functionals there are trade-offs to be made between speed and performance. For example, the choice reflected in (2) leads to simple equations and fast computation. However, the more complicated choice mentioned above produces sharper singularities in Φ and hence less smearing of f near the edges. With slight modifications, the proof of epi-convergence found by Ambrosio & Tortorelli (1990, 1992) can be made to go through for a large class of Ψ and Φ. In particular, one can choose Ψ to be any C^1 function satisfying

$$\Psi(x) > 0, \quad \text{for } x \in (0, 1],$$
$$2\int_0^1 (1-u)\Psi^{1/2}(u)\mathrm{d}u = 1.$$

Note that any function satisfying the first property can be made to satisfy the second property by suitable normalization. Given such a Ψ one can choose Φ to be any C^1 function satisfying

$$\Phi(1) = 1,$$
$$\Phi(0) = 0,$$
$$\Phi(x) \in (0, 1) \quad \text{for } x \in (0, 1).$$

Although the conditions given above are sufficient for the proof of epi-convergence, for algorithms based on 'gradient' descent one should also impose the condition that Ψ be monotonically non-decreasing and Φ be monotonically increasing on $(0, 1)$. Furthermore, for our implementation, which is discussed in § 5, the condition $\lim_{x\to 0} \dot{\Phi}(x)/x = 0$ should be imposed. Even more general Ψ and Φ than defined above are possible. For example, setting

$$\Psi(v) = \Phi(v) = \tfrac{1}{2}e^{-(1-v)^2} \tag{3}$$

also produces an epi-convergent set of functionals. Examples in the class defined above are

$$\Phi(v) = v^{2n} \quad \text{and} \quad \Psi(v) = \frac{(m+1)^2(m+2)^2}{4}v^{2m}, \tag{4}$$

where $m \geq 0$ and $n > 0$. Equation (2) is a special case with $(n, m) = (1, 0)$.

Suppose $E(f, \Gamma) < \infty$. The proof of epi-convergence involves basically two steps. The first is to show that for any sequence $\{f_{c_i}, v_{c_i}\}$ where $c_i \to 0$, $f_{c_i} \to f$,

and $v_{c_i} \to 1$ in an appropriate sense (not pointwise), that $\liminf_{i\to\infty} E_{c_i}(f_{c_i}, v_{c_i}) \geq E(f, \Gamma)$. The second is to construct a sequence such that $\limsup_{i\to\infty} E_{c_i}(f_{c_i}, v_{c_i}) \leq E(f, \Gamma)$. If (f, Γ) minimizes E then this second step requires constructing near minimizers of E_c. If $\liminf_{i\to\infty} E_{c_i}(f_{c_i}, v_{c_i}) < \infty$ then, roughly speaking, if $x \in \Gamma$ one has $\lim_{i\to\infty} \Phi(v_{c_i}(x)) = 0$. (Thus at these points we do not have $v_{c_i}(x) \to 1$, however, Γ is a set of μ measure 0.) On the other hand, the last term in (1) forces $v_c(x)$ to converge to 1 for almost all $x \in \Omega$ (in the sense of Lebesgue measure) and hence one has $\lim_{c\to 0} \Phi(v_c(x)) = 1$ for almost all $x \in \Omega$. The near minimizers of E_c are constructed by setting $\Phi(v_c(x)) \simeq 0$ on Γ and $\Phi(v_c(x)) \simeq 1$ outside some neighbourhood of Γ with a smooth transition in between. The approximations indicated here become equalities in the limit as $c \to 0$. The width of the transition depends on Ψ and on c. We give a brief heuristic description of how this occurs. In the transition region we expect f to be relatively smooth so only the terms not involving f in E_c will have a significant effect on the form of v there. In the following inequality,

$$c\Psi(v)|\nabla v|^2 + \frac{(1-v)^2}{4c} \geq \Psi^{1/2}(v)|\nabla v|(1-v),$$

the equality holds only if $|\nabla v| = \Psi^{-1/2}(v)[(1-v)/2c]$. This suggests that (in one dimension) if $u_c(t)$ satisfies $\partial u_c(t)/\partial t = [(1-u_c(t))/2c]\Psi^{-1/2}(u_c(t))$ with $\Phi(u(0)) \simeq 0$ then setting $v(x) = u_c(\text{dist}\,(x, \Gamma))$, for dist $(x, \Gamma)) \leq \tau_c$ where $\Phi(u_c(\tau_c)) \simeq 1$ (with $u_c(\tau_c) \to 1$ as $c \to 0$), will produce near optimal transitions. This is how the near optimal v_c are constructed by Ambrosio & Tortorelli (1990, 1992). Note that assuming that $u_c(0)$ does not depend on c we obtain $u_c(t) = u_1(t/c)$. Thus the edge width is proportional to c. Let $\gamma(s) = \int_0^s (1-r)\Psi^{1/2}(r)\mathrm{d}r$. We now compute

$$\begin{aligned}
&\int_0^{\tau_c} \left(c\Psi(u(t)) \left| \frac{\partial u_c(t)}{\partial t} \right|^2 + \frac{(1-u(t))^2}{4c} \right) \mathrm{d}t \\
&= \int_0^{\tau_c} \left(\Psi^{1/2}(u(t)) \left| \frac{\partial u_c(t)}{\partial t} \right| (1-u(t)) \right) \mathrm{d}t \\
&= \left| \int_0^{\tau_c} \frac{\partial}{\partial t} \gamma(u(t)) \mathrm{d}t \right| = \gamma(\tau_c) \simeq \frac{1}{2}
\end{aligned}$$

with the last approximate equality becoming equality in the limit as $c \to 0$. In the one-dimensional case we now see that the last term in (1) will contribute approximately α times the number of discontinuity points of f. In two dimensions one obtains approximately α times the length of $|\Gamma|$. Thus, we see that E_c approximates E.

Our edge-focussing algorithm will be implemented as a singular perturbation of a descent on a discrete version of E_c. We will briefly consider the continuum equivalent. Define

$$\partial_f E_c = \beta(f-g) - \nabla \cdot (\Phi(v)\nabla f), \tag{5}$$

$$\begin{aligned}
\partial_v E_c &= \dot{\Phi}(v)\alpha^{-1}|\nabla f|^2 - c\nabla \cdot (\Psi(v)\nabla v) \\
&\quad + 2c\dot{\Psi}(v)|\nabla v|^2 + (1-v)/2c, \qquad (6) \\
&= \dot{\Phi}(v)\alpha^{-1}|\nabla f|^2 - 2c\Psi(v)\Delta v) - c\dot{\Psi}(v)|\nabla v|^2 + (1-v)/2c. \qquad (7)
\end{aligned}$$

The Euler–Lagrange equations for f and v are, respectively, $\partial_v E_c = 0$ and $\partial_f E_c = 0$ with Neumann boundary conditions on both v and f. Allowing f and v to depend on t we can write a 'gradient' descent on E_c in the form,

$$\frac{\partial}{\partial t} f(x,t) = -c_f \partial_f E, \tag{8}$$

$$\frac{\partial}{\partial t} v(x,t) = -c_v \partial_f E, \tag{9}$$

where c_f and c_v control the rates of descent; they would be constant for a strict gradient descent, but may not be in general. In our implementation c_v is not constant. Since the functional E_c is not jointly convex in v and f we do not expect to always reach a global minimum by a descent method. Thus the solution obtained will depend on the initial conditions and also on the parameters c_f and c_v.

