January 1991 LIDS-P-2016

# SCALING RESULTS FOR THE VARIATIONAL APPROACH TO EDGE DETECTION

Thomas J. Richardson[1], Sanjoy K. Mitter[2]

## Abstract

In another paper the first author presented an asymptotic result concerning the Mumford–Shah functional,

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

showing that if $g$ is approximately piecewise smooth and $\beta$ is sufficiently large then $\Gamma$ which minimize $E$ will be close to the discontinuity set of $g$ in the sense of Hausdorff metric. In this paper an algorithm is presented which implements the scaling suggested by the asymptotic results to obtain accurate localization of edges even when only large scale edges are being detected. An approximation to $E$ is also considered and the implications of the asymptotic results to this functional are also examined.

[1] AT&T Bell Labs, 2C-370, Murray Hill, NJ, 07974.

[2] Department of Electrical Engineering and Computer Science and Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139. Research partially supported by the Air Force Office of Scientific Research under contract 89-0279B and by the Army Research Office under contract DAAL03-86-K-0171 (Center for Intelligent Control Systems).

# 1 Introduction

Edge detection is regarded as one of the fundamental components of computer vision. An issue which arises in all edge detection schemes is that of scale. One reason for this is the basic insight which says that coarse (i.e. large scale) representations are less complex than more detailed ones. Coarse scale descriptions can effectively be used to select "regions of interest" for detailed processing, thus reducing the demand on computational resources. It is important therefore that coarse scale descriptions retain those features of the data which are required for effective decision making. In the case of edge detection for example, T-junctions and corners play important roles in estimating the depth and the shape of objects in a scene. It is desirable therefore that even at coarse scales these features be accurately represented. The purpose of this paper is to present (without proofs) certain analytical results on the Mumford-Shah functional related to its scaling properties, and then to show how these results support the use of certain approaches to edge detection.

By edge detection we mean the problem of location step discontinuities in an otherwise smooth function, perhaps in the presence of noise and smearing. There are many methods which have been proposed for detecting edges of this type. Two classes which comprise a majority of these methods are:

1. Techniques which consist of linear filtering followed by some non-linear operation such as the locating of the maxima of gradients.

2. Techniques which combine the smoothing and nonlinear operation into a single formulation or process.

Well-known example from the first class include the Marr-Hildreth edge detector [25] and the Canny edge detector [8]. Within the second class we find *Markov Random Field* formulations [17] [26] [14], the *Variational* formulation [7] [28] [29], and *Non-linear Diffusion* approaches [30]. Historically the first class preceded the second. The central idea the second class added to those of the first class was that of performing simultaneous edge detection and smoothing rather than the previously used two step procedures. The motivation for this was the fact that the smoothing procedures blur across boundaries obscuring and distorting the edges, especially at high curvature locations such as corners and T–junctions. The three approaches which we mentioned as belonging to the second class can all be related to an optimization problem; the minimization of some functional typically having three terms. One example, the point of reference for this paper, is the Mumford–Shah functional associated with the variational formulation;

$$E(f,\Gamma) = \beta \int_{\Omega} (f-g)^2 + \int_{\Omega \setminus \Gamma} |\nabla f|^2 + \alpha \ \text{length}(\Gamma).$$

Here $g$ represents the data, which we will think of as a real valued function. The symbol $\Gamma$ denotes the set of edges in the image and $f$ is a piecewise smooth approximation to $g$. The problem is to minimize $E$ over admissible $f$ and $\Gamma$. The parameters $\beta$ and $\alpha$ control the competition between the terms; they decide the "scale" of the edge detection. The the second term of $E$ provides for the interaction between the edges and the smoothing mechanism by allowing $\Gamma$ to control or modulate the smoothness constraint on $f$. It was expected, and demonstrated to some extent in [7] that this approach better localizes the edges than the methods of class 1. However, analysis of optimality conditions revealed that optimal $\Gamma$ can have only certain restricted types of local geometries, implying in particular that T–junctions and corners tend to be distorted [29]. On the other hand a result of the first author showed that these constraints were truly of a local nature and that globally the variational approach should produce reasonable solutions. The result is an asymptotic one stating that as $\beta \to \infty$ optimal $\Gamma$ will converge in an appropriate sense to the discontinuity set of $g$ provided the noise and smearing are removed from the data sufficiently quickly. This, then, is a kind of fidelity result for the variational formulation. Furthermore the results suggest a principle whereby it may be possible to recover coarse scale edges alone with the localization accuracy usually available only on the finest scale.

An algorithm is developed which on a small time scale resembles the minimizing of the energy functional but on a longer time scale evolves by changing the parameters of the functional in the direction of the aforementioned limit. In order to prevent the resolving of microscopic detail as the limit is taken, i.e. in order to retain only large scale features, we systematically remove small scale features as if they were a distortion of the ideal image. The rates and topological structure for this removal (which is achieved by edge location dependent smoothing) are governed by the convergence theorem mentioned above.

The implementation of the variational formulation itself presents several difficulties. It turns out that there is a method for approximating $E$, using the concept of $\Gamma$ convergence that allows for a straight–forward implementation (a gradient descent method). Furthermore we can argue that this approximation may in some sense be superior to the original one in the manner in which it distorts the singularities i.e. the T–junctions and the corners, and this argument is based on the principle referred to above.

## 2 Summary and Outline of the Paper

For the sake of compactness we have chosen to suppress most of the mathematical results. In Section 3 some of what is known about the fundamental question of existence of minimizers to the (continuous) variational problem is reviewed. Some of the implications of the formulation on the structure of minimal boundaries is presented. These results help to motivate our work which aims to circumvent those constraints and to demonstrate their local nature. We view these constraints as undesirable structural restrictions placed on solutions by ad hoc choices in the formulation of the energy functional. What the functional offers in return is a relatively simple structure that admits analysis.

Section 4 summarizes our main contributions to the analytical understanding of the variational formulation of the segmentation problem. The results demonstrate an asymptotic fidelity of the variational approach. The ideas inherent in these results also serve as the primary justification and motivation for the algorithm. To obtain these results we assume that the image is a corrupted version of piecewise constant or piecewise smooth function, depending on the particular problem formulation. The results state that asymptotically as $\beta \to \infty$ the boundaries given as a solution to the variational problem converge (in Hausdorff metric) to the discontinuity set of the underlying image. These results can be thought of as a counterpoint to the results of the calculus of variations. The results of the calculus of variations imply that the minimizers have certain local structure which from the image processing point of view may be undesirable. The limit theorems say that when viewed globally the solutions behave well and asymptotically essentially any structure i.e. any boundary geometry, can be recovered.

In Section 5 we sketch an algorithm. The main ideas on which the algorithm is based are derived from the limit theorems. The structure of the algorithms closely resembles that of the limit theorems; in fact the limit theorems can be interpreted as consistency results for the algorithms.

Section 6 is devoted to presenting the Γ–convergent approximation to $E$. Some basic properties of solutions to this approximation are stated. In Section 7 the details of the computation are given. In Section 8 we present the results of some experiments.

# 3 The Variational Model for Edge Detection

Three techniques for image segmentation and reconstruction based on intensity information which have recently gained considerable attention are Markov Random Fields, Variational Formulations, and Non-linear Filtering. Most researchers in this area have realized that these methods are closely connected (see [15] or [30]); the practical differences lying mostly in the conception of the computation to be carried out. The essential feature which these models are designed to capture; simultaneous smoothing and edge enhancement/boundary detection, is achieved in essentially the same way. Our work is connected with these methods, it is most convenient to relate it to the Variational formulation.

The Variational formulation models edge detection as the minimization of an energy functional. The functional introduced by Mumford and Shah, [28] [29], and (in discrete form) referred to as the weak membrane by Blake and Zisserman [7] is the following,

$$E(f,\Gamma) = \beta \int_{\Omega} (f-g)^2 + \int_{\Omega\setminus\Gamma} |\nabla f|^2 + \alpha \,\text{length}(\Gamma)$$

where $\alpha$ and $\beta$ are positive real scalars (the parameters of the problem) and $f$ is a piecewise smooth approximation to $g$ having discontinuities only on the set $\Gamma$ which one interprets as the edges found in the image. The first term of $E$ penalizes the fidelity of the approximating image $f$ to the data $g$. The second term imposes some smoothness on $f$. The third term penalizes the total length of the boundary (which we think of as the union of curves). The removal of any term results in trivial solutions yet with all three terms the functional captures in a simple way the desired properties of a segmentation/approximation by piecewise smooth functions.

The parameters $\beta$ and $\alpha$ have to be chosen. Since we have not fixed them a priori we have really defined a two dimensional space of functionals. It is of interest to examine certain limiting versions of the functional.

