Biased Anisotropic Diffusion — A Unified Regularization and Diffusion Approach to Edge Detection

K. Niklas Nordström
Department of Electrical Engineering and Computer Sciences
University of California
Berkeley, CA 94720

31 May 1989

Abstract

We present and analyze a global edge detection algorithm based on variational regularization. The algorithm can also be viewed as an anisotropic diffusion method. We thereby unify these two, from the original outlook, quite different methods. This puts anisotropic diffusion, as a method in early vision, on more solid grounds; it is just as well-founded as the well-accepted standard regularization techniques. The unification also brings the anisotropic diffusion method an appealing sense of optimality, thereby intuitively explaining its extraordinary performance.

The algorithm to be presented moreover has the following attractive properties.

1. It only requires the solution of a *single* boundary value problem over the *entire* image domain — almost always a very simple (rectangular) region.
2. It converges to the solution of interest.

The first of these properties implies very significant advantages over other existing regularization methods; the computation cost is typically cut by an order of magnitude or more. The second property represents considerable advantages over the existing diffusion methods; it removes the problem of deciding when to stop, as well as that of actually stopping the diffusion process.

1 Introduction

The purpose of computer vision is to generate useful descriptions of the environment from an *original image function* ζ defined on some open bounded connected image domain B. This image function can represent various kinds of data, collected from the visible surfaces in the scene. Common examples are brightness data, color data, depth data, etc. We will be concerned with image functions representing brightness data. Thus $\zeta : B \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$. For the purpose of describing the environment the most useful information in the scene is often contained in the discontinuities of the (functions representing) depth, surface orientation, reflectance properties and illumination of the visible surfaces, all of which bring about discontinuities in the *true image function*, one would obtain by pure projection of the brightness in the scene onto the image domain. If the true image function was known, this problem would be easy. However because of imperfections, such as blurring, noise, discretization, sensor non-linearities, etc., which are present in any physical image formation

process, the original image function, one is given, is distorted, so that the discontinuities in the true image function are disguised into large gradients. Global edge detection therefore essentially boils down to numerical differentiation — a problem well-known to be ill-posed (in the sense of Hadamard) due to its instability with respect to the initial data [1,2,3]. Since measurement noise and other undesirable disturbances, (for example dirt, dust, bugs and other textures, which are too fine to be relevant for a useful description of the environment,) cannot be avoided, the global edge detection problem thus has to be stabilized, in order to have a meaningful solution. In more practical terms, this means, that the undesirable disturbances must be suppressed, without the disappearance or dislocation of any of the edges. Over the last six years or so many attempts along these directions have appeared in the literature. One can distinguish between two seemingly quite different approaches, viz. regularization and anisotropic diffusion.

The regularization approaches seek formal stabilization of the global edge detection problem (with respect to some topologies on the initial data and solution spaces). This can be done in different ways. In probabilistic regularization [4,5] the problem is reformulated as Bayesian estimation. In variational regularization [6,7,8,9,10,11,12,13,14,15] it is posed as a cost (or energy) functional minimization problem, leading to a variational principle. In spite of the different outlooks of these approaches they essentially end up with the same mathematical and computational problem; given the original image function ζ, minimize a cost functional of the form

$$\mathcal{C}_\zeta(w,z) \doteq \mathcal{E}(w) + \mathcal{D}(z,\zeta) + \mathcal{S}(w,z)$$

where w is some function representing the edges, and $z : B \to \mathbb{R}$ is the so called *estimated* (or reconstructed) *image function.* The function w might be defined in a variety of ways. In this paper we will only be concerned with edge functions of the form $w : B \to \mathbb{R}$. The purpose of the *edge cost* $\mathcal{E}$ is to impose an explicit penalty for the presence of edges, thereby preventing pathological solutions, whose edges fill up large parts of or even the entire image domain. The *deviation cost* $\mathcal{D}$ ensures, that the estimated image function z is a faithful approximation of the original image function ζ. The stabilizing cost $\mathcal{S}$ stabilizes the problem with respect to the initial data. Given the edge function w it is typically the case, that there exists a unique optimal estimated image function $\check{z}_w$, which can be found by solving a linear partial differential equation — a condition generally taken advantage of. For the minimization with respect to w however, all of the regularization approaches, referred to above, resort to some kind of stochastic or deterministic search method, such as the Metropolis algorithm or steepest descent. Because of the tremendous size of the solution space any such search method is by itself quite expensive. In addition the general non-convexity of the cost function causes any converging search algorithm to get stuck at local minima. The common response to this unfortunate situation has been to solve whole sequences of minimization problems, as a mechanism for "gravitating" towards a good local possibly a global minimum. The GNC-algorithm introduced in [12,13] and simulated annealing [4] are both examples thereof. As a consequence every global edge detection method up to date involves some form of repeated iterative minimization process, and because of the tremendous computational cost resulting therefrom, the optimality of the solution is often compromised. In summary each of these methods is extensively expensive and/or yields suboptimal solutions.

Anisotropic diffusion has been introduced in early vision [16,17,18] as a method of detecting edges at a continuum of scales of resolution, without weakening or dislocating any of the edges at the scales of interest — a well-known problem with isotropic diffusion and other

linear filtering operations. The anisotropic diffusion method does not seek an optimal solution of any kind. Like other common scale-space methods it instead operates by repeatedly filtering the image function with a smoothing kernel of small support, thereby producing a sequence of *diffused image functions* of successively lower resolution. In order to retain the strength and correct location of the edges at the resolution of interest the "smoothing power" of the filter kernel is made to depend (inversely) on the magnitude of the image function gradient in a somewhat heuristic fashion. Finally all edges disappear, and the diffused image function converges to a constant. At some stage in the iterated filtering process remarkably impressive edges can be obtained by postprocessing the diffused image function with the most rudimentary local edge detector; it basically suffices to threshold the absolute difference between nearest neighbor pixel values. The task of finding this stage however, has so far been a matter of manual inspection.

In this paper we present a global edge detection algorithm based, on variational regularization. As it turns out however, it can also be viewed as a (new) *biased* anisotropic diffusion method. We thereby unify the from the original outlook quite different methods of regularization and anisotropic diffusion. This puts anisotropic diffusion, as a method in early vision, on more solid grounds; it is just as well-founded as the well-accepted standard regularization techniques. The unification also brings the anisotropic diffusion method an appealing sense of optimality, thereby intuitively explaining its extraordinary performance. The algorithm to be presented moreover has the following attractive properties:

1. It only requires the solution of a *single* boundary value problem on the *entire* image domain — almost always a very simple region.

2. It converges to the solution of interest.

The first of these properties implies a number of advantages over other existing regularization methods. In particular:

(i) No search methods are necessary.

(ii) No sequences of minimization problems have to be solved.

(iii) The computational cost is relatively very low.

The second property represents a couple of advantages over the existing diffusion methods:

(i) It removes the problem of manual selection of, which one in the sequence of diffused image functions, to be postprocessed with the local edge detector.

(ii) It is superior for hardware implementations.

The rest of this paper is organized as follows: In the next section we review Terzopoulos' controlled-continuity stabilizers for early vision problems. In section 3 we propose our modification of his paradigm, and derive the resulting conditions for optimality. In section 4 we compare our variational edge detection method with the anisotropic diffusion algorithm introduced by Perona and Malik. In section 5 and 6 we study some properties of the biased anisotropic diffusion. In section 7 we discuss discretizations of the variational edge detection problem, and propose a numerical solution. Section 8 is devoted to convergence, uniqueness and stability analysis of the discretized problem and the proposed algorithm. Finally section 9 covers our experimental results.

Part of our notation, although here consistent, carries common multiple meanings in the literature, and needs therefore to be settled: The spaces $\{0,1,2,\ldots\}$, $]-\infty,0[$ and $]0,\infty[$ will be denoted by $\mathbf{N}_0$, $\mathbf{R}_-$ and $\mathbf{R}_+$ respectively. The Euclidean norm in $\mathbf{R}^n$ will be written $\|\cdot\|$, while L_∞-norms will be written $\|\cdot\|_\infty$. Difference operators will be denoted by $\blacktriangle$, to distinguish them from the Laplacian Δ. Finally the binary maximum, minimum and function composition operators will be denoted by $\vee$, $\wedge$ and $\circ$ respectively.

2 Controlled-Continuity Stabilization

The purpose of the stabilizing cost $\mathcal{S}$ is to restrict the space of possible estimated image functions, and thereby regularize the (edge detection) problem, so that, as the name suggests, stability with respect to the initial data is achieved. Several classes of such stabilizing functionals have been studied in the mathematical theory of ill-posed problems. This theory was pioneered by Tikhonov [19,20]. He proposed a general class of stabilizers for univariate regularization of the form

$$\mathcal{S}(z) \doteq \int_{\mathbf{R}} \sum_{i=1}^{I} w_i \left(\frac{d^i z}{dx^i}\right)^2 dx$$

where $w_1,\ldots w_I : \mathbf{R} \to \overline{\mathbf{R}_+}$ are prespecified continuous weighting functions. For multivariate regularization generalized spline functionals of the form

$$\mathcal{S}(z) \doteq \int_{\Omega} \sum_{k_1=1}^{K} \cdots \sum_{k_I=1}^{K} \left(\frac{\partial^I z}{\partial x_{k_1} \cdots \partial x_{k_I}}\right)^2 dx$$

have been considered with varying generality of the domain $\Omega \subseteq \mathbf{R}^K$, its dimension K and the "order of regularization" I [21,22,23,24].

Examples of the stabilizers above have been used for regularization of a wide range of early vision problems with varying degree of success [3]. A common flaw of these stabilizers in this context is, that they do not allow the estimated image function z to be discontinuous. This problem was addressed by Terzopoulos [7,8], who proposed further generalizations of the multivariate generalized spline functionals. His stabilizing functionals, referred to as *controlled-continuity stabilizers* are given by

$$\mathcal{S}(w,z) \doteq \int_{\mathbf{R}^K} \sum_{i=1}^{I} w_i \sum_{k_1=1}^{K} \cdots \sum_{k_i=1}^{K} \left(\frac{\partial^i z}{\partial x_{k_1} \cdots \partial x_{k_i}}\right)^2 dx$$

where $w \doteq [w_1 \cdots w_I]^T$, and the weighting functions $w_1,\ldots,w_I : \mathbf{R}^K \to [0,1]$, referred to as *continuity control functions* are in general discontinuous. In particular they are able to make jumps to zero, and edges, where the partial derivatives of z of order $\geq j$ are allowed to be discontinuous, are represented by the sets

$$\bigcap_{i=j+1}^{I} w_i^{-1}(\{0\}) \qquad j = 0,\ldots,I-1 \tag{1}$$

For the edge cost Terzopoulos proposes the functional

$$\mathcal{E}(w) \doteq \int_{\mathbf{R}^K} \sum_{i=1}^{I} \lambda_i (1 - w_i)\, dx$$

where the constants $\lambda_1, \ldots, \lambda_I \in \overline{\mathbb{R}_+}$ satisfy $\sum_{i=1}^{I} \lambda_i > 0$. An apparent problem with this paradigm is, that it fails to support a genuine variational technique for minimizing the total cost with respect to the continuity control function vector w. In fact it does so for a couple of reasons.

