SCALE INDEPENDENT PIECEWISE SMOOTH SEGMENTATION OF IMAGES VIA VARIATIONAL METHODS

by

Thomas Joseph Richardson

B.Ap.Sc. University of Toronto
(1983)

M.Ap.Sc. University of Toronto
(1986)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY
IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 1990

Signature of Author ______________________________
Department of Electrical Engineering and Computer Science
January 12, 1990

Certified by ______________________________
Sanjoy K. Mitter
Thesis Supervisor

Accepted by ______________________________
Arthur C. Smith
Chairman, Departmental Commitee on Graduate Students

SCALE INDEPENDENT PIECEWISE SMOOTH SEGMENTATION OF IMAGES VIA VARIATIONAL METHODS

by

Thomas Joseph Richardson

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY
IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
JANUARY 1990

Abstract

Variational methods that include an explicit representation of edges have been introduced for image segmentation by Mumford and Shah. We develop a paradigm that improves on these methods to allow segmentation on different scales while retaining the accuracy usually attained only for the finest scale. The paradigm leads to several algorithms requiring scheduling of the parameters of the variational formulation and feedback from the approximating image into the data. The feedback rates and the schedule are governed by several limit theorems which have been attained for the variational model. The limit theorems demonstrate an asymptotic fidelity of the variational model to a more general piecewise smooth model. An efficient computational scheme is built on a sequence of approximating problems converging to the variational problem in the sense of Γ-convergence.

Thesis Supervisor: Sanjoy K. Mitter
Title: Professor of Electrical Engineering

To Dr. A. A. Curtin,
for the inspiration.

Acknowledgements

To my supervisor Prof. Sanjoy K. Mitter I owe a debt of gratitude for his unwavering support in all aspects of my work, for a trip to Italy, where this thesis began and for instilling in me some of his sense for the wider vision.

I thank the rest my thesis committee, Prof. David Mumford, Prof. W. Gilbert Strang and Prof. John N. Tsitsiklis for their questions and comments.

Many other people have contributed significantly to this thesis. Those who were especially influential and to whom I offer many thanks are Luigi Ambrosio, Dr. Vivek Borkar, Sanjeev Kulkarni, Dr. Ofer Zeitouni, and Prof. Jayant Shah.

I owe the friends I have made at M.I.T. a great deal. Thanks largely to them I will remember my life here as a full one. Those whose companionship has been of particularly great value to me include Dr. Jerome Abernathy, Karen Adler, Louise Alterman, Ken Chou, Mike Chin, Mark Crews, May Ku, Sanjeev Kulkarni, Daniel Chonghwan Lee, Dr. Bernard LeGoff, Dr. Zhi–Quan Luo, and Dr. Ofer Zeitouni.

I thank Dr. Chee-Seng Chow, Dr. Peter Doerschuk and Emre Telatar for all their help with the computer systems.

A special thanks is due to Kathleen O'Sullivan and the rest of the administrative staff for raising the quality of life in the lab and making the bureaucracy with which the lab deals largely invisible to the graduate student.

I thank my parents for their respective contributions to my education.

This work was in part supported by the Office of Naval Research under contract NOOO14-77-0532, by the Air Force Office of Scientific Research under contract AFOSR-85-0227 and by the Army Research Office under contracts DAAG-29-84-K-0005 and DAAL03–86–K–0171 (Center for Intelligent Control Systems). This funding was greatly appreciated.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

In the past twenty years or so there has arisen a body of work which addresses the problem of how to build a machine capable of processing visual information with a sophistication and versatility comparable to a human being. The human visual system is enormously complex and its equal could not possibly be developed in such a small span of time. As in all engineering or analytical efforts in order to comprehend the problem it has been necessary to decompose it into pieces of sufficiently small scope that quantifiable objects and relations can be identified. In vision a separation was made between *early vision* and higher level vision. High level vision problems are those problems that one would ultimately like to solve. Two examples are recognition of objects and three dimensional representations of an environment, suitable perhaps for path planning. A basic problem only slightly less remote than these is the problem of shape description. It is towards shape description that early vision was expected to contribute. The primary exponent of this line of thought was David Marr [24]. It is with early vision that this thesis is primarily concerned. Some of the standard early vision problems have reached a level of mathematical maturity admitting formal models amenable to quite detailed analysis. Much of the work presented in the thesis, in particular the limit theorems presented in Chapter 4 for the variational formulation of the segmentation problem are in this vein.