Equation (8) strongly resembles the anisotropic diffusion scheme of Perona & Malik (1990) and, even more strongly, the 'biased' anisotropic diffusion scheme of Nordström (1990). Perona & Malik (1990) begin by considering the standard diffusion equation,

$$\frac{\partial}{\partial t} f(x,t) = \Delta_x f(x,t), \quad f(x,0) = g(x),$$

which produces the 'scale-space' smoothings of g, parametrized by t. For fixed t one obtains $f(x,t)$ by convolving $g(x)$ with a Gaussian kernel whose variance is linear in t. Perona & Malik (1990) suggested controlling the diffusion coefficient to prevent smoothing across edges. Thus they were led to consider equations of the form

$$\frac{\partial}{\partial t} f(x,t) = \nabla_x \cdot (h(|\nabla_x f(x,t)|)\nabla_x f(x,t)), \quad f(x,0) = g(x).$$

They experimented with $h(s) = J/[1 + (s/K)^2]$ and $h(s) = e^{-(s/K)^2}$, where J and K are constants. Equation (8) resembles this equation in that it is a diffusion with controlled conductivity. The control of the conductivity depends on $|\nabla f|$ indirectly through (9). The term $\beta(f - g)$ in (8) stabilizes the solution at some particular scale. Such a term also appears in the 'biased' anisotropic diffusion scheme studied by Nordström (1990). Perona & Malik (1990) analyse their scheme to show that the maximum principle holds, i.e. that the solution's extrema never exceed those of the original image. They argue that this implies that no new 'features' (blobs) are introduced into the solution. Here, as in Nordström (1990), this property is a trivial consequence of the formulation. The functionals E_c would *increase* if such new features appeared. (Truncating such a new feature would decrease E_c.) An advantage of the scheme presented here is that it yields an explicit representation of the edges (via the function $\Phi(v)$). The resulting system of equations admits a particularly simple implementation in digital mesh connected parallel machines with simple processors or, potentially, in an analog network, such as discussed by Harris *et al* (1989).

3. Scale, noise, and accuracy

Witkin (1983) introduced the idea of a scale-space representation of an image – smoothing the image and identifying edges on all scales. One of the problems arising from the scale-space concept was the correspondence problem: Which of the fine scale edges correspond to

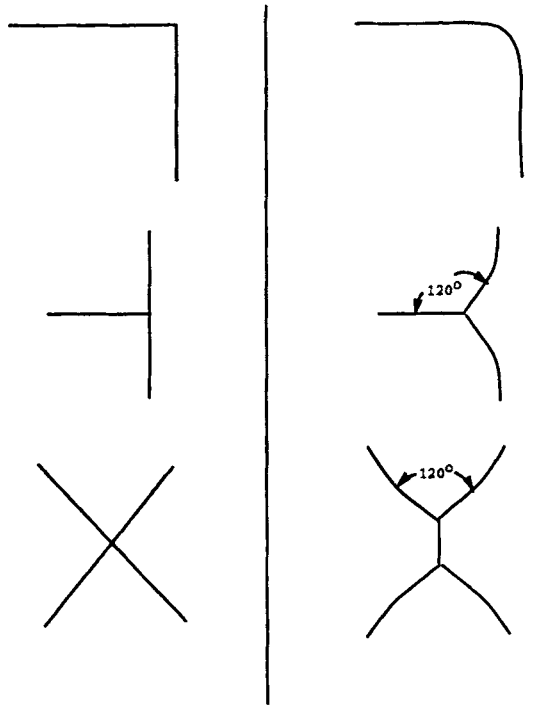

Figure 1. Calculus of variations results.

the coarse scale ones? The problem is aggravated by the fact that the distortion of the edges, mentioned earlier, depends on scale. Typically 'coarse scale' implies more smoothing and, hence, more distortion. This is undesirable in many situations because salient features, corners and T-junctions for example, tend to be obscured. The correspondence problem is therefore of great importance. The edge-focussing algorithm of Bergholm (1987) addresses the correspondence problem by taking small steps in scale. We address it this way too, in effect, but also by using a better underlying edge detection algorithm.

Since the variational approach combats the distortion caused by smoothing one hopes that the correspondence problem will be alleviated by using it. Although this appears to be the case, problems remain; there still are distortions, depending on scale, and the model intrinsically restricts the geometry of possible edge sets in an unnatural way. The analysis of Mumford & Shah (1985) showed that edge sets produced by the variational approach have the following properties, which are illustrated in figure 1.

If Γ is composed of $C^{1,1}$ arcs then

- at the most three arcs can meet at a single point and they do so at 120°,
- they meet $\partial\Omega$ only at an angle of 90°,
- it never occurs that exactly two arcs meet at a point (other than the degenerate case of two arcs meeting at 180°), i.e., there are no corners.

These results derive from the fact that the term $|\Gamma|$ in E locally dominates the behaviour of singularities in Γ. Hence the types of singularities observed are identical to those of

minimal surfaces. Of course 'real' edges are not restricted to these geometries. The dependence on scale derives from the interaction between the singularities of f and those of Γ. Roughly speaking, smaller values of β produce greater distortion. This becomes more clear in light of the theorems quoted below and the analysis presented by Mumford & Shah (1989).

In view of the foregoing it is natural to ask what are the noise/accuracy/scale tradeoffs and, more generally, whether scale-dependent distortions are necessary. Is it necessary that the mechanisms used to combat noise, i.e. smoothing, be tied simultaneously to scale and accuracy? We propose that the accuracy of edge localization need not be limited by scale. The central ideas of the method we will develop can be gleaned from some asymptotic theorems proved by one of us (Richardson 1990, 1992) concerning minimizers of E. We will now present a slightly simplified statement of those theorems.

To measure the disparity between one edge set and another we introduce the Hausdorff metric. For $A \subset \mathbb{R}^2$ the ϵ-neighbourhood of A will be denoted by $[A]_\epsilon$ and is defined by $[A]_\epsilon = \{x \in \mathbb{R}^2 : \inf_{y \in A} ||x - y|| < \epsilon\}$ where $\| \cdot \|$ denotes the Euclidean norm. Denoted by $d_H(\cdot, \cdot)$, the Hausdorff metric is evaluated by

$$d_H(A, B) = \inf\{\epsilon : A \subset [B]_\epsilon \text{ and } B \subset [A]_\epsilon\}.$$

Elementary analysis shows that d_H is a metric on the space of non-empty compact sets in R^2.

Suppose for the moment that we have ideal data: g is a piecewise smooth function. To make this clear we denote it by g_I. We assume that there exists a set Γ_g, a union of curves, satisfying length $(\Gamma_g) < \infty$ such that g_I is discontinuous on Γ_g and smooth elsewhere. More precisely we require that $\int_{\Omega - \Gamma_g} |\nabla g_I|^2 d\mu < \infty$, that there exists a constant L such that g_I, restricted to any straight line segment lying in $\Omega - \Gamma_g$, is a Lipschitz function with Lipschitz constant L, and that g actually has a discontinuity everywhere on Γ_g except, possibly, for a set having zero total length. Under these conditions the following holds.

Theorem 1. *For any fixed $\alpha > 0$ and $\epsilon > 0$ there exists $\beta^\star < \infty$ such that if $\beta \geq \beta^\star$ and Γ_β is minimal for E then*

$$d_H(\Gamma_g, \Gamma_\beta) < \epsilon.$$

The thrust of this theorem is that as $\beta \to \infty$ the set Γ_β can assume arbitrary geometries and will coincide with the discontinuity set of g_I. The theorem can be interpreted as an asymptotic fidelity result for the variational approach. It implies that the distortions resulting from *ad hoc* functions of this type are local, small-scale effects.

From a practical point of view the result is lacking because $\beta \to \infty$ forces f to match g_I exactly; noise in g_I will result in the appearance of many spurious boundaries. However, the theorem can be extended to incorporate noise and smearing effects. Roughly speaking, if the admissible noise magnitude scales as $o(\beta^{-1/2})$ and the admissible smearing acts over a radius of $o(\beta^{-1})$ then their presence can be tolerated and the theorem still holds. To give a more precise statement of this, let h_r and h_λ be any positive real valued functions

satisfying

$$\lim_{\beta\to\infty} \beta h_r(\beta) = 0,$$

$$\lim_{\beta\to\infty} \beta^{\frac{1}{2}} h_\lambda(\beta) = 0,$$

and let $\Psi(\beta)$ be the set of all functions g which can be written as

$$g = \mathcal{S}_r(g_I) + \lambda\omega,$$

where $0 \leq \lambda \leq h_\lambda(\beta)$ is a scalar, $\omega \in L^\infty(\Omega)$ with $|\omega|_\infty \leq 1$, $\mathcal{S}_r$ is any smearing operator satisfying

$$\mathcal{S}_r f(x) \in \left[\min_{y\in B_r(x)} f(y), \max_{y\in B_r(x)} f(y)\right],$$

where $B_r(x)$ is the disc of radius r centred at x with $0 \leq r \leq h_r(\beta)$, and g_I is the ideal data described earlier. We then have the following.