First of all, calculus of variations with respect to w requires, that the space W of admissible continuity control functions is embedded in some topological vector space. Any continuity control function, which can be separated from the set of all strictly positive continuity control functions by this topology, that is any continuity control function, which represents an essential set of edges according to (1), will necessarily belong to the boundary of W. Hence the continuity control function vectors of particular interest, that is those representing edges, can be optimal, without being critical, that is, without resulting in a zero variation of the total cost with respect to w.

Secondly, if the variation of the total cost with respect to w is set to zero, one obtains the ridiculous condition

$$\sum_{k_1=1}^{K} \cdots \sum_{k_i=1}^{K} \left(\frac{\partial^i z}{\partial x_{k_1} \cdots \partial x_{k_i}} \right)^2 \equiv \lambda_i \qquad i = 1, \ldots, I$$

under which the total cost is completely independent of w. Thus the optimal continuity control function vector can not be found by means of calculus of variations, even if it does not represent an essential set of edges. Terzopoulos resolves this problem by first discretizing the entire space of continuity control functions; w is defined on a finite subset D — a dual pixel grid — and only allowed to take the values 0 or 1. The edge cost is modified accordingly to

$$\mathcal{E}(w) \doteq \sum_{x \in D} \sum_{i=1}^{I} \lambda_i [1 - w_i(x)]$$

For a solution he then applies a descent method in the continuity control function vector space W^I. Prior to each update of w, the optimal estimated image function $\check{z}_w$ for the present w is computed, by solving the Euler equation — a partial differential equation in $\check{z}_w$ — associated with the variational principle $\delta_z \mathcal{C}(w, z) = 0$. This method is expensive, and since the update $\blacktriangle w$ is based on the cost difference $\mathcal{C}(w + \blacktriangle w, \check{z}_w) - \mathcal{C}(w, \check{z}_w)$, as opposed to $\mathcal{C}(w + \blacktriangle w, \check{z}_{w+\blacktriangle w}) - \mathcal{C}(w, \check{z}_w)$ — computation of $\check{z}_{w+\blacktriangle w}$ for all possible updates $\blacktriangle w$ would be far too expensive — convergence to a global minimum cannot be guaranteed.

3 Genuinely Variational Edge Detection

For our problem of detecting discontinuities of a bivariate image function, the appropriate deviation and stabilizing costs in the paradigm above are given by

$$\mathcal{D}(z, \zeta) \doteq \int_B (z - \zeta)^2 \, dx$$

and

$$\mathcal{S}(w, z) \doteq \int_B w \|\nabla z^T\|^2 \, dx$$

In order to remedy the difficulties with Terzopoulos' method, we propose the use of a smooth continuity control function $w : B \to \overline{\mathbb{R}_+}$. If w was prespecified, this would amount to the simplest straight forward generalization of Tikhonov stabilization to bivariate regularization. However, as Terzopoulos we will consider w to be a variable, and optimaize the total

cost with respect to both w and z. To avoid the problem with optimal continuity control functions, which are non-critical, and thus impossible to find by means of variational calculus, we will arrange the edge cost, so that for each estimated image function z, the total cost $\mathcal{C}(w,z)$ attains its minimum for exactly one optimal continuity control function $\check{w}_z$, whose range is confined to lie in $]0,1]$. This idea is similar to the use of barrier functions in finite dimensional optimaization [25]. The uniqueness of $\check{w}_z$ for a given z, also allows us to solve for $\check{w}_z$ in terms of z in a way similar to Blake and Zisserman's elimination of their "line process" [12,13]. The edge costs, we propose for this purpose, are of the form

$$\mathcal{E}(w) \doteq \int_B \lambda f \circ w \, dx$$

where the *edge cost coefficient* $\lambda > 0$ is constant, and the *edge cost density function* $f : \mathbf{R}_+ \to \mathbf{R}$ is twice differentiable. Our total cost functional is thus given by

$$\mathcal{C}(w,z) \doteq \int_B [\lambda f \circ w + (z-\zeta)^2 + w\|\nabla z^T\|^2] \, dx \tag{2}$$

It would be appropriate to multiply the stabilizing cost $\mathcal{S}(w,z)$ by the square of a (constant) scale-space parameter $\mu > 0$. However a true magnification of the scale of resolution of the edge detector should be equivalent to a shrinkage of the width and height of the image domain (along with the induced space scaling of the functions defined thereon) by the same factor. For any consistent discretization of the problem the effective scale-space parameter will therefore be inversely proportional to, and might as well be absorbed in, the pixel width h.

Setting the first variation of $\mathcal{C}(w,z)$ to zero yields the Euler equations

$$z(x) - \zeta(x) - \nabla \cdot (w \nabla z)(x) = 0 \qquad \forall x \in B \tag{3a}$$

$$\lambda f'(w(x)) + \|\nabla z(x)^T\|^2 = 0 \qquad \forall x \in B \tag{3b}$$

$$w(x) \frac{\partial z}{\partial e_n}(x) = 0 \qquad \forall x \in \partial B \tag{3c}$$

where $\nabla\cdot$ denotes the divergence operator, and $\partial/\partial e_n$ denotes the directional derivative in the direction of the outward normal. The second variation of $\mathcal{C}$ with respect to w is also easily found to be

$$\delta^2_{ww}\mathcal{C}(w,z) = \int_B \frac{\lambda}{2}(f'' \circ w)(\delta w)^2 \, dx \tag{4}$$

Together with the desired existence of a unique optimal continuity control function $\check{w}_z$ for each possible estimated image function z these equations put some restrictions on the edge cost density f. In fact from (3b) it follows, that $f'|]0,1] \to \overline{\mathbf{R}_-}$ must be bijective, and that $f'(]1,\infty[) \subseteq \mathbf{R}_+$. Likewise from (4) we see, that f'' must be strictly positive on $]0,1[$, and that $f''(1) \geq 0$. the simplest functions, which satisfy these requirements are given by

$$f(\omega) \doteq \omega - \ln\omega \qquad \Rightarrow \qquad f'(\omega) = 1 - \frac{1}{\omega} \tag{5}$$

and

$$f(\omega) \doteq \omega \ln\omega - \omega \qquad \Rightarrow \qquad f'(\omega) = \ln\omega \tag{6}$$

but there are of course many other possibilities, for example:

$$f(\omega) \doteq \omega + \frac{1}{(p-1)\omega^{p-1}} \qquad \Rightarrow \qquad f'(\omega) = 1 - \frac{1}{\omega^p} \qquad p \in \mathbf{R}_+ \setminus \{1\} \tag{7}$$

However, some choices of p might be better than others. In section 6 we will present an argument supporting the further restriction, that $p < 2$.

Given that f satisfies these conditions, $f'|]0,1]$ is invertible, and since w is strictly positive, we end up with the equations

$$z(x) = \zeta(x) + \nabla \cdot (w \nabla z)(x) \qquad \forall x \in B \tag{8a}$$

$$w(x) = g(\|\nabla z(x)^T\|) \qquad \forall x \in B \tag{8b}$$

$$\frac{\partial z}{\partial e_n}(x) = 0 \qquad \forall x \in \partial B \tag{8c}$$

where the function $g : \overline{\mathbf{R}_+} \rightarrow]0,1]$, (for reasons soon to make sense,) referred to as the *diffusivity anomaly*, is defined by

$$g(\gamma) \doteq (f'|]0,1])^{-1}\left(-\frac{\gamma^2}{\lambda}\right) \qquad \gamma \geq 0 \tag{9}$$

The properties of the edge cost density f clearly imply, that g is a strictly positive strictly decreasing differentiable bijection. In particular $g(0) = 1$, and $\lim_{\gamma \rightarrow \infty} g(\gamma) = 0$. For the edge cost density in (5) the diffusivity anomaly takes the form

$$g(\gamma) \doteq \frac{1}{1 + \frac{\gamma^2}{\lambda}} \qquad \gamma \geq 0 \tag{10}$$

while the edge cost density in (6) yields

$$g(\gamma) \doteq e^{-\frac{\gamma^2}{\lambda}} \qquad \gamma \geq 0 \tag{11}$$

Since our method necessarily yields continuity control functions, for which

$$w^{-1}(\{0\}) = \varnothing$$

Terzopoulos' edge representation is inadequate. The simplest and most reasonable modification is to consider the edges to consist of the set $w^{-1}(]0,\theta])$, where θ is a fixed threshold. Since the diffusivity anomaly g is strictly decreasing, we have

$$w^{-1}(]0,\theta]) = \|\nabla z^T\|^{-1}([g^{-1}(\theta), \infty[)$$

whence the edges are obtained by thresholding the magnitude of the gradient of the estimated image function.

Other possibilities are of course possible. One could for example attempt to detect various desired edge patterns by filtering w. One could also try to make the threshold adaptive, and/or let it depend on the position x in some clever way. With attempts along these lines however, one will most likely tend to stray away from the original optimality principle, and end up in the kind of hacker's nest, the introduction of such a principle was initially meant to avoid.

4 Biased Anisotropic Diffusion

Perona and Malik [16,18] have introduced anisotropic diffusion as a method of suppressing finer details, without weakening or dislocating the larger scale edges. The initial value

problem governing their method is given by

$$\frac{\partial z}{\partial t}(x,t) = \nabla \cdot (w \nabla z)(x,t) \qquad \forall x \in B \qquad \forall t > 0 \tag{12a}$$

$$w(x,t) = g(\|\nabla z(x,t)^T\|) \qquad \forall x \in B \qquad \forall t > 0 \tag{12b}$$

$$\frac{\partial z}{\partial e_n}(x,t) = 0 \qquad \forall x \in \partial B \qquad \forall t > 0 \tag{12c}$$

$$z(x,0) = \zeta(x) \qquad \forall x \in B \tag{12d}$$

where the diffused image function z and the *diffusivity* w are functions of both position $x \in B$ and time $t \geq 0$, $\nabla\cdot$ and ∇ denote the divergence and the gradient respectively with respect to x, and the diffusivity anomaly $g : \overline{\mathbb{R}_+} \to \overline{\mathbb{R}_+}$ is a decreasing function.

As the name "anisotropic diffusion" suggests, these equations have appealing physical interpretations. The function z can for example be thought of as representing the temperature T in a thin slab S of a material, whose initial temperature is given by ζ, and whose space- and time-varying thermal diffusivity (or thermal conductivity, if time is scaled appropriately,) is given by w. This analogy is depicted in figure 1.

The Euler equations we derived in the previous section are very similar to the initial value problem (12). In fact a solution of (8) is given by the steady state of the initial value problem

$$\frac{\partial z}{\partial t}(x,t) = \zeta(x,t) - z(x,t) + \nabla \cdot (w \nabla z)(x,t) \qquad \forall x \in B \qquad \forall t > 0 \tag{13a}$$

$$w(x,t) = g(\|\nabla z(x,t)^T\|) \qquad \forall x \in B \qquad \forall t > 0 \tag{13b}$$

$$\frac{\partial z}{\partial e_n}(x,t) = 0 \qquad \forall x \in \partial B \qquad \forall t > 0 \tag{13c}$$

$$z(x,0) = \chi(x) \qquad \forall x \in B \tag{13d}$$

which is obtained from (12) by replacing the anisotropic diffusion equation (12a) by the closely related *"biased" anisotropic diffusion* equation (13a). Since our interest is in the steady state solution, the initial condition (12d) can also be replaced by an arbitrary initial condition (13d). The continuity control function w thus plays the role of the diffusivity, and will be referred to as such, whenever the context so suggests.