In the field of early vision there have arisen several basic problems which are considered fundamental. Below we briefly mention a few.

Shape from Shading: A curved surface which is illuminated from various sources with approximately uniform reflective properties will usually present to the viewer a gradually varying light intensity. *Shape from shading* is the problem of deducing the shape or three dimensional geometry of the surface from the varying light intensity.

Depth from Stereo: At any instant of time the normal human viewer has essentially two vantage points on a viewed scene, one for each eye. The disparity between the two views yields information on the distance of various objects from the viewer. A *depth from stereo* algorithm performs computations on a pair of images to determine a depth or distance map.

Motion Tracking: This problem is perhaps one of the most complex early vision problems. It deals with the question of movement of objects and how it is perceived from sequences of images.

Segmentation: The *segmentation problem* is essentially the problem of partitioning the image plane into coherent regions, corresponding to objects, say. What makes a region coherent is the central modeling question. People are capable of detecting regions of uniform texture as well as smoothly varying surfaces. This thesis deals primarily with the segmentation in the latter situation.

Recently there has been a desire to fuse the early vision problems. It is believed that each early vision problem alone is both too prone to error and too limited in scope to provide information which could be used for higher level problems such as recognition. This thesis shares this philosophy somewhat but the synthesis to which it is directed is more of a vertical one, between the basic low level problem and the issue of finding representations suitable for recognition. To reiterate, the function of early vision can be likened to signal processing; a signal is detected and computations are performed, largely independent of the particular signal or any purpose to which the information present in the signal might be put. The selection of the relevant information, representation for categorization or recognition is left to the domain of higher level problems. Perhaps this scenario does have an analog in many biological systems, however, it certainly is true that various regions of the retina perform different computational tasks and control of the direction of the eye depends on intentionality at the highest level.

A question worth consideration is; given the type of information early vision computation can yield, into what forms should the representation of the results be transformed to admit higher level problem solving. This is a very large and difficult problem and one that we do not intend to answer. Rather, we would like to emphasize one possible feature such intermediate representations might possess to considerable advantage, and, in the context of one early vision problem, namely piecewise smooth segmentation, show how the basic computation for such a representation can be approached. The primary structural form we envision the intermediate data structures to possess is that of a hierarchy. A simple example would be a tree or graph like structure in which nodes represent elements of the image or environment and links the relations between them. The ongoing development of computer languages has recently produced the so-called object oriented languages. These structures might also provide a convenient framework in which to represent a scene, its components, and the relations between them.

It is not the aim of this thesis to pursue any of these grand schemes. The thesis addresses in detail one issue, namely that of scale in the context of segmentation. It is the contention of the author that information such as one gets from scale-wise segmentation can be naturally represented in the above mentioned data structures. Furthermore such structures could provide domains for formulating and solving recognition problems. Also, in real time vision systems they could be used as a domain for performing computations to decide what low level information is likely to be relevant thereby admitting the possibility of resource management for the enormous early vision computational demands.

As has been entioned, the segmentation problem is that of partitioning the domain of an image into coherent regions. In this thesis we examine only the problem of segmenting based on intensity information. That is we examine the problem of approximating real valued functions by piecewise smooth functions. The interpretation of this abstract problem as segmentation of images based on intensity information is largely historical. The model can be used for any problem where a piecewise smooth approximation is desirable. The restriction to real valued as opposed to vector valued functions is not intrinsic. Much of the work in this thesis could be extended to vector valued functions. In any case we focus our attention on the problem of segmenting grey level images. The foundation of the segmentation problem in this context has been and remains edge detection. Edge detection alone is insufficient for segmentation because most techniques admit the possibility of finding edge fragments i.e. of not providing a partition. In this thesis we consider both segmentations by piecewise smooth and piecewise constant functions. A segmentation by a piecewise constant functions obviously will provide a true partition of the domain. Nevertheless we will refer to "segmentations" provided by piecewise smooth functions.