Consider allowing $\beta$ and $\alpha$ to tend to zero while keeping their ratio fixed. Relative to the other terms the smoothing term would dominate. Clearly any limit of minimizers would necessarily be a locally constant function on $\Omega\backslash\Gamma$ (where $\Gamma$ would be the limiting boundaries.) Mumford and Shah were thus lead to introduce the following functional,

$$E_0(f,\Gamma) = \sum_i \int_{\Omega_i} (f_i - g)^2 + \alpha \text{ length}(\Gamma)$$

where $\Gamma = \Omega\backslash \cup_i \Omega_i$ and the $f_i$ are constants. This functional, because of its greater simplicity lends itself to more thorough analysis.

The energy functional associated with the variational model is ad hoc. In [29] the structure of minimizers of the functional was studied via the calculus of variations. A summary of a few of the results on the structure of minimal $\Gamma$ is presented here since this paper is partially motivated be the desire to circumvent some of these constraints. The following constraints on $\Gamma$'s which minimize $E$ were proved by Mumford and Shah in [29]. They are illustrated in Figure 1.

- If $\Gamma$ is composed of $C^{1,1}$ arcs then at most three arcs can meet at a single point and they do so at $120^o$.
- If $\Gamma$ is composed of $C^{1,1}$ arcs then they meet $\partial\Omega$ only at an angle of $90^o$.
- If $\Gamma$ is composed of $C^{1,1}$ arcs then it never occurs that two arcs meet at an angle other than $180^o$.

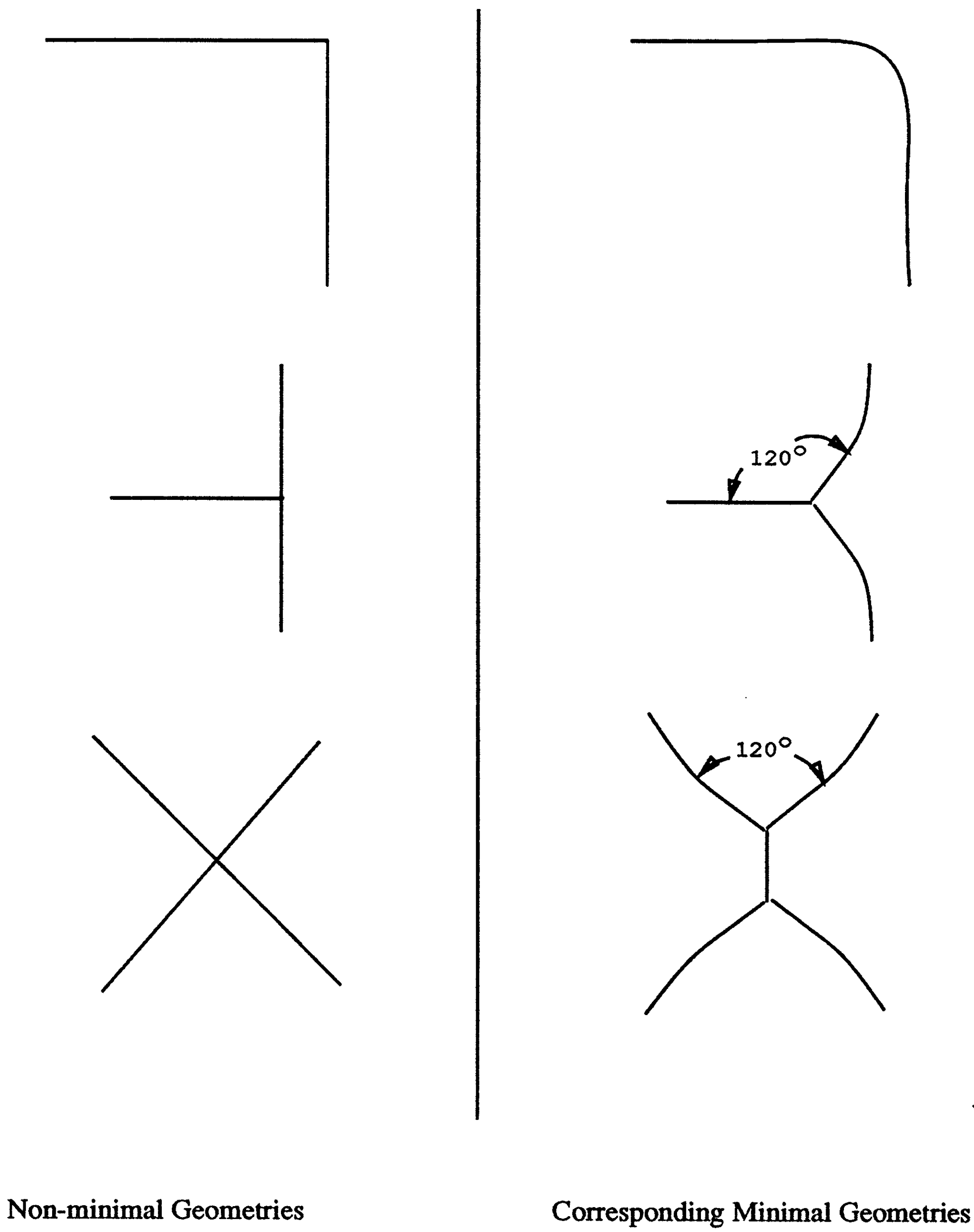


Figure 1: Calculus of Variations Results

- If $x \in \Gamma$ and in a neighborhood of $x$, $\Gamma$ is the graph of a $C^2$ function then $(\beta(f-g)^2+|\nabla f|^2)^+ - (\beta(f-g)^2+|\nabla f|^2)^- + \alpha\,\mathrm{curv}(\Gamma) = 0$ where the superscripts $+$ and $-$ denote the upper and lower trace of the associated function on $\Gamma$ at $x$ and $\mathrm{curv}(\Gamma)$ denotes the curvature of $\Gamma$ at $x$.

The difficulty with these results is that they do not support the use of the variational approach as an image segmenting scheme with respect to the goal of obtaining intuitively appealing segmentations. In particular T–junctions, believed to be important image features, tend to be distorted and corners tend to be rounded out. The restrictions on the geometry of the edges arises out of the model and are artifacts of the particular formulation and do not reflect an intrinsic property of the problem at hand. How then can one improve upon such as hoc models ? One idea is the following. Consider the set of all possible minimizers of the functional $E$, over all possible values of the parameters. Each of these minimizers possess the properties which the model imposes. However, if we take the closure of these functions in an appropriate topology we may widen the class of functions considerably. The asymptotic theorems outline in Section 4 indicate that particular meaningful members of such a closure may be found by taking the parameters associated with the functional to certain limits. In fact in this way one can produce essentially any piecewise smooth function with edges having arbitrary geometries. An idea which clearly presents itself is to develop an algorithm in which the same limit is taken. This is one of the purposes of this paper.

In all known segmentation/edge detection schemes there exist parameters which can be related in some sense to "scale". There is no generally accepted definition of "scale" but often one has notions of size, contrast and geometry in mind. Consider, for example, edge detection techniques based on convolution of the image with Gaussian kernels followed by detection of gradient maxima [24] [25]. Here the relevant parameter is $\sigma$, the "variance" of the kernel. For each value of $\sigma$ one obtains a different set of edges. Ideally one would hope that as $\sigma$ increases that the set of edges would decrease monotonically. However, this is not the case; in general the edges drift as the scale varies. Since on the finest scales it is desirable to know which edges correspond to gross features in the image there arises the problem of finding within the small scale edges those that correspond with large scale features. It is true that the edge sets vary continuously as a function of the parameter $\sigma$ but in general it is computationally too costly to compute boundaries for sufficiently dense a set of $\sigma$ to make the tracking obvious.

In the energy base formulations there are usually 2 free parameters associated with the problem. In this sense the use of the word "scale" is misleading. When Blake an Zisserman [7] speak of varying the scale of the problem they consider varying the coefficient on the smoothing term in $E$ (which is set to 1 in our formulation.) This is equivalent to varying $\alpha$ and $\beta$ while keeping their ratio fixed. While $\frac{1}{\sqrt{\beta}}$ can be interpreted as the analog to $\sigma$, keeping the ratio smoothing conception of "scale" since this has the effect of keeping the total quantity of boundary and the localization errors roughly constant. In our limit theorems we (usually) keep $\alpha$ fixed and let $\beta$ tend to $\infty$. Thus for a fixed $\alpha$ "scale" can be thought of as proportional to $\frac{1}{\sqrt{\beta}}$. However, in general there are two parameters and these parameters describe the range of functionals under consideration. In [36] the significance of the two parameters is studied in the context of the 1-dimensional segmentation problem.