The bias term $\zeta - z$ intuitively has the effect of locally moderating the diffusion as the diffused image function z diffuses further away from the original image function ζ. It is therefore reasonable to believe, that a steady state solution does exist. The following physical interpretation of this initial value problem further substantiates this belief: Let S be a thin slab of some material resting on top of another slab S_0 of some (other) material as in figure 2. Suppose that the space- and time-varying thermal conductivity of S is given by αw, where the constant $\alpha > 0$ is the coefficient of heat transfer between S and S_0. If the initial temperature at each point $x \in B$ of S is given by $\chi(x)$, and the temperature distribution of S_0 is held fixed at ζ, then z represents the space- and time-varying temperature of S.

The possibility of suppressing finer details, while the more significant edges remain intact, or are even strengthened, is a consequence of the anisotropy, which in both the diffusions described above in turn is caused by the non-constancy of the diffusivity anomaly g. If g is constant, the *unbiased* diffusion (12a) reduces to Gaussian blurring, while the steady state of the biased diffusion (13a) in a sense corresponds to filtering with a doubly cascaded first order Butterworth filter. For our variational method governed, by the boundary value problem (8), the choice of g was based on optimality considerations. Perona and Malik select

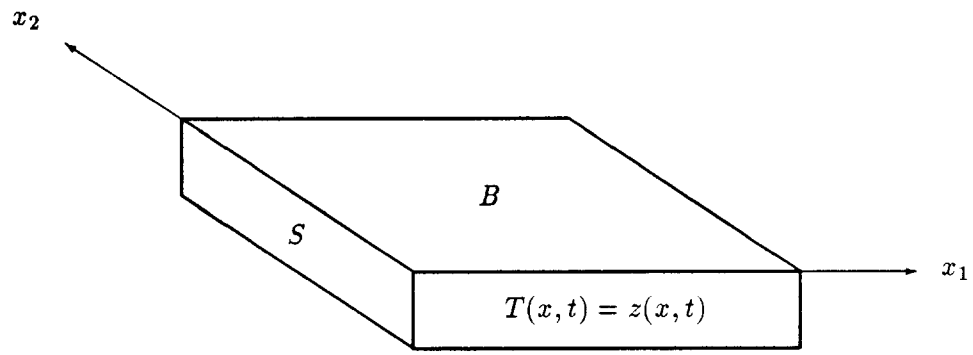


Figure 1: Physical model of unbiased anisotropic diffusion.

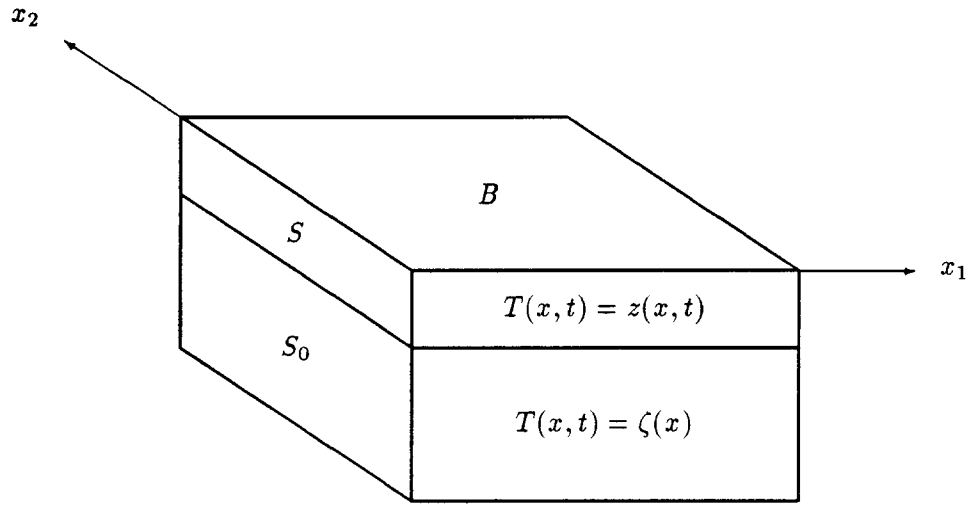


Figure 2: Physical model of biased anisotropic diffusion.

their function g, by demanding, that the resulting unbiased anisotropic diffusion enhances the already pronounced edges, while the less significant edges are weakened. Based on an analysis including only blurred linear step edges — an unnecessary restriction, as we shortly shall see — they vouch for diffusivity anomalies of the form

$$g(\gamma) \doteq \frac{c}{1 + \left(\frac{\gamma^2}{\lambda}\right)^a} \tag{14}$$

where $c, \lambda > 0$ and $a > 1/2$ are constants. It is easy to check, that, if these functions were substituted in the Euler equation (8b), the corresponding edge cost densities would satisfy the requirements of our variational method. (To be precise, the constant c would actually have to be equal to unity. This is however an artifact, which would not have surfaced, had we incorporated the scale-space parameter μ, and vanishes regardless in the discretization process.) Incidentally, for their experimental results, Perona and Malik use exactly the functions, we proposed in (10) and (11), of which only the former belongs to the class specified by (14).

Finally we note, that the heuristically motivated method, that Perona and Malik used for extracting a set of edges from the diffused image function, is practically identical to the method, implied by our edge representation in terms of the continuity control function. While they threshold the absolute difference between four-connected neighbor pixel values, our edge representation leads, as we saw in the previous section, to thresholding of the magnitude of the gradient.

5 The Extremum Principle

The extremum principle is a common tool for proving uniqueness and stability with respect to boundary data for linear elliptic and linear parabolic problems [26]. For quasi-linear equations, such as the Euler equation (8a) and the biased anisotropic diffusion equation (13a), it is not quite as conclusive. Nevertheless it provides bounds on the solution and useful insight for convergence analysis of the numerical methods employed to find such a solution. We will present an extremum principle for the biased anisotropic diffusion problem (13) as well as for the boundary value problem (8). In both cases we will assume, that the diffusivity anomaly $g : \overline{\mathbb{R}_+} \to \overline{\mathbb{R}_+}$ is continuously differentiable.

Theorem 5.1 *Let $z : \overline{B} \times \overline{\mathbb{R}_+} \to \mathbb{R}$ be a solution of the biased anisotropic diffusion problem (13), where it is assumed, that $\zeta : B \to \mathbb{R}$ is uniformly continuous. Assume further, that z and its first and second order partial derivatives with respect to x are continuous (on $\overline{B} \times \overline{\mathbb{R}_+}$). Then the following claims are true:*

(i) *If $\pm y_\tau : \overline{B} \to \mathbb{R} : x \mapsto \pm z(x, \tau)$ has a local maximum at $\xi \in \overline{B}$ for some fixed $\tau > 0$, then*

$$\pm \frac{\partial z}{\partial t}(\xi, \tau) \leq \pm \zeta(\xi) \mp z(\xi, \tau)$$

(ii) *If $\pm z$ has a local maximum at $(\xi, \tau) \in \overline{B} \times \mathbb{R}_+$, then*

$$\pm z(\xi, \tau) \leq \pm \zeta(\xi)$$

(iii) $$\inf_{\xi \in B} [\zeta(\xi) \wedge \chi(\xi)] \leq z(x,t) \leq \sup_{\xi \in B} [\zeta(\xi) \vee \chi(\xi)] \qquad \forall x \in \overline{B} \qquad \forall t \geq 0$$

Proof: From (13a) and the continuity assumptions regarding g, ζ and z one can show, that $\partial z/\partial t$ is uniformly continuous on $B \times T$ for every bounded interval $T \subseteq \mathbb{R}_+$, and therefore has a unique continuous extension to $\overline{B} \times \mathbb{R}_+$. By the bounded convergence theorem of integration it also follows, that this extension equals $\partial z/\partial t$ on the boundary $\partial B \times \mathbb{R}_+$. Hence (13a) is satisfied on all of $\overline{B} \times \mathbb{R}_+$ (with the appropriate one-sided derivatives on $\partial B \times \mathbb{R}_+$). Suppose that $\pm y_\tau$ has a local maximum at $\xi \in \overline{B}$ for some $\tau > 0$. Then by Taylor's formula (and the Neumann condition (13c), if $\xi \in \partial B$) we have, that $\nabla y_\tau(\xi) = 0$, and $\pm\Delta y_\tau(\xi) \leq 0$. Thus

$$\pm\nabla \cdot (w\nabla z)(\xi,\tau) = \pm\nabla w(\xi,\tau)\nabla y_\tau(\xi)^T \pm w(\xi,\tau)\Delta y_\tau(\xi) \leq 0$$

whence (i) follows. Suppose next, that $\pm z$ has a local maximum at $(\xi,\tau) \in \overline{B} \times \mathbb{R}_+$. Then $\pm y_\tau$ has a local maximum at ξ, and $\partial z/\partial t(\xi,\tau) = 0$. Hence (ii) follows from (i). Finally consider the compact set $\overline{B} \times [0,T_1]$, on which the continuous functions $\pm z$ attain their maximal values, say at $(\xi_\pm,\tau_\pm)$. If $\tau_\pm = T_1$, then $\pm\partial z/\partial t(\xi_\pm,\tau_\pm) \geq 0$, and $\pm y_{\tau_\pm}$ has a local maximum at $\xi_\pm$. Hence (i) implies, that $\pm z(\xi_\pm,\tau_\pm) \leq \pm\zeta(\xi_\pm)$. If $\tau_\pm \in]0,T_1[$, the same conclusion follows immediately from (ii). Since $T_1 > 0$ was arbitrarily chosen, this shows, that

$$\sup_{(x,t)\in\overline{B}\times\mathbb{R}_+} \pm z(x,t) \leq \sup_{x\in B} \pm\zeta(x)$$

from which (iii) follows. ■

For the boundary value problem (8), governing our variational edge detection method, a proof similar to that above yields the following extremum principle:

Theorem 5.2 *Let $z : \overline{B} \to \mathbb{R}$ be a solution of the boundary value problem (8). Assume further, that z and its first and second order partial derivatives are continuous (on $\overline{B}$). Then*

$$\inf_{\xi\in B} \zeta(\xi) \leq z(x) \leq \sup_{\xi\in B} \zeta(\xi) \qquad \forall x \in \overline{B}$$

We remark, that in both the theorems above, the assumption, that the derivatives of z are continuous up to *and including* the boundary ∂B ($\times\overline{\mathbb{R}_+}$), (or equivalently uniformly continuous on every bounded subset of the interior of the domain of z,) could have been traded for a weaker bound on z, which in addition to the values of z on B ($\times\{0\}$) also includes those on ∂B ($\times\overline{\mathbb{R}_+}$). However, for the discretized problem, that we eventually will have to solve, the subtle difference between plain vs. uniform continuity of z and its derivatives is of no consequence. The "stronger-assumption-conclusion" versions of the extremum principles presented above are therefore more useful in this context.

According to the two theorems above the solutions of the biased anisotropic diffusion problem are well-behaved, in that they do not stray too far away from the original image function ζ, unless forced to by the initial condition, and even if so, they eventually approach the range of ζ as $t \to \infty$. In plain language condition (i) of the first theorem says, that the diffused image process, at each of its momentary critical points (with respect to x) is headed towards the original image function. Condition (ii) of the same theorem says, that all the non-initial local extrema of the diffused image process are within the range of the original image function, and condition (iii) gives explicit bounds on the entire collection of diffused image functions in terms of the initial and original image functions. The second theorem bounds the steady state diffused image function in terms of the original data alone. In other words, our variational edge detection method produces an estimated image function, whose range is contained inside that of the original image function.