The issue of scale arises naturally in the segmentation problem. A decision must be made concerning what constitutes a useful segmentation. An important question, for example, is: into roughly what size blocks should the domain be segmented? All segmentation and edge detection techniques incorporate parameters which effect this decision. From the outset of early vision it has been realized that one would like to have segmentations on different scales [24] [32] [37], i.e. one should be able to find consistent segmentations on different scales obtaining both spatial and scale-wise segmentation of the image. One of the weaknesses of many segmentation techniques of the type examined in this thesis is the scale dependence of errors. The so called energy–based methods such as the Variational formulation and Markov random field formulations (see the next section) introduce ad hoc energy functionals whose minimizer represents the proposed segmentation. Generally speaking these functionals possess three terms. One is a fidelity term, forcing the solution to track the data somewhat. The second is a smoothing term guaranteeing the approximation is smooth, at least off of the boundaries. The third penalizes the boundaries themselves, controlling the quantity of boundary admitted into the solution. The functionals also possess parameters to selectively weight each term. In particular these parameters control the quantity of boundary in the solution and the degree of smoothing i.e. they control the scale of the segmentation. By the examination of special cases one can easily demonstrate the scale dependence of the errors in the localization of the boundary. The work presented in this thesis is built on a paradigm, which may be of more general interest, that allows one to overcome some of the weaknesses of such ad hoc models. The goal of the paradigm is to be able to provide for the segmenting of large scale features of the image while retaining the accuracy usually reserved only for much smaller scales.

In the context of our particular problem the paradigm has the following form. We consider idealized versions of our data, i.e. data for which we can state explicitly what we want our solution to be. For the segmentation problem we consider corrupted versions of an otherwise piecewise smooth image. The desired solution is the identification of the discontinuity set of the image with the boundaries and the recovery of underlying smooth image. We show that although the variational (energy based) solutions do not yield the desired solution for particular values of their parameters, these solutions will converge to the desired ones in an appropriate topology as the parameters tend to the limit of microscopic scale, provided we control the degree of corruption appropriately. We then develop an algorithm which on a small time scale resembles the minimizing of the energy functional but on a longer time scale changes the parameters of the functional in the direction of the microscopic limit. In order to prevent the resolving of microscopic detail, i.e. in order to retain only large scale features, as the limit is taken 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 boundary dependent smoothing) are governed by the convergence theorem mentioned above.

1.2 Some Segmentation Models and Techniques

In this section we outline some segmentation techniques with which our work is related. Three techniques for image segmentation and reconstruction based on intensity information which have recently gained considerable attention are, Markov Random Fields [17] [26] [14], Variational Formulations [7] [28] [29], and Non-linear Filtering

[30]. 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. In this section we give a sketch of these methods and also one older one, with which we begin.

One of the earliest ideas in image segmentation was the following. Roughly speaking the edges in an image g, which we think of as a function defined on some two dimensional domain Ω, will occur at the maxima of the gradient of the image. In general real images are noise corrupted and the calculation of the gradient is very sensitive to noise. The solution proposed by Canny [8] and Marr and Hildreth [25] was to smooth g by convolving it with a Gaussian kernel G of some width σ before differentiating. Because of the symmetry of the Gaussian kernel the laplacian of the smoothed image is the same as the convolution of the image with the laplacian of the kernel, thus the edges were modeled as the zero crossings of,

$$\nabla^2 G * g.$$

Now, as one increase σ the degree of smoothing increases and the quantity of edges decreases. This was interpreted as a variation of scale in the approximation to g and lead to a concept of a set of approximations to the image on different scales. This concept was referred to as *scale space representations* and is due to Witkin [37]. The locations of the zero crossings of $\nabla^2 G * g$ vary continuously with σ and one can thereby determine which boundaries found on the fine scale correspond with boundaries of large scale features. One of the difficulties with this approach is that the scales would be discretized in a computational scheme making the correspondence problem, between boundaries on different scales, difficult. The localization of the boundaries tends to be poor, especially on larger scales, because of the smoothing across boundaries in the image. Figure 1.1 illustrates the behavior of this approach for a one dimensional case. The basic idea of scale space is to use the localization achieved on a fine scale together with coarse scale representations to recover coarse