## 3.1 Existence Results

For the functional,

$$E_0 = \beta \sum_i \int_{\Omega_i} (f_i - g)^2 + \text{length}(\Gamma)$$

(where the $\Omega_i$ are the connected components of $\Omega \backslash \Gamma$ and the $f_i$ are constants), the following has been proved.

**Theorem 1** *[27] Let $\Omega$ be an open rectangle and let $g \in L^\infty(\Omega)$. For all one-dimensional sets $\Gamma \subset \Omega$ such that $\Gamma \cup \partial\Omega$ is made up of a finite number of $C^{1,1}$ - arcs, meeting each other only at their end-points, and, for all locally constant functions $f$ on $\Omega \backslash \Gamma$, there exists an $f$ and a $\Gamma$ which minimize $E_0$.*

Mumford and Shah [29] proved a similar theorem with the restriction that $g$ be continuous of $\overline{\Omega}$. In this case they showed that $\Gamma$ is composed of a finite number of $C^2$ curves. The proof relied heavily on results from geometric measure theory. The theorem quoted above was proved by Morel and Solimini [27] using direct, constructive methods. Finally, another proof using $\Gamma$ restricted to be unions of line segments and then taking limits as the segment lengths tend to zero was achieved by Y. Wang [37].

To formally state the best known existence theorem for $E$ we will require a little more notation. It is still an open problem to show existence of minima for the functional $E$ with sufficient regularity of the boundary to allow the analysis of Mumford and Shah [29] to go through. The best available existence results necessarily allows the boundary $\Gamma$ to be sufficiently irregular that 'length' cannot be defined for it. This measure is therefore replaced with a more general measure.

### 3.1.1 Hausdorff Measure

A *curve* $\Gamma \subset \Re^n$ is the image of a continuous injection $g : [0,1] \to \Re^n$. The *length* of a curve $\Gamma$ is defined as

$$L(\Gamma) = \sup\{\sum_{i=1}^{m} \|g(t_i) - g(t_{i-1})\| \ : \ 0 = t_0 < t_1 < \cdots < t_m = 1\}$$

and $\Gamma$ is said to be *rectifiable* if $L(\Gamma) < \infty$.

For a non-empty subset $A$ of $\Re^n$, the *diameter* of $A$ is defined by $\text{diam}(A) = \sup\{\|x - y\| : x, y \in A\}$. Define

$$\mathcal{H}^1_\delta(A) = \ \inf\{\sum_{i=1}^{\infty} \text{diam}(U_i)^s \ : A \subset \bigcup_{i=1}^{\infty} U_i, \ \text{diam}(U_i) \leq \delta\},$$

The *Hausdorff 1-dimensional measure* of $A$ is then given by

$$\mathcal{H}^1(A) = \lim_{\delta \to 0} \mathcal{H}^1_\delta(A) = \sup_{\delta > 0} \mathcal{H}^1_\delta(A)$$

Many properties of Hausdorff measure can be found in [11, 12, 32]. The following theorem states that $\mathcal{H}^1$ is a generalization of length, as required.

**Theorem 2** *If $\Gamma \subset \Re^n$ is a curve, then $\mathcal{H}^1(\Gamma) = L(\Gamma)$.*

**Proof** See [11] Lemma 3.2. □

The following theorem is a structure theorem for closed sets of finite $\mathcal{H}^1$ measure. We refer to a compact connected set as *continuum.*

**Theorem 3** *If $\Gamma$ is a continuum with $\mathcal{H}^1(\Gamma) < \infty$, then $\Gamma$ consists of a countable union of rectifiable curves together with a set of $\mathcal{H}^1$-measure zero.*

**Proof** See [11], Theorem 3.14. □

A natural "weak" formulation of the variational principal is thus the following,

$$E(f,\Gamma) = \beta \int_\Omega (g-f)^2 + \int_{\Omega\backslash\Gamma} \|\nabla f\|^2 + \mathcal{H}^1(\Gamma) \tag{1}$$

where $\Gamma$ being a relatively closed subset of $\Omega$ and $f \in W^{1,2}(\Omega\backslash\Gamma)$ where $W^{1,2}$ is the Sobolev space as defined in [1]. An existence result now exists for this formulation due to results of Ambrosio [2] [3] [4] and DeGiorgi–Carriero–Leaci [10] but it is beyond the scope of this paper to describe these results.

# 4 The Asymptotic Theorems

In this section we state the limit theorems. The proofs are beyond the scope of this paper and will be published elsewhere (they may be found in [31]). The theorems are concerned with what happens to solutions of the variational formulation of the segmentation problem as $\beta \to \infty$. Some notation is required so that we may define what we mean by "convergence of a set of edges".

### 4.0.2 The Hausdorff Metric

For $A \subset \Re^n$, the *$\epsilon$-neighborhood* of $A$ will be denoted by $[A]_\epsilon$ and is defined by

$$[A]_\epsilon = \{x \in \Re^n : \inf_{y\in A} \|x-y\| < \epsilon\}$$

where $\|\cdot\|$ denotes the Euclidean norm. In the terminology of mathematical morphology [35], $[A]_\epsilon$ is the dilation of $A$ with the open ball of radius $\epsilon$. The notion of distance between boundaries which we will use is the *Hausdorff metric.* Denoted $d_H(\cdot,\cdot)$, the Hausdorff metric is evaluated by

$$d_H(A_1, A_2) = \inf\{\epsilon : A_1 \subset [A_2]_\epsilon \text{ and } A_2 \subset [A_1]_\epsilon\}.$$

Elementary considerations show that $d_H(\cdot,\cdot)$ is in fact a metric on the space of all non-empty compact subsets of $\Re^n$.

The results of this section assert that if the image is ideal i.e. a piecewise smooth or piecewise constant function depending on whether $E$ or $E_0$ is being considered, then the optimal boundaries $\Gamma$ converge to the discontinuity set of image with respect to the Hausdorff metric. Furthermore the convergence still holds if the image is corrupted by smearing and additive noise provided the smearing effect and the magnitude of the noise decay sufficiently quickly as $\beta$ tends to infinity. We treat the piecewise constant case (i.e. minimizing $E_0$) and piecewise smooth case (i.e. minimizing $E$) separately.

## 4.1 Problem Formulation

We will be examining minimizers of $E$ and $E_0$ the existence of which is asserted by the existence theorems (hence by $E$ we mean the weak form). Solutions are determined by $\Gamma$ (i.e. for a fixed $\Gamma$ the optimal $f$ if unique) and we will often refer to the solution $\Gamma$ meaning the pair $f, \Gamma$. Also, we will be varying the parameter $\beta$ and will use $\Gamma_\beta$ to indicate an optimal solution for a particular value of $\beta$.

The proofs require that we make certain assumptions on the data $g$. The limit theorem has been proved in greater generality than stated in this paper but we choose to avoid some of the more intricate mathematics. The piecewise constant and the piecewise smooth case need to be treated in part separately. The piecewise constant case is described first.

### 4.1.1 The Piecewise Constant Case

The case we are interested in is one in which the image is a corrupted version of a piecewise constant $L^\infty$ function $\hat{g}$. We will define a set which we interpret as the natural candidate for a set of boundaries in the image; this will be the discontinuities of the image.

Suppose that $\Omega$ can be decomposed into a countable number of disjoint sets $A_j$ having piecewise Lipschitz boundaries (say) such that on each $A_j$, $\hat{g}$ is constant. We define the "boundary" $\mathcal{B}_g$ to be $\Omega \cap \bigcup_j \partial A_j$. We assume length$(\mathcal{B}_g) < \infty$ and that if length $(\partial A_i \cap \partial A_j > 0)$ then $\hat{g}(A_i) \neq \hat{g}(A_j)$, i.e. the boundary should be "visible" from $\hat{g}$.

### 4.1.2 The Piecewise Smooth Case

To define the edges in the piecewise smooth case assume $\hat{g} \in L^\infty(\Omega)$. The Lebesgue points of $u$, i.e.,

$$\{x : \exists z : \lim_{\rho \to 0+} \rho^{-n} \int_{B_\rho(x)} |u - z| dx = 0\}$$

are the points of approximate continuity $\hat{g}$. The set of edges, which we denote by $\mathcal{B}_g$ is simply the complement of the Lebesgue points of $u$ in $\Omega$, and for convenience we assume it is relatively closed, (a weaker sufficient assumption is to assume that $\mathcal{H}^1 \overline{\mathcal{B}_g} \backslash \mathcal{B}_g = 0$).

The following summarizes our assumptions in both cases.