Theorem 2. *For any fixed $\alpha > 0$ and $\epsilon > 0$ there exists $\beta^\star < \infty$ such that if $\beta \geq \beta^\star$ and Γ_β is minimal for E for some $g \in \Psi(\beta)$ then*

$$d_H(\Gamma_g, \Gamma_\beta) < \epsilon.$$

The proofs are beyond the scope of this paper (they require a lengthy function analytic development) and have appeared in Richardson (1990, 1992). The relevant mathematical framework is outlined in Ambrosio (1989). This theorem indicates how noise and localization defects should scale with the parameter β to maintain fidelity.

Remark. The requirement $\lim_{\beta\to\infty} \beta^{1/2} h_\lambda(\beta) = 0$ seems to be necessary in general. There are certain images which asymptotically produce spurious edges for the piecewise constant version of the variational formulation if this condition is violated (Richardson 1990, 1992).

Our immediate goal, fulfilled in the next section, is to describe a heuristic which mimics the theorems to produce accurate localization of the edges in a manner independent of scale. Two problems immediately suggest themselves. First, a real image has fixed noise which cannot be scaled since it cannot be identified, and secondly, smearing is fixed and cannot (in general) be removed. The intrinsic noise and smearing of an image limits the recoverable accuracy of the edge locations. We are not proposing a scheme to eliminate that. Edge detection, ostensibly, will be performed on different scales. As we have indicated, operating on coarse scales tends to introduce distortions above and beyond those inherent in the signal. Our goal is to eliminate those distortions and recover, even on coarse scales, the same accuracy of localization usually achieved only at fine scales.

Since we intend to use an approximation to the variational formulation, it is prudent to consider whether the approximation deviates in a significant way from the original formulation with regard to distortion of edges. Although analysis is prohibitively difficult, we expect (and simulations have borne this out) that the spreading of the edges in the epi-convergent approximation actually ameliorates some of the geometric distortion. We recall

that the primary reason for the geometric distortion is that the term $|\Gamma|$ in E determines the structure of the singularities. Roughly speaking, this arises because the length term is one-dimensional and scales linearly while the other terms are two-dimensional integrals and hence scale quadratically in the size of the domain. (Actually this is not precisely true because singularities arise in f, but the dominance of the length term still occurs at singularities in Γ.) When the edges are smeared and length is replaced by a two-dimensional integral the concentration of cost in the length term is alleviated and, hence, we expect the distortion to be relaxed. The price paid for this is the lack of resolution of the edges. The algorithm begins with thick edges, thus relaxing distortion, but ends by sharpening the edges while scaling the parameters in accordance with theorem 2. Thus the edges are focussed as resolution increases.

4. Edge focussing via scaling

In this section we describe in detail the modifications to the descent equations, (8) and (9), which we introduce to focus edges. The essential idea is to satisfy the conditions of theorem 2 by smoothing g in a controlled way while allowing $\beta \to \infty$. Simultaneously we allow $c \to 0$ to sharpen the edges. We draw an analogy between the width of the edges and the diameter of the smearing operators appearing in theorem 2. Thus the rates of change of β and c are coupled.

We consider introducing dynamics into quantities β, g, and c, which in a standard minimization of E_c would be held fixed. These dynamics are intended to come into effect only after the basic descent equations (4) and (5) have essentially converged. Thus we assume $g(x, 0)$ is the initial data and $f(x, 0)$ and $v(x, 0)$ satisfy their respective Euler–Lagrange equations with $\beta = \beta(0)$ and $g(x) = g(x, 0)$. This implies the presence of a nominal set of edges, i.e. a function $v(x, 0)$. We will be guided by the heuristic that the subsequent focussing should only focus the edges already found and not introduce new ones. We make the following correspondences with the quantities which appear in theorem 2,

$$\begin{aligned} g(x,t) &\longleftrightarrow \mathcal{S}_{r(t)}(g_I) + \lambda(t)w(x,t), \\ f(x,t) &\longleftrightarrow \mathcal{S}_{r(t)}(g_I), \\ \text{where} \quad r(t) &= Kc(t) \text{ for some constant } K, \\ |\lambda(t)|_\infty &= |g(x,t) - f(x,t)|_\infty, \\ \text{and} \quad |w(x,t)|_\infty &= 1. \end{aligned}$$

We will discuss the choice of K and the meaning of the correspondences shortly. Consider first the following equations,

$$\frac{\partial}{\partial t} g(x, y, t) = \epsilon(f(x, y, t) - g(x, y, t)),$$

$$\frac{\partial}{\partial t} \beta(t) = \epsilon \beta(t),$$

where ϵ is a small positive constant included to reflect the fact that these equations are perturbations of (8) and (9). We observe that we obtain a solution to these equations such

that $\partial_f E = 0$ and $\partial_v E = 0$ for all t by setting

$$\begin{aligned} v(x,t) &= v(x,0), \\ f(x,t) &= f(x,0), \\ \beta(t) &= \beta(0)\,\mathrm{e}^{\epsilon t}, \\ g(x,t) &= g(x,0)\,\mathrm{e}^{-\epsilon t} + f(x,0)(1-\mathrm{e}^{-\epsilon t}). \end{aligned}$$

In fact, if $(f(x,0), v(x,0))$ minimizes E_c with data $g(x,0)$ and parameter $\beta(0)$ then it is easy to see that $(f(x,t), v(x,t))$ minimizes E_c with data $g(x,t)$ and parameter $\beta(t)$. These equations show that with this scaling we have $\beta(t) \to \infty$ while the minimal solutions $v(x,t)$ and $f(x,t)$ remain fixed. Interpreted in view of theorem 2 this means that g_I corresponds to $f(x,0)$ (i.e. $K = 0$). Note that according to our correspondences we have $\lambda(t) \propto \beta^{-1}(t)$, so the scaling conditions of theorem 2 are satisfied. We will now alter these equations slightly. First, we will introduce some dynamics into c to sharpen the edges. Second, we suppress the smoothing of g in a neighbourhood of the edges which shrinks with time; this is to permit focussing of the edges. Thus we consider

$$\frac{\partial}{\partial t} g(x,t) = \epsilon \rho(v(x,t))(f(x,t) - g(x,t)), \tag{10}$$

$$\frac{\partial}{\partial t} \beta(t) = \epsilon \beta(t), \tag{11}$$

$$\frac{\partial}{\partial t} c(t) = -\epsilon c(t), \tag{12}$$

where $\rho(v(x,t))$ should be approximately zero inside some neighbourhood of the edges and approximately one outside some larger neighbourhood. Furthermore, the width of the larger neighbourhood should shrink as $\beta^{-1}(t)$. A simple and reasonable choice, for example, is $\rho = \Phi$ since in this case the neighbourhood width is proportional to $c(t)$, which in turn is proportional to $\beta^{-1}(t)$. The algorithm takes the form of (8) and (9) until a local minima is reached, and subsequently (10)–(12) come into effect. Assuming $g(x,t)$ and $f(x,t)$ converge, we interpret the limit $g(x,\infty) = f(x,\infty)$ as corresponding to g_I and the set $S_g = \{x : \Phi(v(x,\infty)) \simeq 0\}$ as corresponding to Γ_g. The quantity K does not appear in our equations and is meant only to facilitate the correspondences; it should be interpreted as being sufficiently large so that the set $\{x : \rho(v(x,t)) < 1 - \delta\}$ for some small positive constant δ is contained in $[S_g]_{r(t)}$. We expect that $f(x,t) \simeq f(x,0)$ for all $x \notin [S_g]_{r(0)}$. Given this, the correspondence of $f(x,t)$ with $\Phi_{r(t)}(g_I)$ is consistent with the noise scaling of theorem 2.