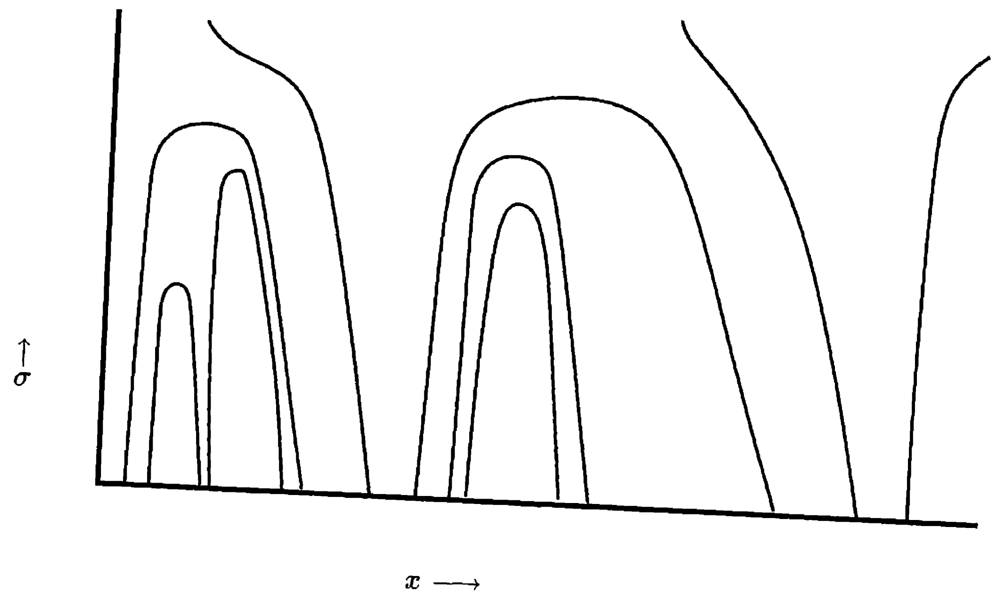


Figure 1.1: Zero-crossings of $\nabla^2 G(\sigma) * g$

scale features with good localization of the boundaries.

The three more recent approaches combat this difficulty (and others) by simultaneously placing boundaries and smoothing the image only off of the boundaries. Although it is true that the more recent methods have better localization then the gaussian smoothing approach it is also true that they make certain systematic errors, such as the rounding of corners and the distorting of t-junctions. Our approach allows one to find coarse scale features with accurate boundary locations directly, without first doing the fine scale computation. This has certain advantages. In a complex visual information processing task it might often be useful to manage early vision computations based on current representations of the environment; where to determine more detailed information can be decided based on coarse scale information. Accurate boundary locations on the coarse scale may facilitate recognition on the coarse scale and also, since their location is accurate, in many cases they need not be recomputed, but can be retained for computation on finer scales. Before we develop our ideas we will describe the three more modern segmentation methods mentioned above.

Markov random field models are posed in a discrete setting. A simple model for

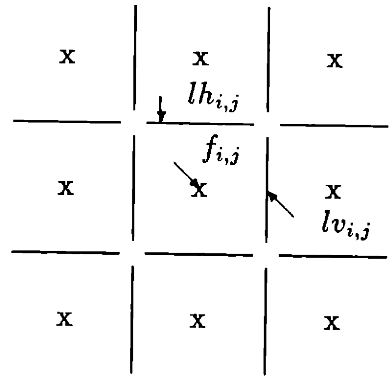


Figure 1.2: Markov Random Field Variable Assignments

segmentation was proposed by Geman and Geman [17]. The model resembles the Ising model of ferromagnetism. We associate to each element of a two dimensional lattice a real valued random variable $f_{i,j}$. This represents the intensity of the image. Another lattice is defined to represent the boundaries. To each pair of neighboring sites in the intensity lattice we associate an edge element; a random variable which takes values in $\{0, 1\}$. When the edge element takes the value 1 it is interpreted as the existence of an edge crossing between the two neighboring intensity sites with which the edge element is associated. When the edge element takes the value 0 the edge is absent. Figure 1.2 illustrates the construction and our notation. Since there are twice as many edges as intensity sites we label them $lv_{i,j}$ and $lh_{i,j}$ according to whether they represent vertical or horizontal edges.