**Assumption 1:** $\hat{g} \in L^\infty(\Omega)$, $\int_{\Omega \backslash \mathcal{B}_g} |\nabla \hat{g}|^2 + \mathcal{H}^1(\mathcal{B}_g) < \infty$ and $\mathcal{B}_g$ has no isolated points i.e. if $x \in \mathcal{B}_g$ then $\forall \rho > 0$, $\mathcal{H}^1(\mathcal{B}_g \cap B_\rho(x)) > 0$.

**Assumption 2:** If $A \subset \Omega$ is an open set satisfying dist$(A, \mathcal{B}_g) > 0$ then there exists an $L < \infty$ such that if $x$ and $y$ are the end points of a line segment lying in $A$ then then $|\hat{g}(x) - \hat{g}(y)| \leq L|x - y|$. We refer to $L$ as the Lipschitz constant associated with $A$.

Essentially we have assumed that $\hat{g} \in C^{0,1}(\Omega \backslash [\mathcal{B}_g]_\epsilon)$ for any $\epsilon > 0$.

### 4.1.3 The Noise Model

In this section we describe the restrictions on the relation between $g$ and $\hat{g}$ we need to make in order for the limit theorems to go through. A succinct statement of the assumptions can be made by defining a parametrized class of images $\Upsilon(\beta)$. The following are our assumptions on this class.

$$\lim_{\beta \to \infty} \sup_{g \in \Upsilon(\beta)} \beta \int_\Omega (g - \hat{g})^2 = 0, \tag{2}$$

and,

$$\forall \epsilon > 0, \lim_{\beta \to \infty} \sup_{g \in \Upsilon(\beta)} \|(g - \hat{g})(1 - \chi_{[\mathcal{B}_g]_\epsilon})\|_\infty = 0 \tag{3}$$

Under some mild additional assumptions we can convert this into a model allowing smearing and bounded additive noise. The main reason for allowing smearing is not to require the image to have actual jumps. To model smearing define $\mathcal{S}_r$ as the class of maps taking $L^\infty(\Omega)$ to $L^\infty(\Omega)$ having the property that the value of the image function at a point $x \in \Omega$ lies within the range of essential values that the argument function takes in a ball of radius $r$ around $x$. This models in a quite general way smearing of the image and hence distortion of the boundaries. More formally $\Phi \in \mathcal{S}_r$ of and only if $\Phi$ has the property

$$\Phi(g)(x) \in [\text{ess inf } g|_{B_r(x)}, \text{ ess sup } g|_{B_r(x)}].$$

An example of such a $\Phi$ would be a smoothing operator defined using a mollifier with support lying inside the ball of radius $r$, but nonlinear perturbations are also allowed. To admit this model of smearing it is convenient to make the following mild assumption,

*There is a constant $c_b < \infty$ such that $|[\mathcal{B}_g]_r \cap \Omega| < c_b r$.*

This assumption is automatically satisfied for a large class of sets containing all closed sets having finite $\mathcal{H}^1$ measure and finitely many connected components. This is a consequence of the following result from the theory of Minkowski content [12],

**Proposition 4** *[12] Let $\Gamma$ be a continuum in $\Re^2$ with $\mathcal{H}^1(\Gamma) < \infty$ then*

$$\lim_{\epsilon \to 0+} \frac{|[\Gamma]_\epsilon|}{2\epsilon} = \mathcal{H}^1(\Gamma).$$

A simplifying assumption for the piecewise smooth case is that the Lipshitz constants referred to in Assumption 2 can be uniformly bounded i.e. $L$ does not depend on $A$. (This is slightly more than we require but it simplifies the statements to follow.)

Now, for either case let $g$ have a representation of the form,

$$g = \Phi(\hat{g}) + \vartheta w \tag{4}$$

for some $\Phi \in \mathcal{S}_r$ and $w \in L^\infty$ with $\|w\|_\infty \leq 1$ and $\vartheta$ a real scalar. Further, assume that there are functions $h_r : (0,\infty) \to [0,\infty)$ and $h_\vartheta : (0,\infty) \to [0,\infty)$ satisfying

$$\lim_{\beta\to\infty} \beta h_r(\beta) = 0$$
$$\lim_{\beta\to\infty} \beta^{\frac{1}{2}} h_\vartheta(\beta) = 0.$$

Define $\Upsilon(\beta)$ to be those functions $g$ which can be written in the form 4 for some $\Phi_r, w$ and $\vartheta$ with $r \leq h_r(\beta)$ and $\vartheta \leq h_\vartheta(\beta)$. It now follows that with this definition of $\Upsilon(\beta)$ the assumptions 2 and 3 are satisfied since

$$\beta \int_\Omega (g - \hat{g})^2 \leq \beta |[\mathcal{B}_g]_r| \, \|g - \hat{g}\|_\infty + \beta(Lh_r(\beta) + \vartheta)^2 |\Omega|.$$

## 4.2 Statement of the Limit Theorems

Given that assumptions 1 and 2 are satisfied, the following holds,

**Theorem 5** *As $\beta \to \infty$ $\{\Gamma_\beta\}$ converges to $\overline{\mathcal{B}_g}$ with respect to the Hausdorff metric, and $\mathcal{H}^1(\Gamma_\beta) \to \mathcal{H}^1(\mathcal{B}_g)$. Furthermore $\sqrt{\beta}(f - \hat{g})$ converges to 0 in $L^2(\Omega)$.*

We mean by this that for any $\epsilon > 0$ there exists $\beta' > \infty$ such that if $\beta > \beta'$ and $\Gamma_\beta$ is a minimizer of $E$ (or $E_0$ as the case may be), for some $g \in \Upsilon(\beta)$, then $d_H(\Gamma_\beta, \mathcal{B}_g) < \epsilon$ and $|\mathcal{H}^1(\Gamma_\beta) - \mathcal{H}^1(\mathcal{B}_g)| < \epsilon$. Note that this shows the variational formulation to be asymptotically faithful to the piecewise smooth assumption. In the limit as $\beta \to \infty$ the geometry of the boundaries $\Gamma_\beta$ becomes essentially unrestricted. It should also be noted that in the limit $\beta \to \infty$ any discontinuity in the image will be detected, i.e. edges on all scales are recovered. The convergence depends strongly on removal or rescaling of the noise (the rates for which are tight). To convert this into an algorithm we need to re-interpret the rescaling of the noise.

# 5 A Scaling Algorithm

It was pointed out in Section 1 that in general there is a trade off between the accuracy of localization of boundaries found by the variational method and the total quantity of boundary admitted into the solution. (This appears to be true for most edge detection techniques.) Suppose the goal of the segmentation was to recover objects only above a certain scale. Consider Figure 2 for example (assume the domain is very large), if one were trying to find objects on the scale of the larger square and not those on the scale of the smaller square by minimizing $E_0$ with appropriately chosen parameters, then it is necessary to incur an error at the corners of at least $(\sqrt{2}-1)b$ as illustrated in Figure 2.

Now, the limit theorem discussed in the preceding section state that as $\beta$ tends to infinity the boundaries which are found by solving the variational problem converge to the correct ones (i.e. the discontinuity set of the image) with respect to the Hausdorff metric. As such, these theorem do not provide us with any means of circumventing the scale/accuracy trade–off because they state that the limit of the minimizing $\Gamma$ includes all of the discontinuity set of the image. We ask whether it is possible to take the limits required by the limit theorem while avoiding the attendant problem of introducing more and more boundary into the solution.

In response to this question we sketch an algorithm which requires within it two key operations or procedures:

**P1:** *The Minimization of $E$ (or $E_0$) to produce $f$ and $\Gamma$ with the parameters $\alpha$ and $\beta$ as input variables*

**P2:** *The updating or altering of the image by smoothing outside some neighborhood of $\Gamma$ and updating the parameters which provide the data for P1 for resolution at smaller scale.*

The interaction between these two procedures is illustrated in Figure 3. An execution of the algorithm begins by minimizing $E$ with the function $g$ set to the original data and the parameters $\alpha$ and $\beta$ chosen to provide edge detection on the desired scale. Once minimal $f$ and $\Gamma$ have been determined they are used to alter the function $g$. The new $g$ is formed by "smoothing" the original $g$ outside of some neighborhood of $\Gamma$ while leaving it unchanged inside the neighborhood. Since $f$ is a piecewise smooth approximation to $g$ which respects the edges $\Gamma$ a simple smoothing technique might be to take a convex combination of $f$ and $g$ (this is what was done in the simulations). Simultaneously, the parameters $\beta$ and $\alpha$ are assigned new values such as would be