In general the choice of ρ is a delicate issue. If g is noisy it may be desirable to allow more smoothing near the edges. The price for this is admitting potential distortion into the edges. In this situation a better smoothing mechanism might be a directionally controlled diffusion of g, suppressing diffusion across edges but enhancing diffusion parallel to the edge. This can easily be implemented within the framework developed here since $\nabla\Phi(v(x))$ will be perpendicular to an edge in a neighbourhood of that edge. (We have not experimented with this alternative.) Even when noise is not an issue ρ must be chosen carefully. We contend that the ideal choice should allow for edge correction and adequate smoothing without the introduction of edges from finer scales.

An alternative scaling, which deviates from the theory but may be advantageous under some conditions, is to allow β to depend on x and to scale it only around the edges. For example, one could replace (11) with $(\partial/\partial t)\beta(x,t) = \epsilon(1-\rho(v(x,t)))\beta(x,t)$ and eliminate the smoothing of g, (10). We have experimented with this variation in our simulations and the results are similar to those presented here.

5. Discretization and parameter choices

In this section we address some of the issues which arise as a consequence of discretization and further refine the algorithm. The lattice spacing (we will consider only square lattices) can be absorbed, via scaling, into the other parameters, β, c, and α. Appropriate step sizes for the discrete versions of the descent algorithm must be found, and the relative rates of the gradient descent and the scaling dynamics must be decided.

For the simulations presented in this paper f, g and v are discretized in a manner described below. Discrete versions of f and g are defined on a rectangular subset of a square lattice while the discrete version of v is defined on a similar subset of a square lattice which is twice as dense and rotated by 45°. This is not necessary, but it facilitates the discrete implementation. We define the following subset of $\mathbb{Z}^2 \subset \mathbb{R}^2$,

$$\mathcal{L}_f = \{(i,j) : i = 1,\ldots,N,\ j = 1,\ldots,M\}.$$

We assume that g is defined on $\mathcal{L}_f$. The nearest neighbours of $x \in \mathcal{L}_f$ are defined by

$$\mathcal{N}_f(x) = \{x' \in \mathbb{Z}^2 : |x' - x| = 1\}.$$

Similarly, we define

$$\mathcal{L}_v = \{(x+x')/2 : x \in \mathcal{L}_f, x' \in \mathcal{N}_f(x) \cap \mathcal{L}_f\},$$

which will support v, and the nearest neighbours

$$\mathcal{N}_v(y) = \{y' = (x+x')/2 : x \in \mathcal{L}_f, x' \in \mathcal{N}_f(x), |y - y'| = 1/\sqrt{2}\}.$$

For discretization we can take the approach of discretizing E_c and then deriving discrete Euler-Lagrange equations, or we can discretize the Euler–Lagrange equations directly. We consider the first approach first. A discrete version of E_c with lattice spacing δ is the following,

$$E_d = \beta \sum_{x \in \mathcal{L}_f} \left(\delta^2 (f(x) - g(x))^2 + \frac{1}{2} \sum_{x' \in \mathcal{N}_f(x)} \Phi((x+x')/2)(f(x) - f(x'))^2 \right)$$

$$+ \alpha \sum_{y \in \mathcal{L}_v} \left(\frac{c}{4} \sum_{y' \in \mathcal{N}_v(y)} (\Psi(y) + \Psi(y'))(v(y) - v(y'))^2 + \frac{\delta^2}{8c}(1 - v(y))^2 \right),$$

where if $x \in \mathcal{L}_f$ and $x' \in \mathcal{N}_f(x) \backslash \mathcal{L}_f$ then we impose the condition $f(x') = f(x)$ (this defines $f(x')$ uniquely), and if $y \in \mathcal{L}_v$ and $y' \in \mathcal{N}_v(y) \backslash \mathcal{L}_v$ then we impose the condition

$v(y') = v(y'')$, where $y'' \in \mathcal{N}_v(y)$ is the unique point satisfying $|y' - y''| = 1$. This is to ensure Neumann type boundary conditions in the descent equations given below. By making the substitutions

$$\beta \to \delta^{-2}\beta, \quad c \to \delta^2 c, \quad \alpha \to \delta^{-2}\alpha,$$

we find that without loss of generality we can set $\delta = 1$, which we do henceforth. A discrete form of the Euler–Lagrange equations can now be found directly by differentiating E_d with respect to $v(y)$ and $f(x)$. To simplify the notation we will write $\Psi(y)$, $\Phi(y)$ instead of $\Psi(v(y))$, $\Phi(v(y))$. For each $y \in \mathcal{L}_v$ the pair x, x' such that $y = (x + x')/2$ and $x \in \mathcal{L}_f, x' \in \mathcal{N}_f(x) \cap \mathcal{L}_v$ is uniquely determined. Thus, for each such y we can set $\mathrm{d}f(y) = (f(x) - f(x'))^2$. The derivatives of Ψ and Φ will be denoted $\dot{\Psi}$ and $\dot{\Phi}$ respectively. For each $x \in \mathcal{L}_f$ and $y \in \mathcal{L}_v$ we define

$$\partial_x E_d = \beta(f(x) - g(x)) + \sum_{x' \in \mathcal{N}_f(x)} \Phi((x + x')/2)(f(x) - f(x')) \tag{13}$$

$$\begin{aligned} \partial_y E_d = {} & \alpha^{-1}\dot{\Phi}(y)df(y) - \frac{1 - v(y)}{4c} + c \sum_{y' \in \mathcal{N}_v(y)} (v(y) - v(y')), \\ & \times \left(\Psi(y) + \Psi(y') + \frac{1}{2}\dot{\Psi}(y)(v(y) - v(y'))\right), \end{aligned} \tag{14}$$

which are proportional to $[\partial/\partial f(x)]E_d$ and $[\partial/\partial v(y)]E_d$ respectively. These equations are discrete analogs of (5) and (6) respectively. (The constants in (14) are slightly different from those in (6) because the lattice spacing of $\mathcal{L}_v$ is $1/\sqrt{2}$.) Now we consider discretizing the Euler–Lagrange equations directly. A discrete version of (7) is

$$\begin{aligned} \partial_y E_d = {} & \alpha^{-1}\dot{\Phi}(y)df(y) - \frac{1 - v(y)}{4c} + c \sum_{y' \in \mathcal{N}_v(y)} (v(y) - v(y')) \\ & \times \left(2\Psi(y) - \frac{1}{2}\dot{\Psi}(y)(v(y) - v(y'))\right). \end{aligned} \tag{15}$$

A third alternative is to average (6) and (7) and discretize, or equivalently, to average (14) and (15). If we do so, we obtain

$$\begin{aligned} \partial_y E_d = {} & \alpha^{-1}\dot{\Phi}(y)df(y) - \frac{1 - v(y)}{4c} \\ & + c \sum_{y' \in \mathcal{N}_v(y)} (v(y) - v(y')) \left(\frac{3}{2}\Psi(y) + \frac{1}{2}\Psi(y')\right). \end{aligned}$$

From the point of view of implementation this form is much more attractive since it simplifies computation. Finally, one could approximate $3/2\Psi(y) + 1/2\Psi(y') \simeq 2\Psi(y)$ to obtain an even simpler form. Although this deviates from the theory, we use it in our implementation.