For each *configuration* which can be thought of as an assignment to each variable of some particular value, one defines an energy,

$$\begin{aligned} E'(f,l) &= \sum_{(i,j)} (f_{i+1,j} - f_{i,j})^2(1 - lv_{i,j}) + (f_{i,j+1} - f_{i,j})^2(1 - lh_{i,j}) \\ &\quad + \alpha(lv_{i,j} + lh_{i,j}) \end{aligned}$$

and (formally) a probability according to a Gibbs distribution,

$$p(f,l) = \frac{e^{\frac{-E'(f,l)}{T}}}{Z'}$$

where Z' is the appropriate normalization constant. This then gives a *prior model* on the space of images. The parameter α is designed to control the quantity of boundary. T is a parameter which plays the same role as temperature in statistical mechanical systems. If we suppose that we observe an image g, i.e. a value $g_{i,j}$ for each site i,j in the lattice and that it is equal to $f_{i,j}$ plus some additive Gaussian noise, independent over the different sites, then we get a posterior conditional distribution on f and l which has the form,

$$p(f,l|g) = \frac{e^{\frac{-E(f,l|g)}{T}}}{Z}$$

where Z is a new normalization constant and,

$$\begin{aligned} E(f,l|g) &= \sum_{(i,j)} (f_{i+1,j} - f_{i,j})^2(1 - lv_{i,j}) + (f_{i,j+1} - f_{i,j+1})^2(1 - lh_{i,j}) \\ &\quad + \alpha(lv_{i,j} + lh_{i,j}) + \beta(f_{i,j} - g_{i,j})^2 \end{aligned}$$

where β is a constant depending on the variance of the observation noise. A maximum a posteriori (MAP) estimate is a configuration which maximizes $p(f,l|g)$. The MAP estimates are the minimizers of E. This formulation suggests the use of some Monte–Carlo for the optimization problem i.e. the finding of the MAP estimate. One simulates the posterior distribution via some Monte-Carlo method while slowly decreasing the temperature T to force the distribution to be weighted relatively more and more on the minimum energy states. This is essentially the simulated annealing algorithm [17] [22]. The difficulty with simulated annealing in general is that it takes excessively long to converge.

In a recent paper Geiger and Girosi [14] considered using mean field theory, a method often applied to statistical mechanical systems in which the interactions between different variables in the system are approximated by relations between their mean values. Using this approach the authors obtained deterministic equations for the variables in the system and they eliminated the line process variables expressing the solution solely in terms of the intensity variables. The equations they obtain have the interpretation of (locally) minimizing the following energy functional,

$$E(f|g) = \sum_{(i,j)} (f_{i+1,j} - f_{i,j})^2 + (f_{i,j+1} - f_{i,j+1})^2 + \beta(f_{i,j} - g_{i,j})^2$$

$$-T \ln [(1+e^{-\frac{\alpha-(f_{i+1,j}-f_{i,j})^2}{T}})(1+e^{-\frac{\alpha-(f_{i,j+1}-f_{i,j})^2}{T}})]$$

It is possible of course to find the Euler equations for the $f_{i,j}$. This has been done in [14]. There appears to be some connection between the result of this approach and the GNC algorithm of Blake and Zisserman [7] which we discuss a little later.

The Variational formulation deals exclusively with the energy functional such as E quoted above, and does not involve probability. The problem can now also be given a continuous domain formulation. The energy functional introduced by Mumford and Shah, [28] [29], and referred to as the weak membrane by Blake and Zisserman [7] is the following,

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

where α and β are positive real scalars, (the parameters of the problem,) f is a piecewise smooth approximation to g, having discontinuities only on the set Γ which one interprets as the boundaries 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 measures 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 β and α 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 β and α 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 Γ 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.