6 Edge Enhancement

It was mentioned earlier, that the biased anisotropic diffusion (13), in similarity with its unbiased counterpart (12), has the important property of suppressing finer details, while strengthening the more significant edges. Indeed, the edges are roughly either sharpened or blurred depending on their present strength, viz. the magnitude of the gradient of the diffused image function z.

In order to see this, we define the edges to consist of the points in the image domain B, at which the magnitude of the gradient of the diffused image function has a strict local maximum along the direction perpendicular to the edge, that is the direction of the gradient. For simpler notation we let $\sigma \doteq \|\nabla z^T\|$. We also define e_ν and e_τ to be the unit vectors in the directions of $[\partial z/\partial x_1 \quad \partial z/\partial x_2]$ and $[\partial z/\partial x_2 \quad -\partial z/\partial x_1]$ respectively, that is e_ν is normal, and e_τ is tangential to the edge. Since $\sigma > 0$ on the edges, e_ν and e_τ are well-defined on the points of interest. The edge points can now be characterized by:

$$\frac{\partial \sigma}{\partial e_\nu} = 0 \tag{15a}$$

$$\frac{\partial^2 \sigma}{\partial e_\nu^2} < 0 \tag{15b}$$

For a typical edge of interest it is reasonable to assume, that its strength σ exhibits a fairly pronounced peak along its *perpendicular* direction, resulting in a large value of $|\partial^2\sigma/\partial e_\nu^2|$. On the other hand σ can be expected to vary quite moderately — at most with a fairly constant derivative (shading component) — *along* the edge, with a small value of $|\partial^2\sigma/\partial e_\tau^2|$ as a consequence. We will therefore at little loss allow ourselves to restrict attention to edge points, at which

$$\frac{\partial^2 \sigma}{\partial e_\nu^2} \approx \Delta\sigma < 0 \tag{16}$$

Our discussion includes in particular all symmetrically blurred (smooth) step edges. For points on such edges the approximation (16) is indeed exact, even if the size of the step varies linearly with arc length along the edge.

We begin by noting, that

$$\nabla\sigma\nabla z^T = \frac{\partial \sigma}{\partial e_\nu}\sigma$$

and

$$\frac{\partial z}{\partial e_\nu} = \sigma \tag{17}$$

Assuming that all functions involved are sufficiently smooth, and that the diffusivity is of the usual form $w = g \circ \sigma$, from (15a) and (16) we then have.

$$\begin{aligned}
&\frac{\partial}{\partial e_\nu}\nabla\cdot(w\nabla z) \\
&\quad = \frac{\partial}{\partial e_\nu}(\nabla w \nabla z^T + w\Delta z) \\
&\quad = \frac{\partial}{\partial e_\nu}[(g'\circ\sigma)\nabla\sigma\nabla z^T] + \frac{\partial w}{\partial e_\nu}\Delta z + w\frac{\partial}{\partial e_\nu}\Delta z \\
&\quad = \frac{\partial}{\partial e_\nu}\left[(g'\circ\sigma)\frac{\partial\sigma}{\partial e_\nu}\sigma\right] + (g'\circ\sigma)\frac{\partial\sigma}{\partial e_\nu}\Delta z + w\Delta\frac{\partial z}{\partial e_\nu}
\end{aligned}$$

$$= (g' \circ \sigma)\frac{\partial^2 \sigma}{\partial e_\nu^2}\sigma + (g \circ \sigma)\Delta\sigma$$
$$\approx [(g' \circ \sigma)\sigma + g \circ \sigma]\Delta\sigma$$

From (17) it also follows that

$$\frac{\partial \sigma}{\partial t} = \frac{\partial}{\partial e_\nu}\frac{\partial z}{\partial t}$$

Hence on the edges, the biased anisotropic diffusion (13) causes the edge strength to vary with time according to

$$\frac{\partial \sigma}{\partial t} \approx \frac{\partial \zeta}{\partial e_\nu} - \frac{\partial z}{\partial e_\nu} + (\varphi' \circ \sigma)\Delta\sigma$$

where

$$\varphi(\gamma) \doteq g(\gamma)\gamma \qquad \gamma \geq 0 \tag{18}$$

Rewriting this equation as

$$\frac{\partial}{\partial t}(\sigma - \sigma_\zeta) \approx -(\sigma - \sigma_\zeta) + (\varphi' \circ \sigma)\Delta\sigma \tag{19}$$

where $\sigma_\zeta \doteq \partial\zeta/\partial e_\nu$, it is clear, that the bias term $-(\sigma - \sigma_\zeta)$ merely has a moderating effect on the enhancement/blurring of the edge, while the decision between enhancement vs. blurring depends on the sign of the "driving" term $(\varphi' \circ \sigma)\Delta\sigma$ associated with the unbiased anisotropic diffusion.

For the desired performance of weakening the weak edges, while strengthening the strong ones in a consistent manner, since $\Delta\sigma < 0$, it is therefore necessary, that there exists an *edge enhancement threshold* $\gamma_0 \in \mathbf{R}_+$, such that

$$\varphi'^{-1}(\mathbf{R}_-) =]\gamma_0, \infty[\tag{20a}$$
$$\varphi'^{-1}(\{0\}) = \{\gamma_0\} \tag{20b}$$
$$\varphi'^{-1}(\mathbf{R}_+) = [0, \gamma_0[\tag{20c}$$

Furthermore, if so, the threshold γ_0 clearly controls the sensitivity of the edge detector, and one would hence expect it to be closely related to the intuitively similarly acting edge cost coefficient λ. Indeed from (9) and (18) it immediately follows, that $\varphi'(\gamma)$ is a function of γ^2/λ. Since $\overline{\mathbf{R}_+} \to \overline{\mathbf{R}_+} : \gamma \mapsto \gamma^2/\lambda$ is strictly monotone, γ_0 must therefore be proportional to $\sqrt{\lambda}$. It is easy to verify, that the diffusivity anomalies given in (10) and (11) satisfy (20) with $\gamma_0 = \sqrt{\lambda}$ and $\gamma_0 = \sqrt{\lambda/2}$ respectively. For the diffusivity anomalies corresponding to the edge cost densities in (7) on the other hand, an edge enhancement threshold $\gamma_0 = \sqrt{p\lambda/(2-p)}$ satisfying (20) will exist if and only if $p \in]0,1[\cup]1,2[$.

Although the discussion above generates some useful insight, and offers guidelines for sensible choices of the diffusivity anomaly g, it is not completely satisfactory, in that it does not account for the change in location and orientation of the edges in the image domain during the diffusion process. In fact, by evaluating the second partial derivative $\partial^2\sigma/\partial t\partial e_\nu$, one can show, that only edges with certain symmetry properties, for example symmetrically blurred linear step edges, will remain fixed in position, during the diffusion. If one neglect this weakness — a forgotten subject in previous papers — one could be misled to believe, that the enhancement/blurring decisions about the edges are completely determined by the local properties of the original image function ζ at the edge points. If this was true, one could just as well detect the edges, by checking these properties, amounting to nothing more,

than thresholding the directional derivative of ζ in the direction of its gradient at points, where this derivative has local maxima, — a previously explored paradigm in local edge detection [27].

If the diffused image function converges to a steady state solution, that is a solution of the boundary value problem (8), the edge enhancement/blurring is of course in the limit independent of the initial condition (13d). Indeed from (19) we immediately obtain the steady state edge enhancement

$$\sigma - \sigma_\zeta = (\varphi' \circ \sigma)\Delta\sigma \tag{21}$$

This equation is clearly valid independently of how the edges move during the diffusion process. On the other hand σ_ζ is not representative of the original edge strength $\|\nabla\zeta^T\|$, if ∇z and $\nabla\zeta$ differ too much in orientation.

Since the range of the steady state solution, by the extremum principle, is confined to lie within the range of the original image function, an amply enhanced edge strength σ can only be maintained along a very short distance across the edge. Such edges are therefore sharpened.

For the numerical solution of the boundary value problem (8) on a regular computer there are, as we shall shortly discuss, good reasons for updating the estimated image function according to a rule, different from a straight forward discretization of the biased anisotropic diffusion equation. However, the final edge enhancement (21) is independent of the path to the solution, so the discussion above is still valid.

Besides being of vital importance for the edge enhancement mechanism, the existence of the edge enhancement threshold γ_0 also provides a natural choice for the threshold to be used in the postprocessing, whereby the edges are finally extracted from the estimated image function. It is intuitively clear, that, for our edge representation to be consistent with the edge enhancement mechanism, the edge representation threshold in section 3 should be given by $\theta \doteq g(\gamma_0)$. The edge set $w^{-1}(]0,\theta])$ will then consist of the points in the image domain, where the magnitude of the gradient of the estimated image function exceeds γ_0, that is exactly those points, where the edge strength has been enhanced. On the other hand, and this is in a sense the essential benefit with our regularization approach, the bistability of the edge enhancement mechanism will deplete the set of points, at which the gradient magnitude of the estimated image function takes values close to γ_0. The edge set $w^{-1}(]0,\theta])$ will therefore be almost indifferent to changes in θ, as long as θ belongs to some substantial neighborhood of $g(\gamma_0)$. These circumstances are clearly ideal for thresholding, and consequently our edge representation is practically consistent with the edge enhancement mechanism for a whole interval of edge representation thresholds, corresponding to a relatively wide range of gradient magnitudes.

7 Discretization

For a numerical solution of the variational edge detection problem in section 3 the boundary value problem (8) has to be discretized. The original image function ζ is most likely already given only on a squared pixel grid. Assuming that this is the case, the simplest way of discretizing the image functions z and ζ for the numerical problem, is obviously to use the same grid. For the evaluation of the expression $\nabla\cdot(w\nabla z)$ there are on the contrary a number of more or less equally sensible choices. One can for example expand $\nabla\cdot(w\nabla z)$ in terms of

z and its derivatives according to

$$\nabla \cdot (w \nabla z) = (g' \circ \sigma) \frac{1}{\sigma} \nabla z \, \mathbf{H} z \, \nabla z^T + (g \circ \sigma) \Delta z$$

where $\sigma \doteq \|\nabla z^T\|$, and $\mathbf{H}$ denotes the Hessian operator. With the numerous discrete approximations of ∇z and $\mathbf{H} z$ at hand, this leaves a multitude of open possibilities. Alternatively one can treat $w \nabla z$ as a single function, readily evaluated at appropriate points in terms of some discrete approximation of ∇z, and then take some discrete approximation of its divergence. We have settled for the latter approach, which has the special quality of highlighting the diffusion mechanism. This in turn naturally leads to expressions, which are particularly convenient for the analysis of the resulting algorithms.