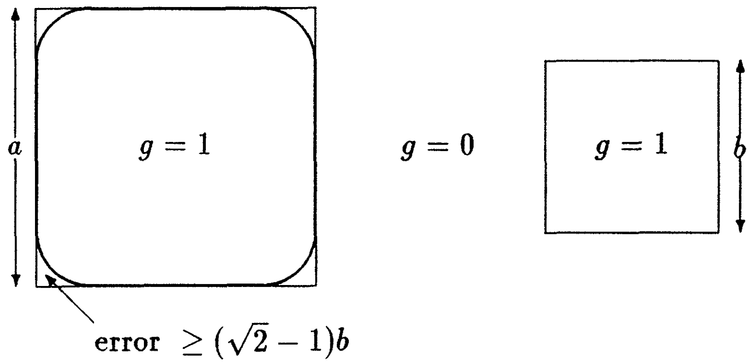


Figure 2: Segmentation of Two Squares with $a > \frac{\alpha}{\beta} > b$

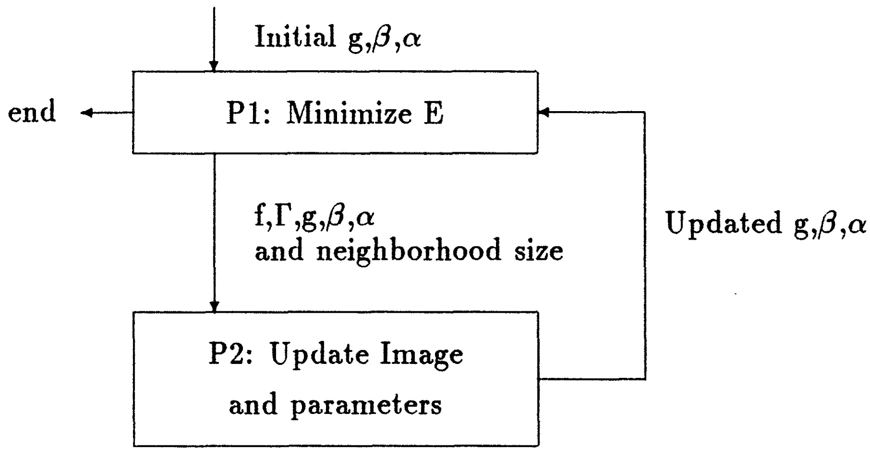


Figure 3: A Schematic for Scale Independent Segmentation

used with the original data if the edge detection were desired on a finer scale. We then re–solve the problem of minimizing $E$, the difference this time being that we use the updated image and parameter values. The hope is that we should detect essentially the same edges as before only now with finer resolution. This procedure can then be iterated on until sufficient accuracy is attained.

As stated the idea of the algorithm is very general and there remains within each procedure considerable flexibility. How one does the smoothing or chooses neighborhoods is unspecified; also, we have not stated how we will find a minimizer of the variational problem. Of course, in order for the algorithm to work the various parameters and operations must be coordinated properly. It is at this point that the limit theorem become useful. The limit theorem effectively governs the evolution of the parameters.

If we remove the smoothing of the image from procedure **P2** then the algorithm would constitute an explicit taking of the same limit as taken in the limit theorem. The deficiency with the limit theorem with regard to their direct application in edge detection, i.e. the reason why this will not be effective is the fact that the limit theorem predicts the convergence of the solution $\Gamma$ to the entire discontinuity set of the image. If the image was noisy this could effectively result in boundaries being put essentially everywhere. The smoothing of the image plays a role in the algorithm analogous to that played by the rescaling of the noise in the limit theorem. Small features and noise are smoothed out, but at the same time the detail needed for accurate localization of the large scale boundaries is retained.

Recall that in the limit theorem the amount of noise which can be allowed while retaining convergence of the boundaries has been quantified. In particular we considered sequences $\beta_n, g_n$ such that $\{g_n\}$ converges to the ideal image $g_\infty$ according to $\lim_{n\to\infty} \beta_n \int_\Omega (g_n - g_\infty)^2 = 0$, where $\Omega$ is the domain of the image, and $\lim_{n\to\infty} \sqrt{\beta_n} \|(1 - \chi_{[\mathcal{S}_{g_\infty}]_\epsilon})(g_n - g_\infty)\|_\infty = 0$ for any $\epsilon > 0$ (where $\mathcal{S}_{g_\infty}$ is the support set of the discontinuities of $g_\infty$). Now, let $\Phi_r$ represent an arbitrary smearing operator such that the value of the result at a point $x \in \Omega$ lies within the range taken by argument in $B_r(x)$ and let $w$ represent an arbitrary function in the unit ball of $L_\infty(\Omega)$. We argued in Section 4, under some mild regularity assumptions on $g_\infty$, that the convergence conditions were satisfied if we could represent the $g_n$ by,

$$g_n = \Phi_{r_n}(g_\infty) + \vartheta_n w_n$$

with $r_n$ being a sequence of constants satisfying

$$\lim_{n\to\infty} \beta_n r_n = 0 \tag{5}$$

and $\vartheta_n$ being a sequence of constants satisfying

$$\lim_{n\to\infty} \sqrt{\beta_n} \vartheta_n = 0. \tag{6}$$

The idea of the algorithm is to produce $g_\infty$ by generating the $g_n$.

The algorithm produces sequences $\{f_n\},\{\Gamma_n\},\{g_n\},\{\beta_n\}$ such that $f_n$ and $\Gamma_n$ are found by minimizing $E(\beta_n, g_n)$ (or $E_0$), and it also produces a scalar sequence $\{u_n\}$ where $u_n$ denotes the size of the neighborhood within which the smoothing is suppressed at stage $n$. The algorithm is initialized by setting $g_0 = g$ where $g$ is the original data and by choosing $u_0$ . The quantities, $g_n, \beta_n$ and $u_n$ are defined according to the following schedule,

$$\begin{aligned} g_{n+1} &= g_n + h_n(x)\,\epsilon(f_n - g_n) \\ \beta_{n+1} &= (1-\epsilon)^{-2}\beta_n \\ u_{n+1} &= (1-\epsilon)^2 u_n = \frac{\beta_0}{\beta_n} u_0 \end{aligned}$$

The function $h_n(x)$ controls the spatial dependence of the smoothing which is effected by partially replacing $g_n$ with $f_n$. For simplicity, and to be consistent with our simulations we consider setting $h_n$ equal to $1 - \chi_{[\Gamma_n]_{u_n}}$. The parameter $u_0$ represents an estimate of the error in the initial boundary locations. For the simulations $u_0$ has been selected heuristically (see also Section 6). The formalism just presented is a discrete one i.e. a discrete sequence of images and parameters is produced. One can in principle also vary the parameter $n$ continuously and represent the algorithm as differential equations. In this context we can say that $\epsilon$ represents the step size of the algorithm.

One can argue heuristically why the set of boundaries should remain essentially unchanged throughout the iterations of the algorithm for reasonable small values of $u_0$. For simplicity, consider what would happen if we set $h_n(x) = 1$ for all $n$ i.e. $u_0 = 0$. It is not very difficult to check that the solution $f_0, \Gamma_0$ would be a local minimum for the functional $E(g_1, \beta_1)$ i.e. if we consider small local variations in the boundary we find that the original locations are optimal. We conjecture that $f_0$ is in fact a global minimum and this can easily be seen for certain special cases such as the image of a square. Because the original solution is in some sense being reinforced by the feedback we expect that it becomes a 'deeper' minima then previously. This then lends a certain robustness to the algorithm.

There are several observations which should be make concerning the proposed algorithm. First, since the algorithm requires minimizing $E$ many times the computational load will obviously be higher than simply minimizing $E$. The algorithm has been proposed to demonstrate the possibility of overcoming the scale/accuracy tradeoff inherent in the original model. It has been structured to parallel as much as possible the asymptotic theorem. There are several improvements and speed–ups that can be considered. Many approaches to minimizing $E$ would admit at each stage the possibility of using solution from one step as an initial condition for the next, thus reducing time required to find the new solution. Another possibility would be to modify the parameter $\beta$ locally within the image domain and to dispense with the smoothing step entirely. This could also substantially save on computation. A second observation concerns the magnitude of $\epsilon$. Because of the manner in which we spatially control the smoothing of the data we cannot afford to make $\epsilon$ large. The sharp cutoff of the region in which smooth (by convex combination) could create small discontinuities in the data of order $\epsilon$. It is important therefore in our particular implementation that $\epsilon$ not be too large. Having smaller $\epsilon$ on the other hand forces more iterations of the sequence P1–P2 in order to reach the desired resolution. A simple improvement might be force the feedback to have vary smoothly spatially.

Although these ideas and others could conceivably improve the computational viability of the proposed algorithm substantially we have chosen to defer these refinements to a later time. Our goal here is to build on the asymptotic results for the variational formulation in the simplist way possible demonstrating the potential for improving on the formulation via certain dynamics in scale.