Other similar functionals have been proposed and are being considered. For example, Grimson [20], and Blake and Zisserman [7] proposed the addition of higher order smoothing terms. Their "weak plate" functionals are the following,

$$E(f,\Gamma) = \beta \int_{\Omega\setminus\Gamma} (f-g)^2 + \int_{\Omega\setminus\Gamma} f_{xx}^2 + f_{xy}^2 + f_{yy}^2 + \alpha_1 \text{length}(\Gamma_1) + \alpha_2 \text{length}(\Gamma_2)$$

$$E(f,\Gamma) = \beta \int_{\Omega\setminus\Gamma} (f-g)^2 + \int_{\Omega\setminus\Gamma} |\Delta f|^2 + \alpha_1 \text{length}(\Gamma_1) + \alpha_2 \text{length}(\Gamma_2)$$

Where $\Gamma = \Gamma_1 \cup \Gamma_2$, Γ_1 being the set of discontinuities of f and Γ_2 being the set of discontinuities of ∇f. We will not pursue these variants further in this thesis.

One of the methods which has been proposed for solving the minimization problem is the Graduated Non–Convexity (GNC) algorithm of Blake and Zisserman [7]. The method is developed in the discrete setting. We indicate the basic idea in the one dimensional case. To be consistent with the continuous domain variational formulation we explicitly use the lattice spacing which we denote δ to approximate the continuous formulation by finite elements. The line variables are eliminated to write the energy in the form,

$$E(f) = \sum_i G(f_{i+1} - f_i) + \beta\delta(f_i - g_i)^2$$

where G is defined by (see Figure 1.3),

$$G(x) = \begin{cases} \frac{x^2}{\delta} & \text{for } |x| \leq \sqrt{\alpha\delta} \\ \alpha & \text{for } |x| > \sqrt{\alpha\delta} \end{cases}$$

It is not difficult to see the correspondence between minima of this functional and minima of the original one. Blake and Zisserman's proposal was to approximate the problem of minimizing the non–convex functional E by finding a convex approximation to E. This is accomplished by modifying G. Basically the idea is to lower bound the second derivative of G in order to lower bound the eigenvalues of the Hessian of E. The solution they obtained is illustrated in Figure 1.4. The parameter c^* is ideally set as large as possible and controls the eigenvalues mentioned above. In each dimension there is a strict upper bound on the value of c^* for which the approximating functional remains convex. Thus effectively c^* is a constant. (We refer to [7] for

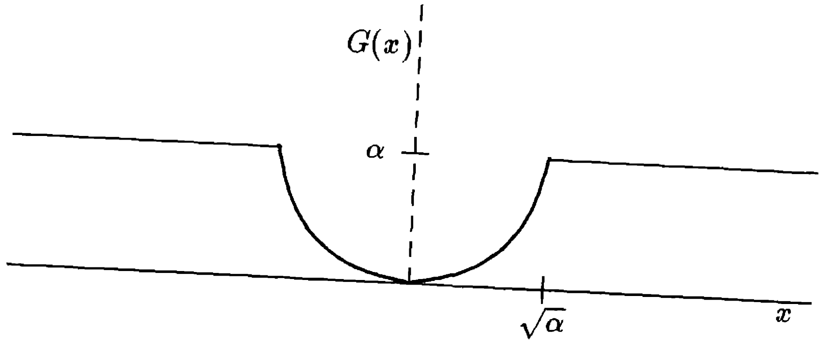
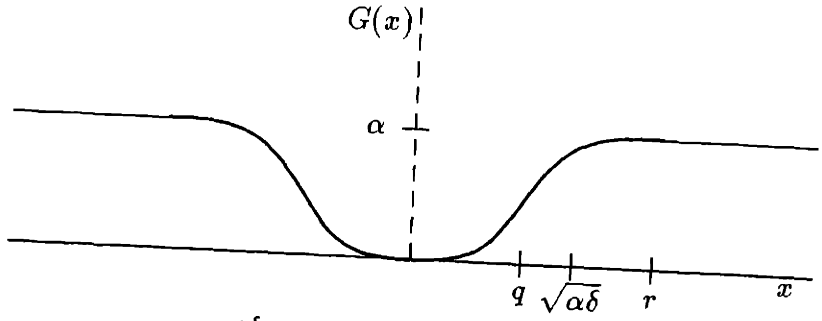


Figure 1.3: The Standard Interaction Function



$q = \frac{\alpha\delta}{r}$

$r^2 = \alpha\delta(1 + \frac{2}{\delta^2 \beta c^*})$

Figure 1.4: GNC Interaction Function

details.) Note that as $\delta \to 0$, we have $q \to 0$ and $r \to \infty$. Apparently this approach cannot be applied to the problem when it is posed in a continuous domain setting.