To be more specific, let us consider an original image function ζ, given on a grid $\{jh : j \in J\}$, where $h > 0$ is the pixel width, and $J \doteq \{1, \ldots, J_1\} \times \{1, \ldots, J_2\}$ for some $J_1, J_2 \in \mathbb{N}$. The corresponding image domain is thus given by $B \doteq]\frac{1}{2}, J_1 + \frac{1}{2}[\times]\frac{1}{2}, J_2 + \frac{1}{2}[$. We then define the discretized shifted functions

$$\begin{array}{lcll} \zeta_0(j) & \doteq & \zeta(jh) & j \in J \\ z_q(j) & \doteq & z(\arg\min_{l \in J} \|l - j - q\| \cdot h) & j \in J \quad q \in S \\ w_q(j) & \doteq & g(\sigma_q(j)) & j \in J \quad q \in S \end{array}$$

where $S \doteq \{-1, 0, 1\}^2$, and $\sigma_q(j)$ is some discrete approximation of $\|\nabla z((j + \frac{q}{2})h)^T\|$. It is reasonable to demand, that $\sigma_q(j)$ be specified in terms of z at the smallest possible symmetric set of neighboring grid points of $(j + \frac{q}{2})h$. This requirement leads to the discrete approximations:

$$\sigma_q^2 \doteq \frac{(z_{q_1,0} - z_{0,1} + z_{q_1,-1} - z_{0,0})^2 + (z_{q_1,0} - z_{0,-1} + z_{q_1,1} - z_{0,0})^2}{8h^2} \qquad q_1 = \pm 1 \qquad (22a)$$

$$\sigma_q^2 \doteq \frac{(z_{0,q_2} - z_{1,0} + z_{-1,q_2} - z_{0,0})^2 + (z_{0,q_2} - z_{-1,0} + z_{1,q_2} - z_{0,0})^2}{8h^2} \qquad q_2 = \pm 1 \qquad (22b)$$

$$\sigma_q^2 \doteq \frac{(z_{q_1,q_2} - z_{0,0})^2 + (z_{q_1,0} - z_{0,q_2})^2}{2h^2} \qquad q_1, q_2 \in \{-1, 1\} \qquad (22c)$$

where we have dropped the dependence of $j \in J$ for shorter notation, and written z_{q_1,q_2} for z_q. The two discrete approximations of (8), which immediately come to mind, can after some manipulation (from a variety of starting points) be written as

$$\zeta_0 - z_0 + \frac{1}{\rho^2 h^2} \sum_{q \in S_\rho} w_q (z_q - z_0) = 0 \qquad \rho^2 = 1, 2 \qquad (23)$$

where $S_\rho \doteq \{q \in S : \|q\| = \rho\}$, $\rho^2 = 1, 2$. Note, that the Neumann condition (8c) is conveniently taken care of by the "arg min"-adjustment in (22b), which systematically replaces any otherwise required value of z at a grid point outside B, by the value of z at the closest grid point inside B, thereby ensuring that

$$\left.\begin{array}{l} z_{-1,q_2}(1, j_2) - z_{0,q_2}(1, j_2) = 0 \\ z_{1,q_2}(J_1, j_2) - z_{0,q_2}(J_1, j_2) = 0 \end{array}\right\} \qquad j_2 = 1, \ldots, J_2 \qquad q_2 = -1, 0, 1$$

$$\left.\begin{array}{l} z_{q_1,-1}(j_1, 1) - z_{q_1,0}(j_1, 1) = 0 \\ z_{q_1,1}(j_1, J_2) - z_{q_1,0}(j_1, J_2) = 0 \end{array}\right\} \qquad j_1 = 1, \ldots, J_1 \qquad q_1 = -1, 0, 1$$

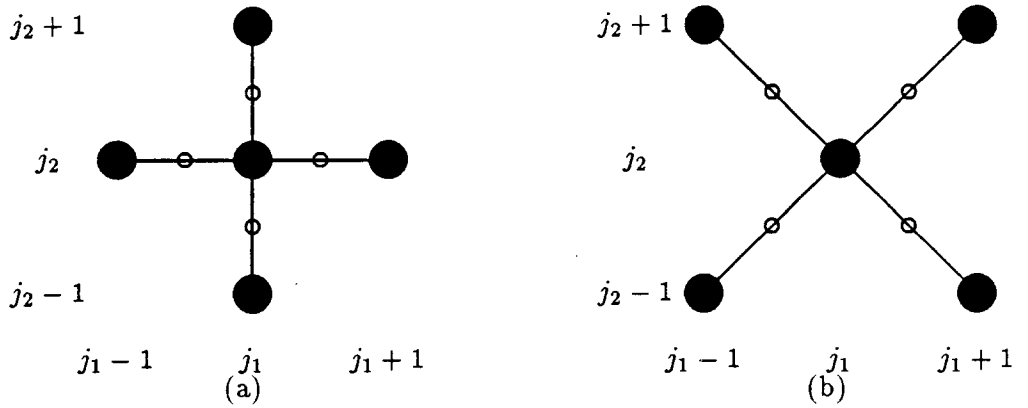


Figure 3: Discrete approximation molecule structures. (a) "Cartesian"; $\rho^2 = 1$. (b) "Diagonal"; $\rho^2 = 2$.

The computational molecules associated with the two approximations, $\rho^2 = 1, 2$, in (23) have the structures depicted in figure 3, where the filled circles (atoms) mark the sites associated with the evaluation of ζ and z, and the empty circles (bond centers) mark the sites associated with the evaluation of w. In each case the sum involved contains four terms.

The "Cartesian" approximation has a couple of apparent advantages. First of all it provides tight coupling between all pairs of eight-connected pixel neighbors. In contrast, as one can see from figure 3, the "diagonal" approximation results in two interleaved but separated computational lattices. An algorithm based on this approximation therefore models two more or less separate diffusion processes, which are coupled only through the shared diffusivity function, that is the coefficient function(s) of the quasi-linear equation (23). For original image functions with mildly well-behaved statistics however, the smoothing effect of the diffusion will, as one would guess, and as our experimental results also indicate, cure this problem. A second minor advantage of the Cartesian approximation is, that its associated truncation error is only $\sqrt{2}/4$ times that of the diagonal approximation.

The diagonal approximation also has a couple of advantages: As figure 3 reveals, it requires only half as many evaluations of the control continuity function, as does the Cartesian approximation. In addition these evaluations are simpler, as they are governed by (22c) as opposed to (22a) and (22b) in the Cartesian case. The diagonal approximation thus leads to faster and simpler software implementations. Our experiments further show, that it, despite its drawbacks, yields excellent results.

There are several possible ways of solving the discretized equation (23) numerically. One method, which is obvious in the light of the discussion in the previous sections, is to propagate the corresponding discretized biased anisotropic diffusion equation

$$\begin{aligned} z_0^{(0)} &\doteq \chi_0 \\ z_0^{(i+1)} &\doteq z_0^{(i)} + k\left[\zeta_0 - z_0^{(i)} + \frac{1}{\rho^2 h^2}\sum_{q\in S_\rho} w_q^{(i)}(z_q^{(i)} - z_0^{(i)})\right] \qquad \rho^2 = 1, 2 \end{aligned}$$

where the *initial image function* $\chi_0 : J \to \mathbf{R}$ is arbitrary, most naturally chosen equal to ζ_0, $k > 0$ is the time step size, and $i \in \mathbf{N}_0$ is an iteration index, representing the time variable

t according to: $t = ik$. However this algorithm is numerically stable only for sufficiently small values of the step size k, and safe play will necessarily bring down the convergence rate. Since we are not interested in the diffusion per se, but merely its steady state solution, this problem can be avoided, by choosing some robuster iteration method. Such methods are easily generated by treating the quasi-linear equation (23) as a linear elliptic equation, and applying any of the commonly used Jacobi, Gauss-Seidel or successive over-relaxation methods. The Jacobi method for example yields the iteration scheme:

$$z_0^{(0)} \doteq \chi_0 \tag{24a}$$

$$z_0^{(i+1)} \doteq \frac{1}{\rho^2 h^2 + \overline{w}_\rho^{(i)}} \left[\rho^2 h^2 \zeta_0 + \sum_{q \in S_\rho} w_q^{(i)} z_q^{(i)} \right] \tag{24b}$$

$$\overline{w}_\rho^{(i)} \doteq \sum_{q \in S_\rho} w_q^{(i)} \tag{24c}$$

8 Convergence

In this section we will study some rudimentary convergence properties of the Jacobi-like iteration method (24). For certain parameter values we manage to show, that this iteration converges to a limit point, which satisfies (23), depends continuously on the original image function ζ_0, and is independent of the initial image function χ_0. Besides convergence of the iteration we thus obtain both uniqueness and a sense of stability with respect to the initial data. This sounds to good to be true, and as a matter of fact the assertions are valid only for parameter values, far from those of major interest for edge detection purposes. Albeit this serious weakness our analytical results give some indication, that solutions exist, and that these solutions are reasonably well-behaved, — a hypothesis further supported quite strongly by our experiments. They might also serve as a starting point for future theoretical development. One could possibly obtain better results, if one applied some more sophisticated iteration method. However this would most likely drastically compromise the simplicity of the algorithm. We have therefore confined our analysis to the Jacobi-like method, which after all yields remarkably satisfactory experimental results.

We begin our discussion with a couple of observations closely related to the extremum principles from section 5.

Proposition 8.1 *Let z_0 be a solution of the discretized boundary value problem (23). Then*

$$\bigwedge_{l \in J} \zeta_0(l) \leq z_0(j) \leq \bigvee_{l \in J} \zeta_0(l) \qquad \forall j \in J$$

Proof: Let $j_\pm \doteq \arg\max_{l \in J} \pm z_0(l)$. Then $\pm[z_q(j_\pm) - z_0(j_\pm)] \leq 0$, $\forall q \in S$. Hence by (23)

$$\bigvee_{l \in J} \pm z_0(l) = \pm z_0(j_\pm) \leq \pm \zeta_0(j_\pm) \leq \bigvee_{l \in J} \pm \zeta_0(l)$$

■

Proposition 8.2 *Let $z_0^{(i)}$, $i \in \mathbb{N}_0$ be defined by the iteration scheme (24). Then*

$$\bigwedge_{l \in J} [\zeta_0(l) \wedge \chi_0(l)] \leq z_0^{(i)}(j) \leq \bigvee_{l \in J} [\zeta_0(l) \vee \chi_0(l)] \qquad \forall j \in J \qquad \forall i \in \mathbb{N}_0$$

Proof: Let $i \in \mathbb{N}_0$, and $j \in J$. From (24a) and (24b) we see that $z_0^{(i)}(j)$ is a convex combination of $\zeta_0(j)$ and $z_q^{(i-1)}(j)$, $q \in S_\rho$, and thus in the convex hull of $\{\zeta_0(l), \quad z_0^{(i-1)}(l) : l \in J\}$. Since this is true $\forall j \in J$, we have

$$\bigwedge_{l \in J} [\zeta_0(l) \wedge z_0^{(i-1)}(l)] \leq \bigwedge_{l \in J} z_0^{(i)}(l) \leq \bigvee_{l \in J} z_0^{(i)}(l) \leq \bigvee l \in J [\zeta_0(l) \vee z_0^{(i-1)}(l)]$$

The proposition then follows by induction. ■

Using the bounds provided by the proposition above we can show the following two convergence results, of which proofs are given in appendix A.

Lemma 8.3 *Let $z_0^{(i)}$, $i \in \mathbb{N}_0$, be defined by the iteration scheme (24) in terms of an initial image function χ_0 and an original image function ζ_0. Let $y_0^{(i)}$, $i \in \mathbb{N}_0$, be defined in a completely analogous manner, but with χ_0 and ζ_0 replaced by ψ_0 and η_0 respectively. Assume that the dependency on the edge cost coefficient λ is reflected by the diffusivity anomaly g given by (10). If λ is sufficiently large, then*

$$\limsup_{i \to \infty} \|y_0^{(i)} - z_0^{(i)}\|_\infty \leq c\|\eta_0 - \zeta_0\|_\infty \qquad \textit{exponentially}$$

for some known finite constant c.