Allowing all quantities to depend on t, our basic descent equations take the form,

$$\begin{aligned} f(x, t + 1) - f(x, t) &= -c_f \partial_x E_d, \\ v(x, t + 1) - v(x, t) &= -c_v \partial_y E_d, \end{aligned}$$

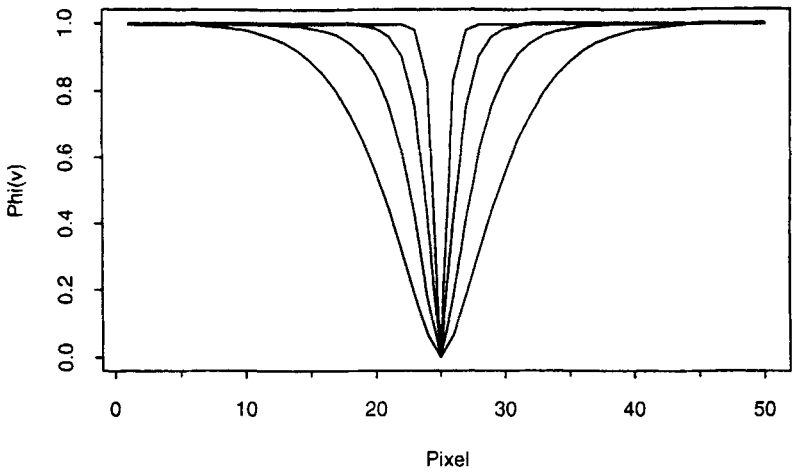

Figure 2. Graphs of minimal $\Phi(v(25, j + 0.5))$ for step edge for $c = 2.0, 1.0, 0.5, 0.2$. In each case $\alpha = 0.05$, $\beta = 2.0$, and Φ and Ψ were chosen as in (2).

where on the right hand side all quantities are evaluated at time t. We now address the question of choosing c_f and c_v. A standard gradient descent would have both c_f and c_v constant. If we try to set c_v constant then it must be chosen small since $(f(x) - f(x'))^2/\alpha$ can be quite large and hence convergence will be slow. A computationally efficient choice that gives much faster convergence is to approximate a Newton-type descent. If we set

$$c_v(y) = \frac{1}{2}\left(\frac{\dot{\Phi}(y)}{\alpha v(y)}\mathrm{d}f(y) + \frac{1}{4c} + 8c\Psi(y)\right)^{-1},$$

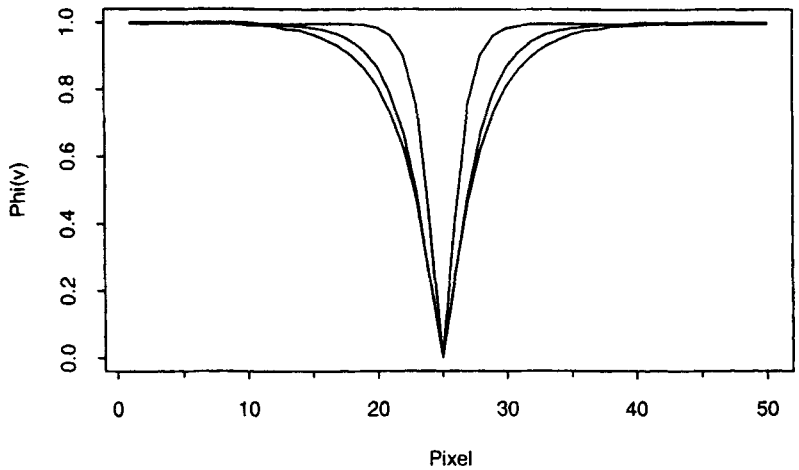

Figure 3. Graphs of minimal $\Phi(v(25, j + 0.5))$ for step edge for $\alpha = 0.05$, $\beta = 2.0$, and $c = 0.5$ where Φ and Ψ are given by (4) with (n, m) given by $(2, 0)$ $(2, 1/2)$, and $(2, 1)$ respectively where increasing m increases the spread of the function.

where we have suppressed dependence on t, and make the approximations mentioned above, then we obtain

$$v(y, t+1) = \frac{1}{2}\left(v(y,t) + \frac{1/4c + 2c\Psi(y)\sum_{y' \in \mathcal{N}_v(y)} v(y')}{\Phi(y)\mathrm{d}f(y)/(\alpha v(y)) + 1/4c + 8c\Psi(y)}\right). \quad (16)$$

Our implementation employs this form of update. It has particularly attractive properties. Note, in particular, that $v(y, t+1)$ is an average of $v(y,t)$ and a well-behaved quantity which lies between 0 and 1. Setting c_f constant is much less problematic, and for our simulations we have set it to $(2\beta + 8)^{-1}$ since this gives a good rate of convergence without allowing overshoot in f. The initialization of f and v will affect which local minimum is reached by the initial gradient descent. We expect this will have little effect on the edge-focussing part of the algorithm. In our simulations we have set $f(0) = g(0)$ and $v(0) = 1$. With these choices we observe that the basic descent on f and v converges in about 30 iterations for the range of parameters we have experimented with. (Larger values of c and smaller values of β will reduce the rate of convergence.)

We now consider the discrete scaling dynamics. We recall that these equations come into effect only after the gradient descent has nearly converged. The following are discrete analogs of (10)–(12),

$$\begin{aligned} g(x, t+1) &= g(x,t) + \epsilon\rho(x,t)(g(x,t) - f(x,t)), \\ \beta(t+1) &= (1-\epsilon)^{-1}\beta(t), \\ c(t+1) &= (1-\epsilon)c(t), \end{aligned}$$

where ρ and ϵ are to be specified in each case. Approximations that save computation can be made here. For example one could compute this update once every n time steps instead of every time step and increase ϵ appropriately. Also, since $f(x,t)$ remains essentially fixed outside of some neighbourhood of the edges, one could restrict the update on β to some neighbourhood of the edges and completely drop the update on g producing essentially the same effect, as mentioned earlier. This would yield substantial time savings in serial implementations. Our implementation employs the equations given above.

If $c(t)$ becomes too small then the discrete approximation to E_c will break down. The value of $c(t)$ can be used as a stopping criteria. For the choices of Ψ and Φ used in this paper we allow $c(t)$ to become small enough so that the effective edge width is one pixel. (Effective edge width can be defined as the width of the set $\{\Phi(y) < 1/2\}$ for example.)

6. Simulation results

For all of our simulation results we have scaled g so that $g(x) \in [0, 1]$. In particular, solid white is 1 and solid black is 0. In the two-dimensional plots and in the images we plotted $\Phi(v)$ on the same mesh as f, i.e. the plotting mesh corresponds to $\mathcal{L}_f$. For each $x \in \mathcal{L}_f$ we plot the minimum value of $\Phi(v)$ among the four nearest neighbours. (Note that an edge 'set' could be defined as $\{y \in \mathcal{L}_v : \Phi(v(y)) < 1/2.\}$

The first two simulations illustrate basic properties of the epi-convergent approximation. Here g is defined on a 50×50 square mesh with $g(i, j) = 0$ for $j \leq 25$ and 1 otherwise. Figure 2 illustrates dependence of the edge smearing on c. We plot $\Phi(v(25, j+0.5))$, $j =$