A final method we wish to mention is non-linear filtering. The non-linear filtering or anisotropic diffusion approach as described in [30] considers as an approximation to the data g the solution to a non-linear diffusion process whose initial condition is g. A *linear* approximation in this sense would be the solution to,

$$\frac{\partial f}{\partial t} = \Delta f, \;\; f_0 = g.$$

The solution f_t for fixed t turns out to be g convolved with a Gaussian kernel whose variance is linear in t. Thus scale-space representations can be viewed as the solution to this diffusion equation. The major deficiency of this approach as applied to segmentation, as we have pointed out earlier, is that it blurs the image uniformly resulting in poor localization of the edges. The basic idea of Perona and Malik [30] is to do the diffusion in such a way that edges do not get blurred; the conductance of the diffusion process should depend on the gradient of the image so that large gradients become enhanced while smaller ones are smoothed. Thus Perona and Malik were led to diffusions of the form,

$$\frac{\partial}{\partial t} = \nabla \cdot (c(x,y,t)\nabla f) \;\; f_0 = g. \tag{1.2.1}$$

where c is the conductance and can be written as a function of $|\nabla f|$,

$$c(x,y,t) = h(|\nabla f(x,y,t)|).$$

In [30] a couple of forms for h were suggested and experimented with,

$$\begin{aligned} h(x) &= \frac{C}{1+(\frac{x}{K})^2} \\ h(x) &= e^{-(\frac{x}{K})^2}. \end{aligned}$$

For either of these we see that the effect is to decrease the conductance c when the gradient ∇f is large.

Perona and Malik point out a connection between their approach and energy based methods such as the Markov random field approach and the variational formulation. The method can be viewed as a descent on an energy functional having only the smoothing and length terms. Consider a functional of the form,

$$E(f) = \sum_i G(f_{i+1} - f_i)$$

where G represents a function such as the one defined for the GNC algorithm, capturing both the smoothing and the length terms. A descent algorithm would set,

$$\frac{\partial}{\partial t} f_i = -c \sum_{j \in \mathcal{N}(i)} \dot{G}(f_i - f_j)$$

for some constant c. Since G is an even function $\dot{G}$ is odd and $\dot{G}(0) = 0$. Thus for some even positive function h we have,

$$\frac{\partial}{\partial t} f_i = -c \sum_{j \in \mathcal{N}(i)} h(f_i - f_j)(f_i - f_j)$$

which is essentially a non–linear diffusion equation.

1.3 Improved Techniques

The energy functional associated with the variational model and the Markov random field models is ad hoc. It seems necessary for producing models for vision to make ad hoc choices at some level unless one is specifically interested in reproducing human vision in which case one can appeal to empirical evidence. Among the attractions of a particular model will usually be it's analyzability, since it is this which makes the model's behavior predictable. Having an understandable model is almost as important as having intuitively appealing results especially if one intends to imbed the model in a much larger problem such as a recognition problem. In the case of the energy functional associated with the variational approach this predictability takes form partly in the analysis provided by the calculus of variations. 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 (whether it is based on intensity information or otherwise.) The restrictions on the geometry of the boundaries which arise out of the model 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 would be to make the model more complex, trying to produce more desirable behavior. Another would be to propose different ad hoc models. A third approach is the one taken in this thesis. 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. What we show in Chapter 4 is that particular meaningful members of such a closure may be found by taking the parameters associated with the functional to certain limits. In fact, one can produce essentially any piecewise smooth function such as one might suppose for a more general model of image. An idea which follows naturally on this one is to develop an algorithm in which the same limit is taken. Roughly speaking this is what is done in this thesis.

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" in but often one has notions of size, contrast and geometry in mind. Take for example the Gaussian smoothing technique. Here the relevant parameter is σ. For each value of σ one obtains a set of edges, the zero crossings of $\nabla^2 G * g$. Ideally one would hope that as σ 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 corresponding to large scale features. It is true that the edge sets vary continuously as a function of the parameter σ but in general it is computationally too costly to compute boundaries for sufficiently dense a set of σ to make the tracking obvious. Thus the correspondence problem presents considerable difficulties.