Theorem 8.4 *For sufficiently large values of the edge cost coefficient λ the discretized variational edge detection problem (23) has a unique solution, which is L_∞-norm-stable with respect to the initial data, and to which the iteration method (24) converges independently of its initial state χ_0.*

Unfortunately the theorem above gives a too pessimistic view of, what our experiments undoubtedly confirm, is really going on; it is only conclusive for values of the edge cost coefficient λ, far greater, than those, for which the algorithm is most useful for edge detection. There are two reasons for this shortcoming.

First of all at most locations $j \in J$ the constant R in the proof of lemma 8.3 (see appendix A) is an overly conservative bound for the *local* differences of $z_0^{(i)}$, it is meant to estimate. If the iteration scheme was linear, this problem could easily be remedied, by replacing the L_∞-norm in the convergence analysis by a Sobolev norm, which incorporates the evolution of these local differences as well as that of $z_0^{(i)}$ itself. However, as we discussed in section 6, the non-linearity, inherited from the boundary value problem (8), is by our choice such, that the local differences are strengthened, wherever initially sufficiently pronounced. Local differences of magnitude of the order R are therefore eventually to be expected. The intuitive reason for the success of the scheme lies in the earlier demonstrated fact, that the strengthened edges are simultaneously sharpened, so that the set of slow convergence shrinks during the iteration, — a mechanism, which is not captured by the L_∞-style of the

proof above. Since the non-linearity prohibits Fourier techniques, this problem might be hard to fix.

Secondly the theorem suggests, that λ be chosen proportional to h^4. In contrast, (as one would also guess from, the way λ enters the defining expressions of the diffusivity anomaly,) our experiments indicate, that λ be chosen proportional to h^2, as if the unity term inside the parenthesis in (25) (see appendix A) was missing. The intuitive reason for this discrepancy has to do with another case of competing processes. A closer examination of the proof above shows, that the source of this term is the unit bound on the continuity control function $w_q^{(i)}$, inherited from the properties of the diffusivity anomaly g. Since $w_q^{(i)}$ actually takes values close to unity at the abundant locations of almost vanishing image function gradient, this bound is tight. However rewriting (24b) as

$$z_0^{(i+1)} - z_0^{(i)} = \frac{1}{1 + \overline{w_2}^{(i)}} \left[2h^2(\zeta_0 - z_0^{(i)}) + \sum_{q \in S_2} w_q^{(i)}(z_q^{(i)} - z_0^{(i)}) \right]$$

we see, that at such locations

$$z_0^{(i+1)} - z_0^{(i)} \approx \frac{2h^2}{5}(\zeta_0 - z_0^{(i)})$$

Thus it seems, like the large values of $w_q^{(i)}$ destroy the exponential convergence rate locally, and only at those locations $j \in J$, where $z_0^{(i)}(j)$ has already practically converged.

9 Experimental Results

In this section we present some experimental results regarding our variational edge detection method, governed by the cost functional (2). In all the experiments the edge cost density was given by $f(\omega) \doteq \omega - \ln \omega$, corresponding to the diffusivity anomaly $g(\gamma) \doteq 1/(1+\gamma^2/\lambda)$. The images involved were obtained by solving the diagonal ($\rho^2 = 2$) discrete approximation (23) of the boundary value problem (8). For computational simplicity we used a Gauss-Seidel-like iteration method, rather than the Jacobi-like scheme (24).

As mentioned earlier, the iteration method converges to the solution of interest. In general, as one should expect, the convergence is faster, if the initial image function χ_0 is set equal to the original image function ζ_0. The sequence of images in figure 4 illustrates this condition. It shows, that reasonably good results are obtained well before 50 iterations, and that convergence in the "sense of insignificant perceptible changes" is reached after about 100 iterations. These observations, are as far as we can tell from our experiments, valid, whenever $\chi_0 = \zeta_0$. In particular, they seem to hold independently of the choice of the edge cost coefficient λ and the pixel width h, at least in the range of interest for edge detection.

The variational edge detection method itself as well as the iteration method, we employed to solve it, appear to be remarkably robust with respect to changes in the initial image function. Indeed if $\chi_0 \neq \zeta_0$, the iteration method still converges, if yet at a slower rate. To demonstrate this behavior, we tried the algorithm on the same original image function, as in figure 4, but with the particularly unfavorable initial image function $\chi_0 = 0$. Some samples from the resulting sequence of images are shown in figure 5. The fact that the limit image functions in the figures 4 and 5 are perceptually so close, also indicates, that the solutions, even though multiple, in large exhibit the desirable type of behavior, that mathematically stringent uniqueness would warrant. As one should expect, the significant differences seem

Fig. 4: (a)

Fig. 4: (b)

Fig. 4: (c)

Fig. 4: (d)

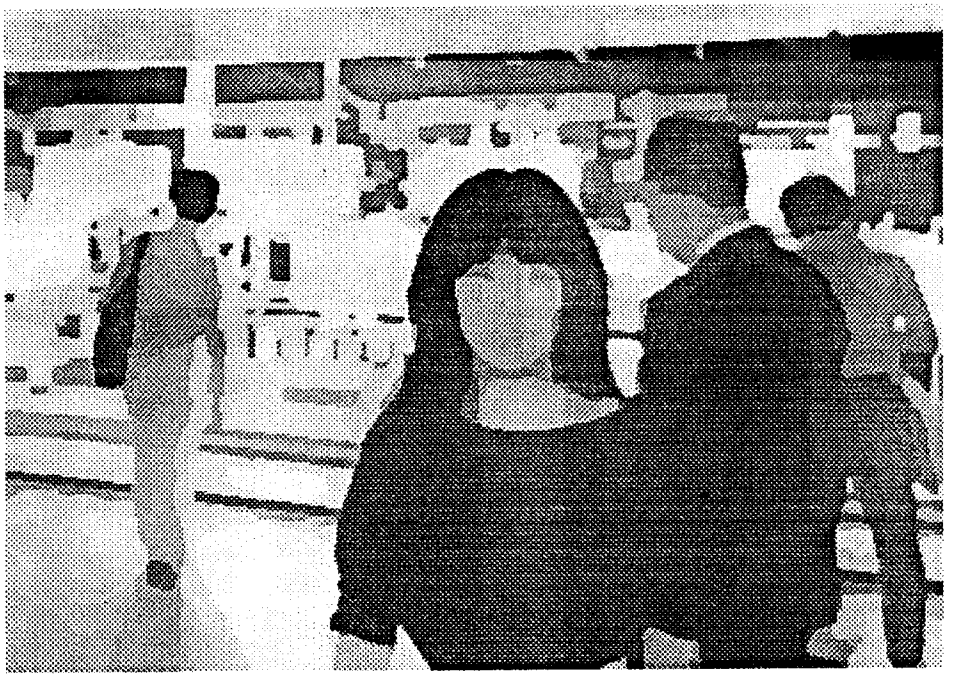


Fig. 4: (e)

Fig. 4: (f)

Figure 4: Estimated image after i iterations, when $\chi_0 = \zeta_0$. (a) $i = 0$ (original image). (b) $i = 25$. (c) $i = 50$. (d) $i = 100$. (e) $i = 200$. (f) $i = 800$.

Fig. 5: (a)

Fig. 5: (b)

Fig. 5: (c)

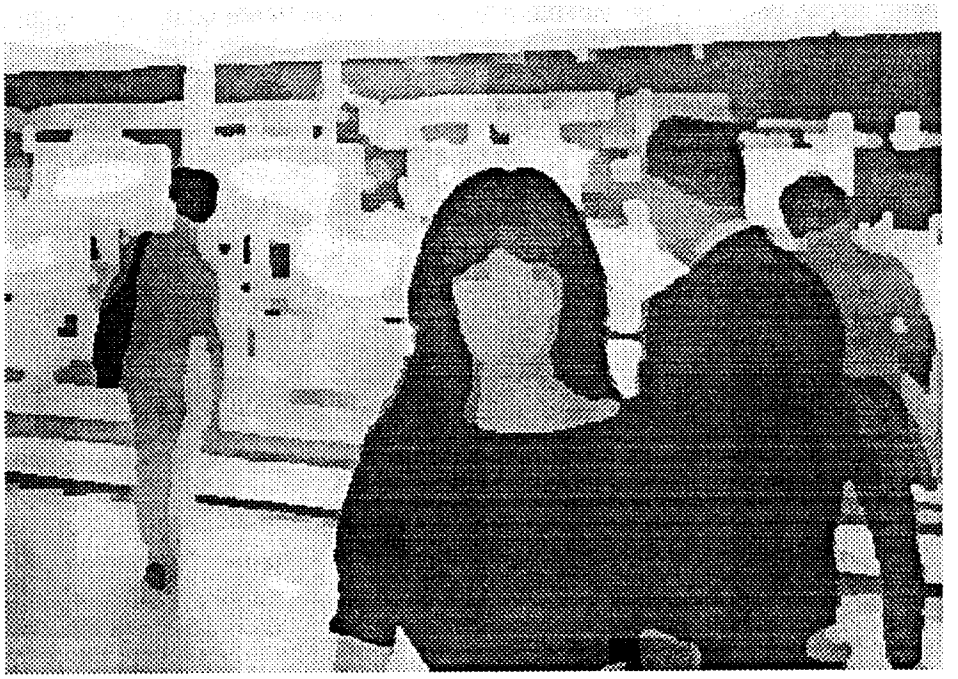


Fig. 5: (d)

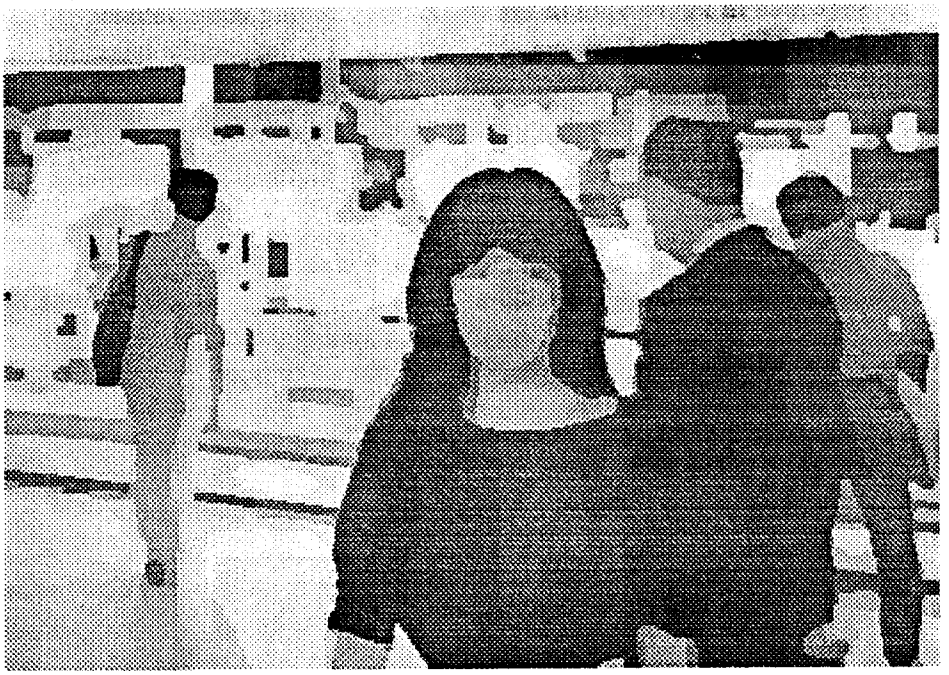


Fig. 5: (e)

Fig. 5: (f)

Figure 5: Estimated image after i iterations, when $\chi_0 = 0$. (a) $i = 25$. (b) $i = 50$. (c) $i = 100$. (d) $i = 200$. (e) $i = 400$. (f) $i = 800$.

to be limited to affect small blobs of high contrast relative to the local background. It is interesting to note, that the little dark blobs in the center of figure 4 (f), which are missing in figure 5 (f), represent pixel values, which are closer to the zero initial image, used to generate the sequence in figure 5, than are the corresponding pixel values (of the non-blobs) in figure 5 (f). This indicates, that the solution, which is implicitly selected by choosing a particular initial image function, tends to reflect the smoothness properties, rather than the actual values of the initial image function.