## 5.1 An Example

To illustrate the algorithm we consider minimizing $E_0$ when the image is that of a light square on a dark background such as in Figure 4. Let $S$ be the support set of the square and define $C = \Omega \backslash S$. For simplicity assume that $\Omega$ is a set much larger then $S$ such that $\mathrm{dist}(\partial\Omega, \partial S) \gg a$. In this case the $\Gamma$ which minimizes $E_0$ is either the emptyset or the contour indicated in Figure 4 which is symmetric about the center of $S$ and consists of 4 straight line segments of length $2(a-r)$ one centered on each side of $S$ and the 4 components one obtains from intersecting a circle of radius $r$ with the

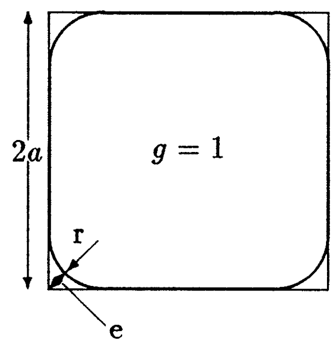


Figure 4: Segmentation of a Square

four quadrants in $\Re^2$. Obviously $r \leq a$. We will denote such a $\Gamma$ by $\Gamma^r$. The following two expressions are easily derived.

$$\begin{aligned} E_0(\Gamma^r) &= \beta \frac{(4-\pi)r^2|C|}{|C|+(4-\pi)r^2} + 8(a-r) + 2\pi r \\ E_0(\emptyset) &= \beta \frac{4a^2|C|}{4a^2+|C|} \end{aligned}$$

For simplicity consider the case $|C| = \infty$. By minimizing $E_0(\Gamma^r)$ over $r$ one obtains $r = \beta^{-1}$. Comparing the two possible solutions one discovers that there is a threshold $t = \frac{2+\sqrt{\pi}}{2}$ such that $\Gamma = \emptyset$ is optimal when $a\beta \leq t$ and $\Gamma = \Gamma^{\beta^{-1}}$ is optimal when $a\beta \geq t$. From this we conclude that the maximum error, i.e. the maximum possible value of,

$$e = d_H(\partial S, \Gamma) = (\sqrt{2}-1)r,$$

which can occur when $\Gamma$ is not the empty set is $\frac{(\sqrt{2}-1)a}{t}$. In general $e$ is proportional to $\beta^{-1}$ thus as long as $u_0 \geq (\sqrt{2}-1)\beta_0^{-1}$ then the $\Gamma_n$ produced by the algorithm will either converge to $\partial S$ in Hausdorff metric as $n \to \infty$ or will equal the empty set for all $n$. When $a\beta = t$ then either $\Gamma_n = \emptyset$ for all $n$ or for some $n$ $\Gamma_n = \Gamma^{\beta_n^{-1}}$ and then $\Gamma_m = \Gamma^{\beta_m^{-1}}$ for all $m \geq n$. If $0 < u_0 < (\sqrt{2}-1)\beta_0^{-1}$ and $a\beta > t$ then the $\Gamma_n$ will not converge to $\partial S$ however $d_H(\Gamma_n, \partial S)$ will be decreasing.

# 6 A Γ–Convergent Approximation

A idea which plays an important role in our study and implementation of the variational approach to the edge detection problem is the notion of Γ-*convergence* due to E. De Giorgi. The same concept was developed independently in France under the name *epi-convergence* by H. Attouch [6]. It concerns variational convergence, i.e. the approximation of one variational problem by another. In this section we provide a definition of Γ-convergence, state some of its basic properties and provide an application to our problem.

Let $(S, d)$ be a separable metric space and let $F_n : S \to [0, +\infty]$ be functions. We say $F_n$ $\Gamma(S)$ – converges to $F : S \to [0, +\infty]$ if the following two conditions hold for all $x \in S$,

$$\begin{aligned} \forall x_n \to x \quad & \liminf_{n\to\infty} F_n(x_n) \geq F(x) \\ \text{and } \exists x_n \to x \quad & \liminf_{n\to\infty} F_n(x_n) \leq F(x) \end{aligned}$$

The limit $F$ when it exists is unique and lower–semicontinuous. The following proposition characterizes the main properties of Γ–convergence.

**Proposition 6** *(see [5] for example) Assume that $F_n$ $\Gamma(S)$–converges to $F$. Then, the following statements hold.*
*(i) $F_n + G$ $\Gamma(S)$–converges to $F + G$ for every continuous function $G : S \to \Re$.*
*(ii) Let $t_n \downarrow 0$. Then, every cluster point of the sequence of sets*

$$\{x \in S : F_n(x) \leq \inf_S F_n + t_n\}$$

*minimizes $F$.*
*(iii) Assume that the functions $F_n$ are lower semicontinuous and for every $t \in [0, \infty)$ and there exists a compact set $K_t \subset S$ with*

$$\{x \in S : F_n(x) \leq t\} \subset K_t \ \forall n \in \aleph$$

*Then, the functions $F_n$ have minimizers in $S$, and any sequence $x_n$ of minimizers of $F_n$ admits subsequences converging to some minimizer $F$.*

A significant contribution of Γ convergence to our problem is that via a Γ-convergent approximation one can represent the boundaries by a function defined on the same domain as the image. The functional given below, for example is a modified version of an approximation scheme due to Ambrosio and Tortorelli [5].

$$E^n(f,v) = \beta \int_\Omega (f-g)^2 + \int_\Omega e^{-nv^2}|\nabla f|^2 + \alpha \int_\Omega (e^{-nv^2}|\nabla v|^2 + \frac{n^2v^2}{16})$$

In this approximation one replaces the set $\Gamma \subset \Omega$ with a function $v : \Omega \to [0,\infty]$. The location of boundaries is in general given by $e^{-nv^2} \simeq 0$.

It was proved in [5] that the sequence of functionals very similar to $\{E^n\}$ (with $(1-v^2)^n$ replacing $e^{-nv^2}$), Γ-*converges* to $E$. This proof can be modified and extended to include the functional given above. The metric space in this case is a product of function spaces elements of which are pairs $(f,v)$.

The functionals $E^n$ can be discretized by finite elements thus obtaining a simple representation of the edges suitable for computation. Our simulations employ a gradient descent to find local minimizers of $E^n$. This approach closely resembles the an-isotropic filtering approach due to Perona and Malik [30]. In our case however we obtain an explicit representation of the boundary which would be useful for subsequent processing.

The function $v$ appearing in the Γ-convergent approximation is such that $1-e^{-nv^2}$ has the appearance of a smoothed neighborhood of the boundaries. The boundaries themselves can be identified with those locations where $e^{-nv^2} \simeq 0$. It can be shown (see [31]) that by thresholding $e^{-nv^2}$ at a level $t$ one obtains a set which can be interpreted as a neighborhood of Γ of size $u_t$ where,

$$\frac{nu_t}{2} \simeq \int_{-\ln\sqrt{t}}^{\infty} \frac{\exp -r}{r}\, dr \tag{7}$$

This means that we can define a neighborhood of Γ to be $\{x : e^{-nv^2} < t\}$ and that this should approximate $[\Gamma]_{u_t}$ where $u_t \simeq \frac{2}{n}\int_{-\ln\sqrt{t}}^{\infty} \frac{\exp -r}{r}\, dr$. Thus sublevel sets of $e^{-nv^2}$ can provide an approximation to the neighborhoods of Γ required by the algorithm outlined in the previous section. For the simulations presented in this paper $t$ has been set to $\frac{1}{2}$. The equation 7 then yields the relation $u_{\frac{1}{2}} \simeq \frac{1.6}{n}$.

# 7 Computation

In this section we describe how the ideas presented in the Section 5 are refined in the piecewise smooth case by introducing the Γ–convergent sequence of approximations to E as a means to generate both the boundary locations and their neighborhoods.

The procedure *P1* of the algorithm will now be implemented by the minimization of the following functional over $f$ and $v$,

$$E(f,v,g,\beta,\alpha,n) = \beta\int_\Omega (f-g)^2 + \int_\Omega (1-v^2)^n|\nabla f|^2 + \alpha\int_\Omega((1-v^2)^n|\nabla v|^2 + \frac{n^2v^2}{16}).$$

The procedure P2 will be implemented by altering the other variables. One can write down the Euler-Lagrange equations for $v$ and $f$ associated with this functional and then the parabolic equations which would be associated with a descent algorithm. For $E$ as above we obtain,

$$\begin{aligned}\frac{\partial f}{\partial t} &= c_f\left[\nabla\cdot(e^{-nv^2}\nabla f) - \beta(f-g)\right]\\ \frac{\partial v}{\partial t} &= c_v\left[\alpha\nabla\cdot(e^{-nv^2}\nabla v) + n(|\nabla f|^2 + \alpha|\nabla v|^2)e^{-nv^2}v - \frac{\alpha n^2}{16}v\right]\end{aligned}$$

with Neumann boundary conditions. The parameters $c_f$ and $c_v$ are arbitrary positive constants which controls the relative rate of descent. These equations resemble the non-linear filtering scheme of Perona and Malik [30]. The differences are worth noting. The equations presented here have a term dependent on $g$; unlike Perona and Malik we do not necessarily converge to a piecewise constant function. Also the control of the conductivity associated with the diffusion of the image is effected by the function $v$ rather than by an explicit function of the magnitude of gradient of $f$. The function $v$ is governed by another partial differential equation whose driving term is related to the gradient of $f$.