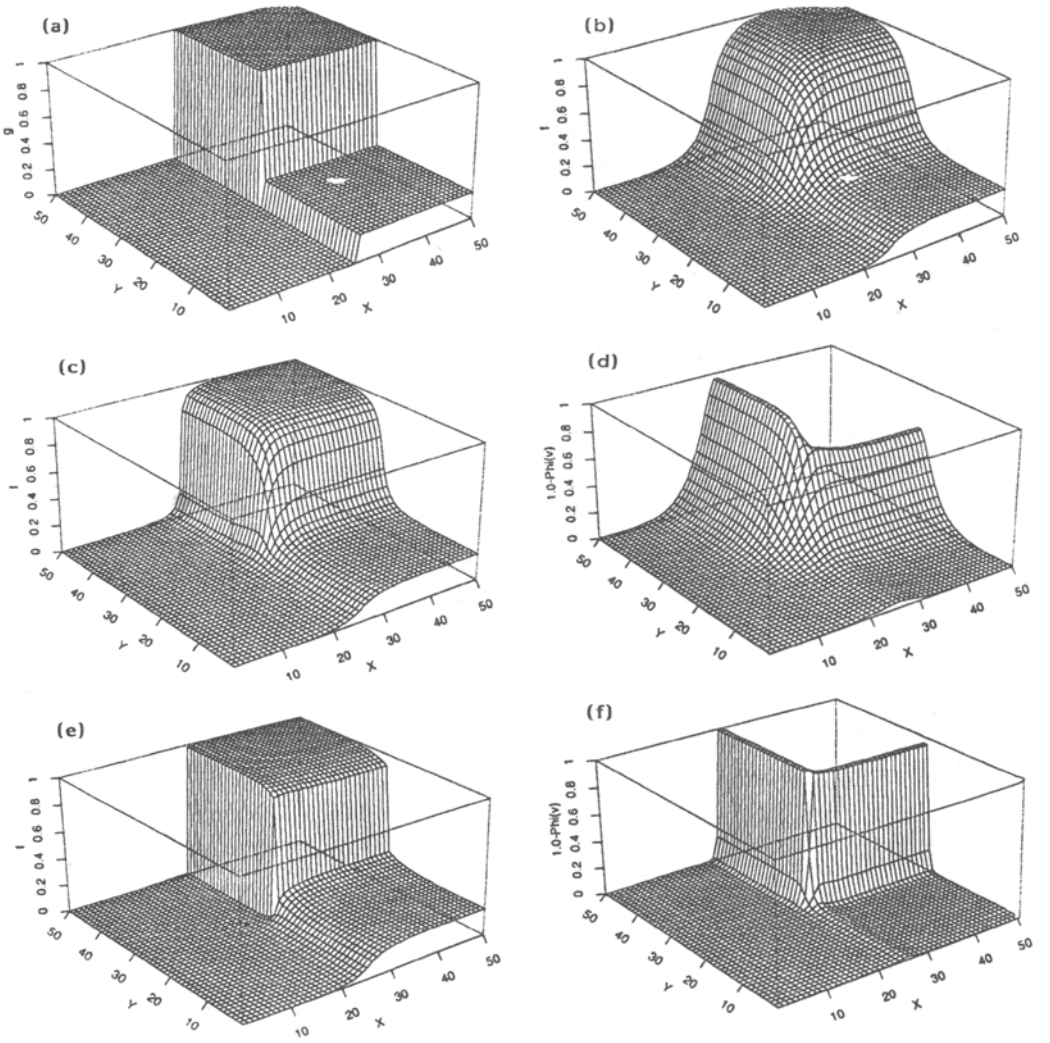

Figure 4. (**a**) Synthetic T-junction/corner data. (**b**) Minimal f with $v = 1$ with $\beta = 0.1$. (**c**) Minimal f with $\beta = 0.1$, $\beta = 0.1$, and $c = 2.0$ (**d**) Minimal $\Phi(v)$ with $\beta = 0.1$, $\alpha = 0.1$, and $c = 2.0$ (**e**) Minimal f after edge focussing with final values of $\beta = 1.0$, $\alpha = 0.1$, and $c = 0.2$. (**f**) Minimal $\Phi(v)$ after edge focussing with final values of $\beta = 1.0$, $\alpha = 0.1$, and $c = 0.2$.

$1, \ldots, 50$ for $c = 2.0, 1.0, 0.5, 0.2$ after convergence of the descent equations (without scaling). In each case $\alpha = 0.05$, $\beta = 2.0$, and Φ and Ψ were chosen according to (2). Figure 3 is similar except that we have fixed $c = 0.5$ and varied Φ and Ψ. They are given by (4) with $(n, m) = (1, 0), (1, 1/2), (1, 1)$ respectively (increasing m tends to increase the width of the edge). This is to illustrate how the shape of the edges can be changed by altering Φ and Ψ.

To demonstrate the behaviour of the edge focussing algorithm on high curvature edges we have simulated the algorithm on the data presented in figures 4 a–f. The functions Φ and Ψ are as in (2) and we set $\rho(v) = \Phi(v)$. We have carefully chosen the parameters to make the detection of the corner marginal. The initial values are $\beta(0) = 0.1$, $\alpha = 0.1$,

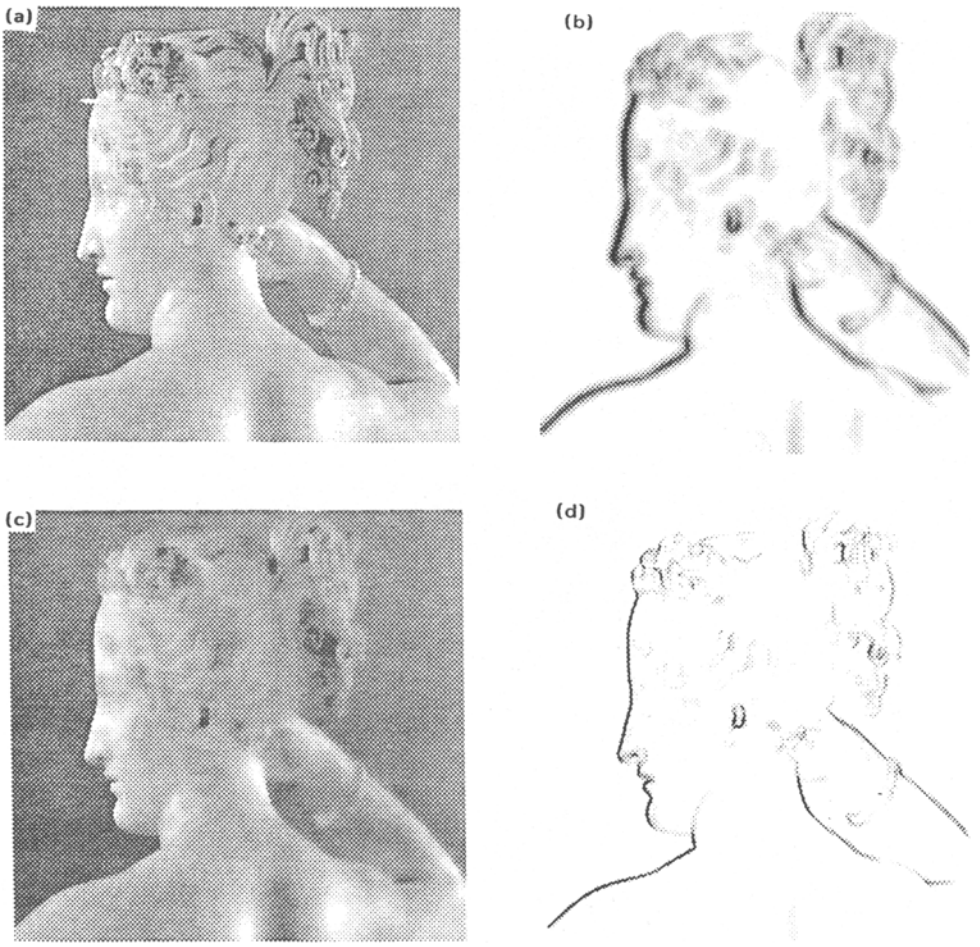

Figure 5. (**a**–**d**) (Caption on facing page.)

and $c(0) = 2.0$. The final values are $\beta(T) = 1.0$ and $c(T) = 0.2$ (where T is the time of termination), and ϵ was chosen so that 200 iterations with scaling are required. Figure 4a is the data set g. Figure 4b plots the optimal f when $\alpha = \infty$, i.e. no edges are allowed. This is to indicate the degree of smoothing associated with this value of β. Figure 4c shows $f(0)$, i.e. the f obtained after allowing a descent to converge with the parameters held fixed at their initial values. Figure 4d shows $\Phi(v(0))$. We have chosen the parameters so that the three edges in the image are detected to different degrees. Figures 4e and 4f are $f(T)$ and $\Phi(v(T))$ respectively. Note that the two larger edges and the corner they form is unambiguously detected while the smallest edge has been smooth out. The slight smoothing visible across the second largest edge is due to slight smoothing feedback at that edge. The function $g(T)$ is not distinguishable from $f(T)$ so we have not included it.