The non-uniqueness of the solutions of (23), stemming from the existence of multiple local minima of the total cost functional (2), should not be very surprising. In fact for most of the other existing regularization based edge detection methods it is relatively easy to construct examples of original image functions, for which the total cost functional exhibits this behavior. It is clear from the experimental results shown in figure 5, if not by intuition, that all the local minima of the total cost are potentially satisfactory solutions to the edge detection problem. Moreover by choosing the initial image function χ_0 to equal either the original image function or a constant, it seems like we have found a method of selecting those local minima, which correspond to the cases of the most and least detailed estimated images respectively. These extreme cases might actually be of more interest, than the solution corresponding to the global minimum.

For our observations regarding the parameter dependence of the solution, that is the influence of the edge cost coefficient λ and the pixel width h on the estimated image function z, we recall, that $r \doteq 1/h$ is a true scale-space parameter governing the spatial resolution of the edge detector, and that $\sqrt{\lambda}$, proportional to the edge enhancement threshold γ_0, controls its sensitivity in a linear fashion. Since the local differences of the (original) image function, unlike the discrete approximations of its derivatives, remain invariant under scale-space variations in terms of h, a more meaningful sensitivity parameter is in this context given by $s \doteq \sqrt{\lambda}h$, which is proportional to the corresponding local difference enhancement threshold $\gamma_0 h$. (The same conclusion would have been obtained, had we kept h constant and instead incorporated the explicit scale-space parameter μ in the total cost functional, as discussed in section 3.) Figure 6 shows an example of how the estimated image function (after 100 iterations) depends on the scale-space parameter r for a fixed sensitivity parameter ($s = \sqrt{20}$). Its dependence on the sensitivity parameter s for a fixed scale-space parameter ($r = \sqrt{50},$) is illustrated in figure 7.

In order to extract a set of edges from the estimated image function z, we followed the strategy outlined in section 3, and simply thresholded the gradient magnitude. Figure 8 shows the edges extracted from the estimated image function in figure 6 (b) using two different thresholds, one lower than, and the other one equally much higher than the edge enhancement threshold γ_0. As predicted by the discussion in section 6, the experiments confirm, that the edge extraction is very robust with respect to changes in the threshold ϑ for a wide range of thresholds around γ_0. In fact, if one allows a couple of edge segments to change, the range in question in this case extends well beyond, that spanned by the three examples in the figure.

(a)

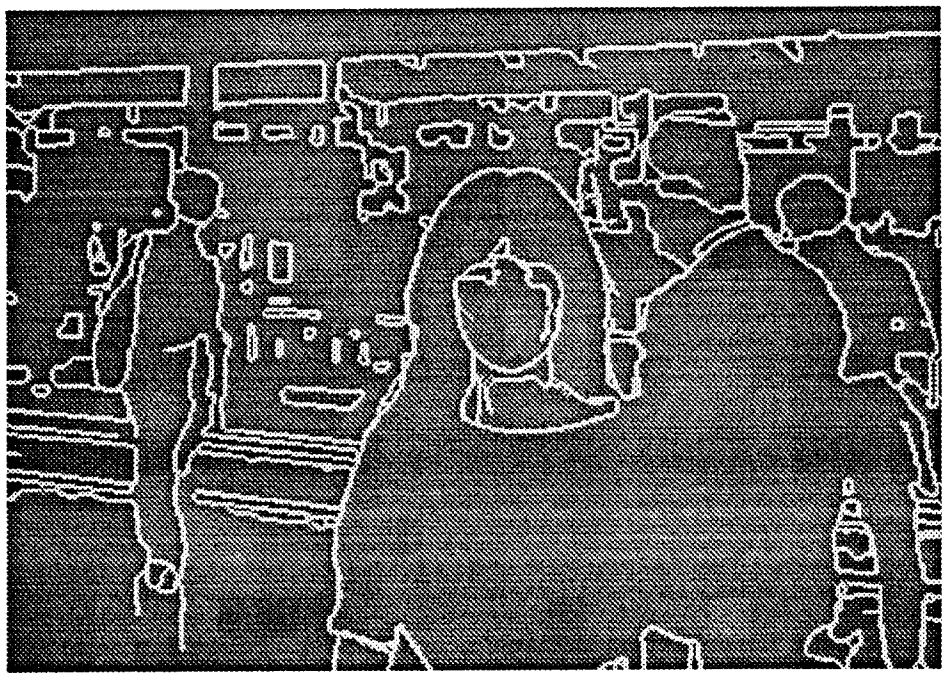
(b)

Figure 6: Estimated images for different values of the scale-space parameter r. (a) $r^2 = 12.5$. (b) $r^2 = 100$.

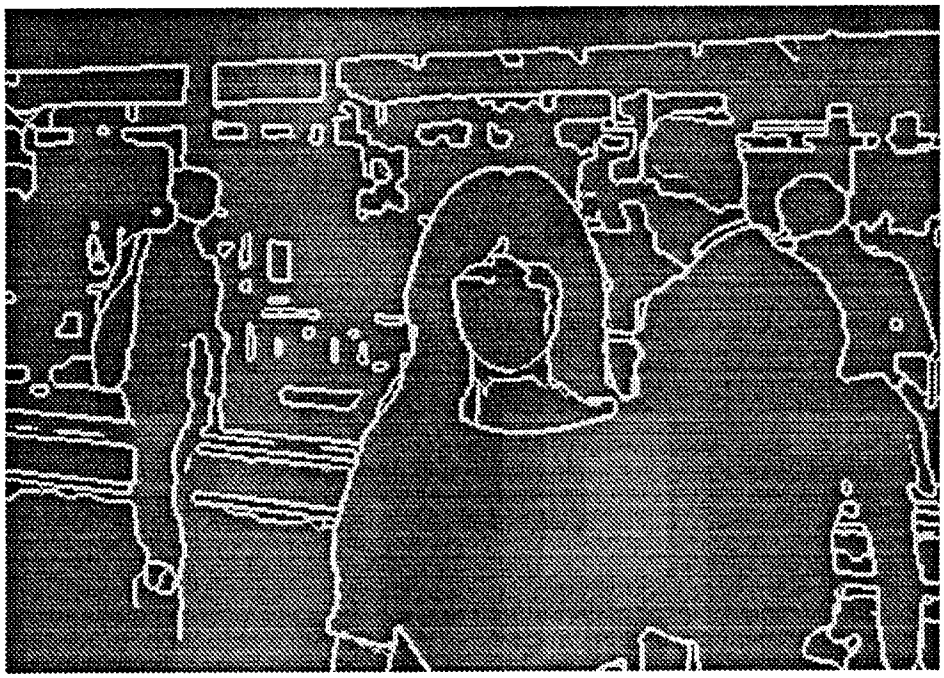
(a)

(b)

Figure 7: Estimated images for different values of the sensitivity parameter s. (a) $s^2 = 10$. (b) $s^2 = 40$.



(a)



(b)

Figure 8: Extracted edges for different values of the threshold $\vartheta \doteq g^{-1}(\theta)$. (a) $\vartheta = 22$. (b) $\vartheta = 67$.

10 Conclusion

We have proposed a global edge detection method based on variational regularization. We have also observed, that this method can be viewed as a biased anisotropic diffusion method. This circumstance exemplifies the close connection between the regularization and diffusion approaches in early vision, and we hope, that our analysis has shed some fruitful light on this interesting subject. Besides being of general interest, the coincidence of the two paradigms has also allowed us to analyze our variational edge detection method in the diffusion context. We have for example showed, that it shares the attractive edge enhancement property, characteristic of the unbiased anisotropic diffusion method.

Unlike other existing regularization approaches to edge detection, our method is tailored to support calculus of variations, not only with respect to the estimated/reconstructed image function, but also with respect to the continuity control function representing the edges. This modification of the paradigm leads to substantial computational savings in comparison with the other regularization methods, as far as we understand, without impairing the performance. The sharpness of the edges, which is seemingly given up from the outset, is regained during the iteration by the edge enhancement mechanism. This was demonstrated by our theoretical analysis as well as by our experimental results.

The most notable difference between our method and other existing anisotropic diffusion methods is, that our method converges to the solution of interest. This fact removes the problem of deciding when to stop the diffusion process as well as that of actually stopping it — an obvious and important advantage, if hardware implementations are to be considered. The price, that one pays for this improvement, is that the estimated image functions for different values of the scale-space parameter no longer can be generated recursively.

For the solution of the variational edge detection problem we have proposed an iterative algorithm. For a practically limited range of parameter values this algorithm has been found to be extremely well-behaved; it converges to a unique solution of the discretized problem, independently of the initial image function, that is the initial state of the iteration process.

In contrast to the limitations of our theoretical convergence analysis, our experimental results have clearly demonstrated, that our method works very well for typical parameter values of interest for edge detection. The algorithm does indeed converge to a solution of interest, that is an estimated image function, which is remarkably robust with respect to the initial image function. Furthermore the edges, which are obtained by postprocessing the estimated image function with a rudimentary local edge detector — thresholding of the gradient — are insensitive to changes in the threshold — the goal of the regularization. In addition to the convergence and robustness issues our experiments have exhibited the dependence of the solution on the values of the scale-space and sensitivity parameters embedded in our paradigm.

Although a number of theoretically relevant problems have been left open, we are convinced, that the method presented in this paper, represents a most significant contribution to the general theory and methods of edge detection.

11 Acknowledgements

The author is grateful to M. Singer for writing most of the user interface for the software implementation of the algorithm and to K. Pister for generous help with generating the images.