Since the functional is not convex in $v$ we do not expect to always reach a global minimum by a descent method. Also the dependence of the solution on the initial conditions, and the constants $c_f$ and $c_v$ is significant. For example if $c_f$ is of much smaller than $c_v$ and we initialize $f$ by setting it equal to $g$ then the system of equations places emphasis on the height of edges. The evolution of $v$ is initially governed essentially by $|\nabla g|^2$ and thus will tend to place boundaries at edges in the image largely ignoring the size of the feature of which the edge is a boundary. Conversely if $c_v$ is much smaller than $c_f$ or if $f$ is initialized by a smoothed version of $g$ then the geometry of the features will play an important role since a smaller features will produce a smaller gradients in $f$ even for the same height of the discontinuity. Thus boundaries will be more likely to appear at the edges of larger objects, everything else being equal. Consider for example the image of the two squares Figure 2. Suppose we initialize the descent equations with $f$ set to the solution of $\Delta f = \beta(f - g)$ and $\beta$ satisfies $\frac{1}{\sqrt{\beta}} \simeq b$. The smaller square (with sides of length $b$) will have a smaller effect on the initial $f$. That is, the gradient of the initial $f$ will be smaller near the edges of the smaller square than near those of the larger square. Consequently if $\alpha$ is chosen appropriately the equations above will have a greater tendency to increase $v$ i.e. to place an edge near the edges of the larger square than near those of the smaller.

Whatever choice is made concerning the selection of the various parameters associated with the computation it is important that they be kept consistent throughout the iterations of the algorithm. The intent of the algorithm is to refine the boundaries found in the early stages, not to radically change them. To achieve this a consistent computational approach is necessary. We mentioned in Section 5.1.1 that the feedback will have the effect of reinforcing the solution found during the earlier stages of the algorithm, tending to make that solution a 'deeper' minima than initially. The same argument holds for the local minima which will be found by a computational procedure such as we have described. As long as the computation remains consistent from iteration to iteration then the feedback should make the algorithm more robust in the sense that it encourages the finding of essentially the same solution.

## 7.1 Discretization

In this section a particular discretization the functionals $E_n$ is presented. This formulation may have value in that it lends itself to possibility of an analog hardware VLSI implementation such as considered in Harris et. al. [21].

For the simulations presented in this paper $f, g$ and $v$ were discretized by finite elements in a manner described below. Discrete versions of $f$ and $g$ are defined on a square lattice with lattice constant $\delta$ while the discrete version of $v$ is defined on one twice as dense. This is not necessary but it facilitates the implementation of the discrete problem. Figure 5 indicates the assignment of lattice values. For convenience the variables are labeled as in standard matrix notation; thus $f_{i,j}$ denotes the variable associated with $f$ at the lattice location row $i$, column $j$. To keep the notation for the function $v$ consistent with the lattice used for $f$ and $g$ the variables associated with $v$ are partitioned into two sets, $vh$ and $vv$, corresponding in some sense to horizontal and vertical edge elements. The assignments are as in Figure 5. A suitable discrete version of $E$ in terms of these variables is the following,

$$\begin{aligned} E &= \delta^2 \sum_{i,j} (f_{i,j} - g_{i,j})^2 + \\ & \sum_{i,j} (f_{i,j} - f_{i+1,j})^2 e^{-n\, vv_{i,j}^2} + (f_{i,j} - f_{i,j+1})^2 e^{-n\, vh_{i,j}^2} \\ & + \frac{\alpha}{2} \left( \sum_{i,j} e^{-n\, vv_{i,j}^2} \sum_{(i',j') \in \mathcal{N}_v(i,j)} (vv_{i,j} - vh_{i',j'})^2 + \sum_{i,j} e^{-n\, vh_{i,j}} \sum_{(i',j') \in \mathcal{N}_h(i,j)} (vh_{i,j} - vv_{i',j'})^2 \right) \\ & + \alpha \left( \sum_{(i',j') \in \mathcal{N}_v(i,j)} e^{-n\, vv_{i,j}^2} (vv_{i,j} - vh_{i',j'}) + \sum_{(i',j') \in \mathcal{N}_h(i,j)} e^{-n\, vh_{i',j'}^2} (vv_{i,j} - vh_{i',j'}) \right) \\ & + \frac{\alpha}{2} \frac{\delta^2 n^2}{16} \sum (vv_{i,j}^2 + vh_{i,j}^2) \end{aligned}$$

where $\mathcal{N}_h(i,j)$ is the set of indices for the nearest vertical edge element neighbors of $vh_{i,j}$ and similarly $\mathcal{N}_v(i,j)$ is the set of indices for the nearest horizontal edge element neighbors of $vv_{i,j}$. A discrete form of the Euler–Lagrange equations for this system is found by differentiating the expression above with respect to the various elements. Associated with this one obtains a gradient descent equations of the following form,

$$f_{i,j}^{t+1} - f_{i,j}^{t} = -c_f \frac{\partial}{\partial f_{i,j}} E \tag{8}$$

$$vv_{i,j}^{t+1} - vv_{i,j}^{t} = -c_v \frac{\partial}{\partial vv_{i,j}} E \tag{9}$$

$$vh_{i,j}^{t+1} - vh_{i,j}^{t} = -c_v \frac{\partial}{\partial vh_{i,j}} E$$

where the variables $c_f$ and $c_v$ control the stepsizes of the algorithm and $t$ denotes 'time' or steps in the algorithm. For our simulations we updated $f$ by using the

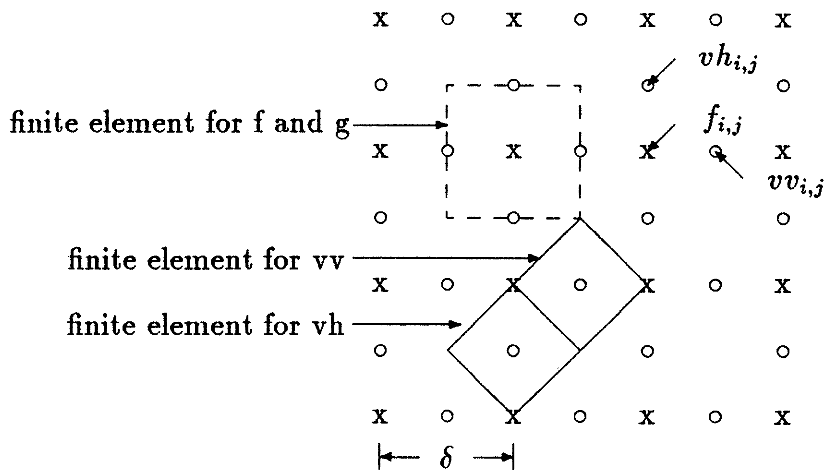


Figure 5: Lattice Variables and Finite Elements

assignment,

$$c_f \;=\; 2\left(\delta^2\beta + e^{-n\,vv_{i,j}^2} + e^{-n\,vv_{i-1,j}^2} + e^{-n\,vh_{i,j}} + e^{-n\,vh_{i,j-1}}\right)^{-1}$$

in keeping with standard relaxation algorithms. The constant $c_v$ should be scaled with $n$. The details can be found in [31]; $c_v$ should be scaled as $n^{-1.5}$.

In general some upper limit on $n$ must be imposed for a given lattice spacing. Consider the behavior of $v$ in the $\Gamma$–convergent approximation for large $n$. For each point $i$ in the array there is a term in the cost proportional to $n^2v_i^2$. Now, as $n$ becomes large it is necessary for the cost to remain bounded that $v_i$ decrease like $\frac{1}{n}$. However $\lim_{n\to\infty}(1-\frac{K}{n^2})^n = 1$ for any $K < \infty$, i.e. as $n$ tends to infinity all boundaries will be removed. There is a secondary positive feedback effect which aggravates this problem. An increase in $e^{-nv^2}$ results in an increase in the smoothness of $f$ i.e. a reduction in $|\nabla f|$. This causes yet a further increase in $e^{-nv^2}$. Thus it is apparent that the discretized version of this approximation becomes unreliable for large $n$. Another reason for keeping $n$ small is that if the wider the 'support' of the edges in terms of the lattice constant the smaller the effect of the discretization. In particular the rotational invariance of the continuum formulation is better retained.