Figures 5 and 6 demonstrate the algorithm on 'real' images. Figure 5 is 480×512 pixels and figure 6 is 512×512 pixels. In general ϵ is chosen so that 200 iterations with scaling are required. The data are in figure (a) in each case. Each image has been processed for two different sets of parameters to indicate the stability of the edges under a change in

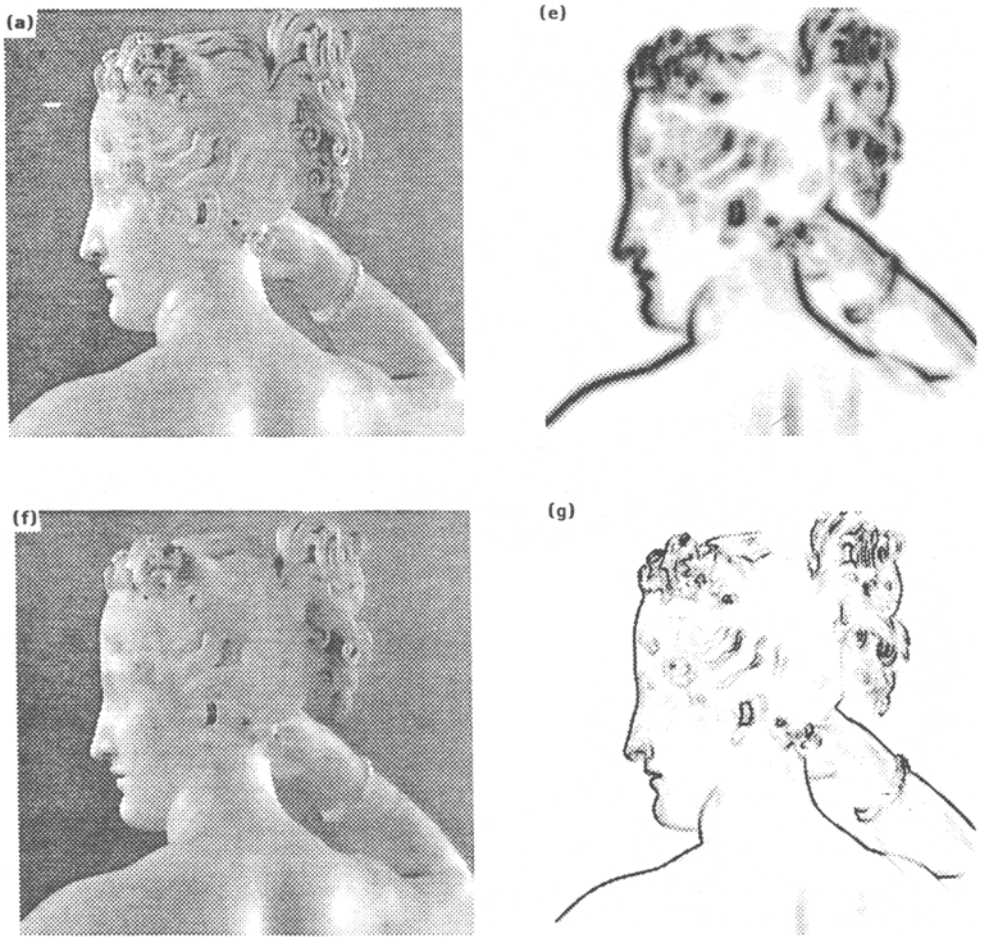

Figure 5. (**a**) Statue image (Paolina Borghese, Canova circa 1800) 480 × 512. Prescaling $\Phi(v)$ (**b**), final $\Phi(v)$, (**c**), and final f (**d**), with initial parameters $\beta = 0.04$, $\alpha = 0.01$, $c(0) = 1.0$ and final parameters $\beta = 0.2$, $c = 0.2$. Prescaling $\Phi(v)$ (**e**) final $\Phi(v)$ (**f**), and final f (**g**), with initial parameters $\beta = 0.1$, $\alpha = 0.01$, $c(0) = 1.0$ and final parameters $\beta = 0.5$, $c = 0.2$.

scale. The parameters α and β are an order of magnitude smaller in figure 6. This admits much greater smoothing of the image. Even so, edge localisation is accurate. In both cases the displayed images are the following. Figures (b), (c), and (d) are $\Phi(v(0))$, $f(T)$, and $\Phi(v(T))$ respectively. Figures (e), (f), and (g) reiterate (b), (c), and (d) for the second set of parameters in each case. In figure 5 the first set of parameters is given by

$$\beta(0) = 0.1, \quad \alpha = 0.25, \quad c(0) = 2.0, \quad \beta(T) = 1.0, \quad c(T) = 0.2,$$

and the second by

$$\beta(0) = 0.1, \quad \alpha = 0.1, \quad c(0) = 2.0, \quad \beta(T) = 1.0, \quad c(T) = 0.2.$$

In figure 6 the first set of parameters is given by

$$\beta(0) = 0.0025, \quad \alpha = 0.05, \quad c(0) = 2.0, \quad \beta(T) = 0.025, \quad c(T) = 0.2.$$

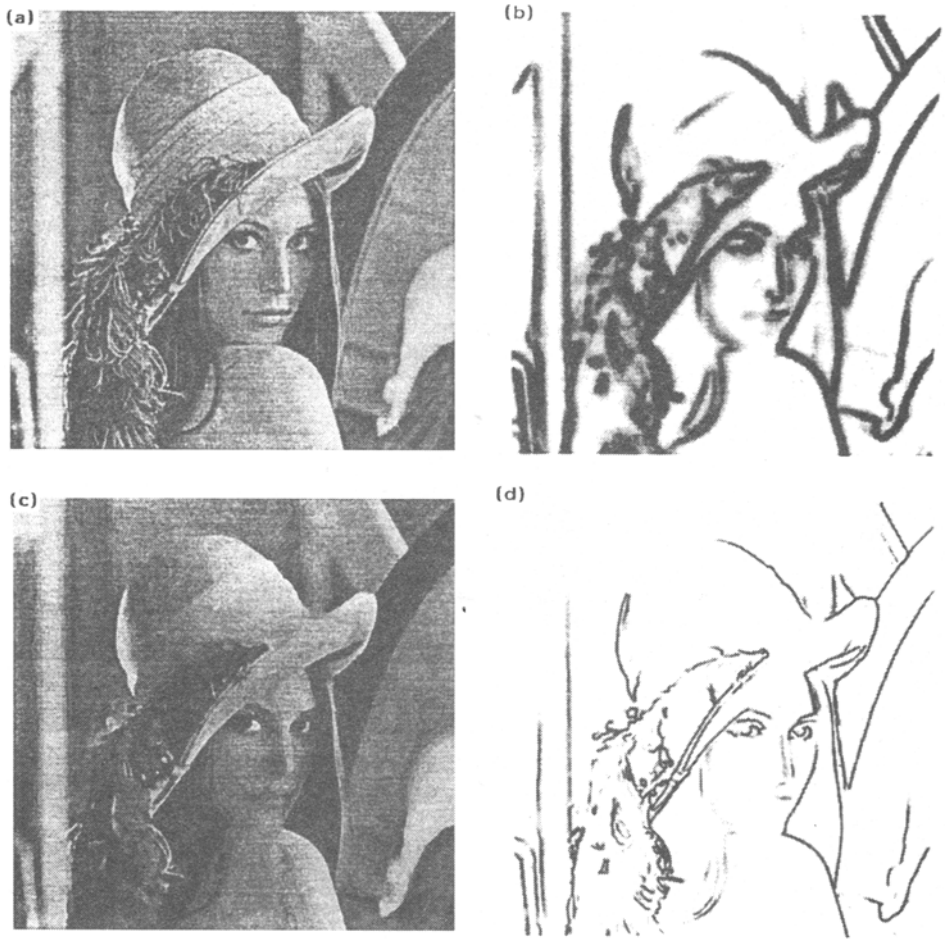

Figure 6. (a–d) (Caption on facing page.)

and the second by

$$\beta(0) = 0.004, \quad \alpha = 0.03, \quad c(0) = 2.0, \quad \beta(T) = 0.04, \quad c(T) = 0.2.$$

We observe that as scale increases the set of edges detected increase monotonically – virtually no scale-dependent distortion is visible.