A Appendix

Proof of Lemma 8.3: To be specific we will only prove the lemma for the diagonal approximation ($\rho^2 = 2$). The proof can however easily be reconstructed to cover the Cartesian case as well. For $i \in \mathbb{N}_0$, let $y_q^{(i)}$, $\tau_q^{(i)}$, $v_q^{(i)}$, $q \in S_2$ and $\overline{v}^{(i)}$ denote the functions associated with $y_0^{(i)}$ corresponding to let $z_q^{(i)}$, $\sigma_q^{(i)}$, $w_q^{(i)}$, $q \in S_2$ and $\overline{w}^{(i)} \doteq \overline{w}_2^{(i)}$ respectively. For simpler notation also define the following bounds:

$$\begin{aligned}
D^{(i)} &\doteq \|y_0^{(i)} - z_0^{(i)}\|_\infty \qquad i \in \mathbb{N}_0 \\
E &\doteq \|\eta_0 - \zeta_0\|_\infty \\
R_z &\doteq \bigvee_{j\in J}[\zeta_0(j) \vee \chi_0(j)] - \bigwedge_{j\in J}[\zeta_0(j) \wedge \chi_0(j)] \\
R_y &\doteq \bigvee_{j\in J}[\eta_0(j) \vee \psi_0(j)] - \bigwedge_{j\in J}[\eta_0(j) \wedge \psi_0(j)] \\
R &\doteq R_z + (E + D^{(0)}) \\
M &\doteq \|\zeta_0\|_\infty \vee \|\chi_0\|_\infty
\end{aligned}$$

From the definitions of the shifted functions z_q, y_q, $q \in S$, and proposition 8.2 we further note that

$$\begin{array}{rcll}
D^{(i)} & \geq & \|y_q^{(i)} - z_q^{(i)}\|_\infty & \forall q \in S \quad \forall i \in \mathbb{N}_0 \\
R_z & \geq & \bigvee_{j\in J} z_q^{(i)}(j) - \bigwedge_{j\in J} z_q^{(i)}(j) & \forall q \in S \quad \forall i \in \mathbb{N}_0 \\
R_y & \geq & \bigvee_{j\in J} y_q^{(i)}(j) - \bigwedge_{j\in J} y_q^{(i)}(j) & \forall q \in S \quad \forall i \in \mathbb{N}_0 \\
R & \geq & \frac{R_y + R_z}{2} & \\
M & \geq & \|z_q^{(i)}\|_\infty & \forall q \in S \quad \forall i \in \mathbb{N}_0
\end{array}$$

Dropping the dependence on $j \in J$ and $i \in \mathbb{N}_0$ for shorter notation, from (22c) we then have

$$\begin{aligned}
&|\tau_q^2 - \sigma_q^2| \\
&\quad = \frac{1}{2h^2}|(y_{q_1,q_2} - y_{0,0})^2 + (y_{q_1,0} - y_{0,q_2})^2 - (z_{q_1,q_2} - z_{0,0})^2 - (z_{q_1,0} - z_{0,q_2})^2| \\
&\quad = \frac{1}{2h^2}|(y_{q_1,q_2} - y_{0,0} + z_{q_1,q_2} - z_{0,0})(y_{q_1,q_2} - y_{0,0} - z_{q_1,q_2} + z_{0,0}) \\
&\qquad + (y_{q_1,0} - y_{0,q_2} + z_{q_1,0} - z_{0,q_2})(y_{q_1,0} - y_{0,q_2} - z_{q_1,0} + z_{0,q_2})| \\
&\quad \leq \frac{2(R_y + R_z)2D}{2h^2} \\
&\quad \leq \frac{4RD}{h^2} \qquad \forall j \in J \qquad \forall i \in \mathbb{N}_0
\end{aligned}$$

Thus from (10) we see that

$$|v_q - w_q| = |g \circ \tau_q - g \circ \sigma_q| = \frac{v_q w_q |\sigma_q^2 - \tau_q^2|}{\lambda} \leq \frac{4 v_q w_q RD}{\lambda h^2} \qquad \forall j \in J \qquad \forall q \in S_2 \qquad \forall i \in \mathbb{N}_0$$

which in turn implies that

$$|\overline{v} - \overline{w}| \leq \sum_{q\in S_2} |v_q - w_q| \leq \sum_{q\in S_2} \frac{4 v_q w_q RD}{\lambda h^2} \leq \frac{4\overline{v}RD}{\lambda h^2} \qquad \forall j \in J \qquad \forall i \in \mathbb{N}_0$$

Hence

$$
\begin{aligned}
&|y_0 - z_0| \\
&\quad = \left| \frac{1}{2h^2 + \overline{v}} \left(2h^2 \eta_0 + \sum_{q \in S_2} v_q y_q \right) - \frac{1}{2h^2 + \overline{w}} \left(2h^2 \zeta_0 + \sum_{q \in S_2} w_q z_q \right) \right| \\
&\quad \leq \left| \frac{1}{2h^2 + \overline{v}} - \frac{1}{2h^2 + \overline{w}} \right| \left| 2h^2 \zeta_0 + \sum_{q \in S_2} w_q z_q \right| \\
&\qquad + \frac{1}{2h^2 + \overline{v}} \left| 2h^2 (\eta_0 - \zeta_0) + \sum_{q \in S_2} (v_q y_q - w_q z_q) \right| \\
&\quad \leq \frac{|\overline{w} - \overline{v}|}{2h^2 + \overline{v}} M + \frac{1}{2h^2 + \overline{v}} (2h^2 + E + |\overline{v} - \overline{w}| M + \overline{v} D) \\
&\quad \leq \frac{\overline{v}}{2h^2 + \overline{v}} \left(\frac{8RM}{\lambda h^2} + 1 \right) D + E \qquad \forall j \in J \qquad \forall i \in \mathbf{N}_0
\end{aligned}
$$

Since $\overline{v}^{(i)}(j) \leq 4$, $\forall j \in J$, $\forall i \in \mathbf{N}_0$, we therefore conclude that

$$
D^{(i+1)} \leq \frac{1}{1 + \frac{h^2}{2}} \left(\frac{8RM}{\lambda h^2} + 1 \right) D^{(i)} + E \qquad \forall i \in \mathbf{N}_0 \tag{25}
$$

If $\lambda > 16RM/h^4$, the assertion of the lemma then follows. ■

Proof of Theorem 8.4: Assume that λ is large enough for lemma 8.3 to be conclusive. Let $z_0^{(i)}$, $y_0^{(i)}$, $i \in \mathbf{N}_0$, be given as in lemma 8.3 with $\psi_0 = z_0^{(1)}$ and $\eta_0 = \zeta_0$. Then $y_0^{(i)} = z_0^{(i+1)}$, $\forall i \in \mathbf{N}_0$, and $E = 0$. By lemma 8.3, a simple Cauchy sequence argument (in $L_\infty(J)$) and the observation, that the left hand side of (23) is a continuous function of z_0 with respect to the L_∞-topology, (z_q is a continuous function of z_0, $\forall q \in S$,) it then follows, that $z_0^{(i)}$ converges to a *solution* of (23) (as $i \to \infty$) independently of its initial value χ_0. Next let χ_0 and ψ_0 be two possibly different solutions of (23), and let $z_0^{(i)}$, $y_0^{(i)}$, $i \in \mathbf{N}_0$, be given as in lemma 8.3 with $\eta_0 = \zeta_0$. Then $z_0^{(i)} = \chi_0$, $y_0^{(i)} = \psi_0$, $\forall i \in \mathbf{N}_0$, and $E = 0$. Thus by lemma 8.3

$$
\|\psi_0 - \chi_0\|_\infty = \lim_{i \to \infty} \|y_0^{(i)} - z_0^{(i)}\|_\infty = 0
$$

which shows, that the solution of (23) is unique. Finally let χ_0 and ψ_0 be the solutions of (23) given, that the corresponding original image functions are ζ_0 and η_0 respectively, and let $z_0^{(i)}$, $y_0^{(i)}$, $i \in \mathbf{N}_0$, be given as before. Again $z_0^{(i)} = \chi_0$, $y_0^{(i)} = \psi_0$, $\forall i \in \mathbf{N}_0$. From proposition 8.1 and lemma 8.3 it thus follows that

$$
\|\psi_0 - \chi_0\|_\infty = \lim_{i \to \infty} \|y_0^{(i)} - z_0^{(i)}\|_\infty \leq \|\eta_0 - \zeta_0\|_\infty
$$

which proves the stability of the solution. ■

References

[1] V. Torre and T. Poggio, "On edge detection," A.I. Memo 768, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Cambridge, MA, Aug. 1984.

[2] T. Poggio, H. Voorhees, and A. Yuille, "A regularization solution to edge detection," A.I. Memo 833, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Cambridge, MA, May 1985.

[3] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 371, pp. 314 – 319, Sep. 1985.

[4] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721 – 741, Nov. 1984.

[5] J. L. Marroquin, *Probabilistic Solution of Inverse Problems.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, Sep. 1985.

[6] D. Terzopoulos, "Multilevel computational processes for visual surface reconstruction," *Computer Vision, Graphics, and Image Processing*, vol. 24, pp. 52 – 95, 1983.

[7] D. Terzopoulos, "Computing visible-surface representations," A.I. Memo 800, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Cambridge, MA, March 1985.

[8] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 413 – 424, July 1986.

[9] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Conference on Computer Vision and Pattern Recognition*, IEEE, 1985.

[10] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," 1986. Unpublished.

[11] A. Blake and A. Zisserman, "Invariant surface reconstruction using weak continuity constraints," in *Conference on Computer Vision and Pattern Recognition*, pp. 62 – 67, IEEE, 1986.

[12] A. Blake and A. Zisserman, "Some properties of weak continuity constraints and the GNC algorithm," in *Conference on Computer Vision and Pattern Recognition*, pp. 656 – 660, IEEE, 1986.

[13] A. Blake and A. Zisserman, *Visual Reconstruction.* Cambridge, MA: The MIT Press, 1987.

[14] D. Lee and T. Pavlidis, "One-dimensional regularization with discontinuities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 822 – 829, Nov. 1988.

[15] K. N. Nordström, *A Variational Approach to Edge Detection.* PhD thesis, University of California, Berkeley, CA, 1989. To be filed.

[16] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," in *Workshop on Computer Vision—Miami*, pp. 16 – 22, IEEE Computer Society, June 1987.

[17] P. Perona and J. Malik, "A network for edge detection and scale space," in *International Symposium on Circuits and Systems—Helsinki*, pp. 2565 – 2568, IEEE, June 1988.

[18] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," Report UCB/CSD 88/483, Computer Science Division University of California, Berkeley, CA, Dec. 1988.

[19] A. N. Tikhonov, "Regularization of incorrectly posed problems," *Sov. Math. Dokl.*, vol. 4, pp. 1624 – 1627, 1963.

[20] A. N. Tikhonov and V. I. Arsenin, *Solutions of Ill-Posed Problems.* Washington, D.C.: Winston, 1977.

[21] M. Atteia, "Fonctions "spline" et noyaux reproduisants d'Aronszajn-Bergman," *Revue Française d'Informatique et de Recherche Operationelle*, vol. 4, pp. 31 – 43, Oct. 1970. Série Rouge.

[22] J. Duchon, "Splines minimizing rotation-invariant semi-norms in sobolev spaces," in *Conference Held at Oberwolfach April 25 – May 1*, (W. Schempp and K. Zeller, eds.), (Berlin), pp. 85 – 100, Springer, 1976. Series: Constructive Theory of Functions of Several Variables, edited by A. Dold and B. Eckmann.

[23] J. Meinguet, "An intrinsic approach to multivariate spline interpolation at arbitrary points," in *Polynomial and Spline Approximation—Theory and Applications*, (B. N. Sahney, ed.), (Dortrecht, Holland), pp. 163 – 190, NATO Advanced Study Institute, D. Reidel, 1978.

[24] J. Meinguet, "Multivariate interpolation at arbitrary points made simple," *Journal of Applied Mathematics and Physics (ZAMP)*, vol. 30, pp. 292 – 304, 1979.

[25] D. G. Luenberger, *Linear and Nonlinear Programming.* Reading, MA: Addison-Wesley, second ed., 1984.

[26] R. Courant and D. Hilbert, *Methods of Mathematical Physics.* Vol. II, New York: John Wiley & Sons, 1962.

[27] D. Marr and E. Hildreth, "Theory of edge detection," A. I. Memo 518, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Cambridge, MA, Apr. 1979.