# 8 Simulation Results

We have simulated the algorithm developed in this Section on the image shown in Figure 6. The size of the image is $230 \times 216$ pixels. The image is a well known one but the version used is an unusally noisy one. It was taken from a bitmap i.e. a $0\backslash 1$ image of size $920 \times 864$ bits and each $4 \times 4$ block was mapped onto a single pixel. The value of the image $g$ at a given pixel is proportional to the number of 1's in the associated $4 \times 4$ block and is scaled so that the range of $g$ lies within $[0, 2]$. The image is thus quite noisy. Our displays are also bitmaps where we have reversed this procedure. Thus our resolution is essentially 4 bits per pixel although the computations were done using 64 bit floating arithmetic. This image is a rather problematic one for edge detection because many of the edges are blurred and there are regions of texture. We have performed the simulation for several scales. For one scale which we denoted 'Scale 2' we have sampled the functions, $g$ and $e^{-nv^2}$ at various stages of the algorithm. What is worth noticing is how the fine detail such as sharp corners and T–junctions are recovered in the final stages. This can be seen particularly in the details of the eyes. Also, as predicted the global properties of the solution remain essentially unchanged. Because the edges themselves are blurred and noisy at fine scales we begin to observe multiple edges. Even though $\beta$ is increased by a factor of 13 the particular boundaries which are found do not change except in the fine detail of the localization. (In other experiments $\beta$ was allowed to increase over a much larger range and similar results were obtained.) In figure 9 we display $e^{-nv^2}$ from solutions obtained for several scales. Notice that the set of boundaries found is essentially monotonic in scale.

| | |
|---|---|
| $\delta$ | 0.05 |
| Range of $g$ | $[0.0, 2.0]$ |
| Stepsize $c_v$ | $\frac{0.3}{\alpha n^{1.5}}$ |
| $\epsilon$ | 0.05 |

Table 1: General Parameters for 'Lenna' Simulations

| 'Scale' | $\alpha$ | Initial $\beta$ | Final $\beta$ | Initial $n$ | Final $n$ |
|---|---|---|---|---|---|
| 1 | 0.008 | 2.0 | 26.0 | 3.0 | 39.0 |
| 2 | 0.006 | 3.0 | 39.0 | 3.0 | 39.0 |
| 3 | 0.0045 | 7.0 | 91.0 | 8.0 | 104.0 |
| 4 | 0.003 | 10.0 | 130.0 | 10.0 | 130.0 |

Table 2: Parameters for Simulation 'Lenna: All Scales'

| Sample Number | $\alpha$ | $\beta$ | n |
|---|---|---|---|
| 1 | 0.006 | 3.0 | 3.0 |
| 2 | 0.006 | 5.0 | 5.0 |
| 3 | 0.006 | 14.0 | 14.0 |
| 4 | 0.006 | 39.0 | 39.0 |

Table 3: Parameters for Simulation 'Lenna: Scale 2'

Figure 6: "Lenna" Data Used for Simulations

Sample 1

Sample 2

Sample 3

Sample 4

Figure 7: Lenna Scale 2: Samples of $e^{-nv^2}$

Sample 1 Sample 4

Figure 8: Lenna Scale 2: Samples of Updated $g$

Scale 2 Scale 4

Figure 9: Original Image with Edges


Figure 10: Boundaries at Various Scales

# References

[1] R. Adams, *Sobolev Spaces*
Academic Press N.Y. 1975

[2] L. Ambrosio, Compactness for a Special Class of Functions of Bounded Variation
To appear.

[3] L. Ambrosio, Existence Theory for a New Class of Variational Problems
To appear.

[4] L. Ambrosio, Variational Problems on SBV, Center for Intelligent Control Systems Report CICS-P-86. MIT. 1989

[5] L. Ambrosio, V. Tortorelli Approximation of functionals depending on jumps by elliptic functionals via $\Gamma$-convergence.
To appear.

[6] H. Attouch, *Variational Convergence for Functions and Operators*
Pitman Publishing Inc., 1984.

[7] A. Blake and A. Zisserman, *Visual Reconstruction*
MIT Press, Cambridge, 1987

[8] J. Canny, *Finding edges and lines in images.* Tech. Rep. 720, MIT A.I.Lab., 1983.

[9] G. Dal Maso, J-M. Morel, and S. Solimini, A Variational Method in Image Segmentation: Existence and Approximation Results
S.I.S.S.A. 48 M, April 1989.

[10] E. De Giorgi, M. Carriero, and A. Leaci, Existence Theorem for a Minimum Problem with Free Discontinuity Set
To appear.

[11] K.J. Falconer, *The Geometry of Fractal Sets*
Cambridge University Press, 1985

[12] H. Federer, *Geometric Measure Theory*
Springer-Verlag, 1969

[13] E.B. Gamble and T. Poggio Visual integration and detection of discontinuities: The key role of intensity edges.
A.I. Memo No. 970, MIT A.I. Lab, 1987.

[14] D. Geiger, and F. Girosi Parallel and deterministic algorithms for MRFs: surface reconstruction and integration
MIT A.I.Lab., Memo No.1114, June 1989.

[15] D. Geiger, and A. Yuille A Common Framework for Image Segmentation
Harvard Robotics Laboratory, Tech. Rep. no. 89-7.

[16] D. Geiger, and F. Girosi An optimal scale for edge detection.
Proceedings IJCAI, Aug. 1987.

[17] S. Geman, and D. Geman Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images
*IEEE Trans. Pattern Analysis and Machine Intelligence* 6, 721-741, 1984

[18] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation.*
Birkhäuser, Basel, 1983.

[19] B.K.T. Horn, *Robot Vision*
MIT Press, Cambridge, 1986

[20] W.E.L. Grimson *From Images to Surfaces*
MIT Press, Cambridge, 1981

[21] J. Harris, C. Koch, J. Luo, J. Wyatt Resistive Fuses: Analog Hardware for Detecting Discontinuities in Early Vision,
Analog VLSI Implementations of Neural Systems, Kluwer, Norwell, MA., pp. 27–55, 1989

[22] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing
Science, 220:219-227, 1983.

[23] S.R. Kulkarni, S.K. Mitter, T.J. Richardson, An Existence Result and Lattice Approximations for a Variational Problem Arising in Computer Vision, Proceedings of the Signal Processing Workshop, Inst. for Mathematics and Applications, U. of Minnesota, 1989. Also, Center for Intelligent Control Systems Report CICS-P-123. MIT. 1989.

[24] D. Marr, *Vision*
W.H. Freeman an Co. 1982

[25] D. Marr, and E. Hildreth Theory of edge detection
Proc. Roy. Soc. Lond. B, 207, 187–217, 1980

[26] J.L. Marroquin, Probabilistic Solution of Inverse Problems.
Ph.D. Thesis, Dept. of E.E.C.S., MIT, 1985

[27] J-M. Morel, and S. Solimini, Segmentation of Images by Variational Methods: a Constructive Approach.
Revista Matematica de la Universidad Complutense de Madrid, Vol. 1, num 1,2,3; 1988.

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

[29] D. Mumford, and J. Shah, Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems.
*Communications on Pure and Applied Mathematics.* vol. XLII No. 4 July, 1989. pgs. 577-685.

[30] P. Perona and J. Malik, Scale-Space and Edge Detection Using Anisotropic Diffusion
Report No. UCB/CSD 88/483, EECS UCBerkeley, December 1988

[31] T.J. Richardson, *Scale Independent Piecewise Smooth Segmentation of Images via Variational Methods.* Ph.D. Thesis Dept. of E.E. & C.S., M.I.T. 1990.

[32] C.A. Rogers, *Hausdorff Measure.* Cambridge University Press. 1970

[33] A. Rosenfeld and M. Thurston. Edge and curve dectection for visual scene analysis. *IEEE Trans. on Computers,* C-20, pgs. 562-569, May 1971

[34] W. Rudin, *Functional Analysis.* McGraw Hill. 1973

[35] J. Serra, *Image Analysis and Mathematical Morphology* Academic Press Inc., 1982.

[36] J. Shah, Segmentation by Minimizing Functionals: Smoothing Properties. To appear.

[37] Y. Wang, Ph.D. Thesis Harvard University Math. Dept. 1989.

[38] A. Witkin, Scale-space filtering, *International Joint Conference on Artificial Intelligence, Karlsruhe* pgs. 1019-1021, 1983

[39] A. Yuille and T. Poggio, Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis, and machine Intelligence*, 8 , Jan. 1986.