This work was partially supported by the Office of Naval Research under contract N00014-77-0532, by the Air Force Office of Scientific Research under contract AFOSR-85-0227 and by the Army Research Office under contracts DAAG-29-84-K-005 and DAAL03-86-K-0171. Some of this work was done while TJR was at the Laboratory for Informations and Decision Systems, MIT.

We are indebted to Pietro Perona and to Stefano Casadei for help with images and image processing software and for helpful discussions.

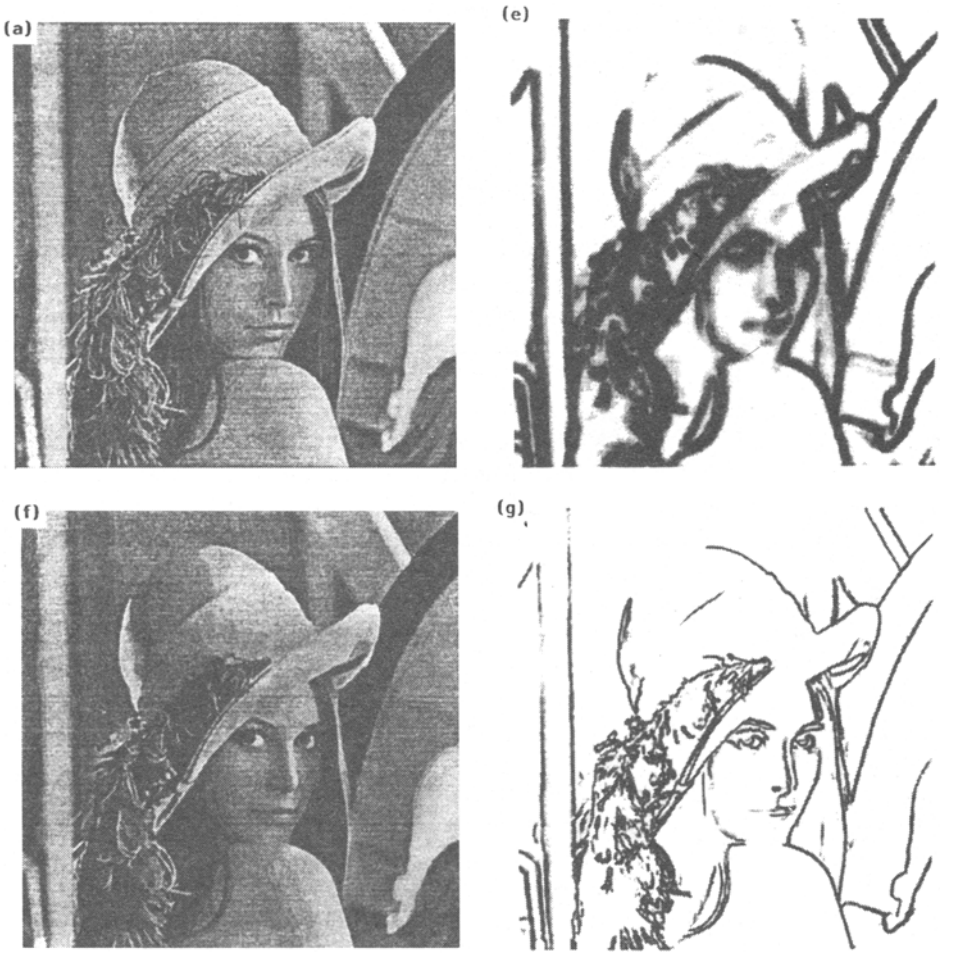

Figure 6. (**a**) Lenna image 512 × 512. Prescaling $\Phi(v)$ (**b**), final f (**c**), and final $\Phi(v)$ (**d**), with initial parameters $\beta = 0.0025$, $\alpha = 0.05$, $c(0) = 2.0$ and final parameters $\beta = 0.025$, $c = 0.2$. Prescaling $\Phi(v)$ (**e**), final f (**f**), and final $\Phi(v)$ (**g**), with initial parameters $\beta = 0.004$, $\alpha = 0.03$, $c(0) = 2.0$ and final parameter $\beta = 0.04$, $c = 0.2$.

References

Ambrosio L 1989 Variational problems on SBV. *Acta Appl. Math.* 17: 1–40

Ambrosio L, Tortorelli V 1990 Approximation of functionals depending on jumps by elliptic functionals via Γ–convergence. *Commun. Pure Appl. Math.* 43: 999–1036

Ambrosio L, Tortorelli V 1992 On the approximation of functionals depending on jumps by elliptic functionals. *Boll. Un. Mat. Ital.* 7: 105–123

Attouch H 1984 *Variational convergence for functions and operators* (London, UK: Pitman)

Bergholm F 1987 Edge Focusing. *IEEE Trans. Pattern Anal. Machine Intell.* 9: 726–741

Bilbro G L, Snyder W E, Garnier S J, Gault J W 1992 Mean field annealing: A formalism for constructing GNC-like algorithms. *IEEE Trans. Neural Networks* 3: 131–138

Blake A, Zisserman A 1987 *Visual reconstruction* (Cambridge, MA: MIT Press)

Canny J 1986 A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.* 8: 679–698

De Giorgi E, Franzoni T 1979 Su un tipo di convergenze variazionale. *Ren. Sem. Mat. brescia* 3: 63–101

Gamble E B, Poggio T 1987 Visual integration and detection of discontinuities: The key role of intensity edges. A I Memo No. 970, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA

Geiger D, Girosi F 1989 Parallel and deterministic algorithms for MRFs: surface reconstruction and integration, Memo No. 1114

Geiger D, Yuille A 1989 A common framework for image segmentation. Tech. Rep. no. 89–7, Harvard Robotics Laboratory, Harvard University, Cambridge, MA

Geman S, Geman D 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.* 6: 721–741

Grimson W E L 1981 *From images to surfaces* (Cambridge, MA: MIT Press)

Horn B K T 1986 *Robot vision* (Cambridge, MA: MIT Press)

Harris J, Koch C, Luo J, Wyatt J, 1989 Resistive fuses: Analog hardware for detecting discontinuities in early vision. *Analog VLSI implementations of neural systems* (Norwell, MA: Kluwer) pp 27–55

March R 1988 Computation of stereo disparity using regularization. *Pattern Recognition Lett.* 8: 181–187

March R 1989 A regularization model for stereo vision with controlled continuity. *Pattern Recognition Lett.* 10: 259–263

March R 1992 Visual reconstruction with discontinuities using variational methods. *Image Vision Comput.* 10:

Marr D, Hildreth E 1980 Theory of edge detection. *Proc. R. Soc. London B* 207: 187–217

Marroquin J L 1985 *Probabilistic solution of inverse problems.* Ph D thesis, Dept. of Electrical Engineering and Computer Science, Massachussetts Institute of Technology, Cambridge, MA

Mumford D, Shah J 1985 Boundary detection by minimizing functionals. *IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco

Mumford D, Shah J 1989 Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* 42: 577–685

Nordström K N 1990 Biased anisotropic diffusion: a unified regularization and diffusion approach to edge detection. *Image Vision Comput.* 8: 318–327

Perona P, Malik J 1990 Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.* 12: 629–639

Richardson T J 1990 *Scale independent piecewise smooth segmentation of images via variational methods*, Ph D thesis, Dept. of Electrical Engineering and Computer Science, Massachussetts Institute of Technology, Cambridge, MA

Richardson T J 1992 Limit theorems for a variational problem arising in computer vision. *Annali della Scuola Normale* 19: 1–49

Rosenfeld A, Thurston M 1971 Edge and curve detection for visual scene analysis. *IEEE Trans. Comput.* C-20: 562–569

Shah J 1991 Segmentation by non-linear diffusion. *Proc. IEEE Comput. Vision Pattern Recognition 91, Hawai*

Shah J 1992 Segmentation by minimizing functionals: Smoothing properties. *SIAM J. Control Optimization* 30: 99–111

Witkin A 1983 Scale-space filtering. *International Joint Conference on Artificial Intelligence* Karlsruhe, pp 1019–1021