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 α and β while keeping their ratio fixed. While $\frac{1}{\sqrt{\beta}}$ can be related to σ which represented the width of the Gaussian kernel used for smoothing in that approach, keeping the ratio $\frac{\alpha}{\beta}$ fixed does not correspond with the Gaussian smoothing conception since this has the effect of keeping the total boundary and the localization errors roughly constant. In our limit theorems we (usually) keep α fixed and let β tend to ∞. Thus for a fixed α "scale" can be thought of as $\frac{1}{\sqrt{\beta}}$. However, in general there are two parameters and these parameters describe the range of functionals under consideration.

Blake and Zisserman [7] point out that in the case of the one dimensional variational problem the location of boundary points will tend to remain fixed as the parameters are varied (with $\frac{\beta}{\alpha}$ held constant.) Although this is not true in general they have proved the stability of isolated discontinuities. First of all as we pointed out they use the wrong notion of scale for comparison with the Gaussian smoothing case. Secondly, this property does not carry over to the two dimensional problem.

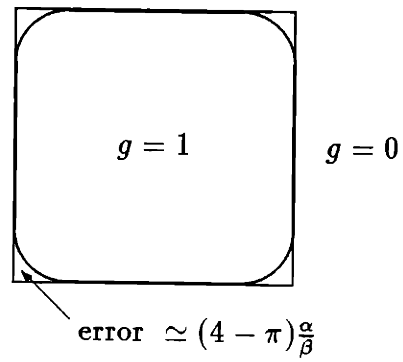


Figure 1.5: Segmentation of a Square

The example in Figure 1.5 shows how the variational solutions round corners. If the parameters are varied with the ratio of $\frac{\alpha}{\beta}$ held fixed then the solution may not change very much but it will change. It is likely in fact that for larger values of α and β the segmentation of the square depicted in Figure 1.5 will become fragmented as it becomes cost effective to smooth over the corners rather than put boundaries there. For small α and β solutions will tend to behave more like the piecewise constant case in which the boundary must form a closed contour. (These observations have been suggested by simulation results.) However, as we mentioned this does not correspond with scale–space in the Gaussian smoothing approach since with that method the rounding tends to zero as the scale is tends to zero. We have argued that with α fixed "scale" is best identified with $\frac{1}{\sqrt{\beta}}$. In this sense Figure 1.5 illustrates the scale dependence of the errors in the localization of the boundaries. The calculus of variations results summarized in Chapter 3 further indicate shortcomings of the formulation. The goal of the work presented in this thesis is to systematically develop a method for segmentation which has the advantages of the variational formulation but is able to accurately locate boundaries on all scales.

Our algorithm is developed in the following way. We first prove that asymptotically as parameters tend to certain limits (usually $\beta \to \infty$) that the solutions provided by the variational method will converge to the discontinuity set of an image. This implies, in particular, that one can recover t–junctions and corners at least asymptotically by the variational method. We also characterize the degree of corruption of the image which can be allowed before these results break down. The limit theorems are not enough to fulfill our goals because in effect they require the "scale" to tend to the microscopic, thus details on a smaller and smaller scale will be segmented in the limit. In general the errors one obtains vary directly with the "scale" of the segmentation. Thus the relative errors do not improve as one tends towards the microscopic scale. Another difficulty with the limit theorem is that as in traditional scale–space representations more boundaries appear on finer scales so the correspondence problem is again extant. The goal of the algorithm then is to take the limit, retaining coarse scale boundaries, letting them tend to limit positions while preventing the segmenting of smaller scale features. We accomplish this goal by smoothing out small scale features while retaining the detail needed to accurately place the boundaries corresponding to the large scale features. The mechanism for doing this is discussed in detail in Chapter 5. Essentially we feedback information from a smoothed version of the image to the image in a manner depending on our current estimate of the boundary locations.

In principal our method for improving on the boundary locations can be applied to any of the segmentation methods mentioned. We prefer to concentrate on the variational formulation because mathematically it appears most fundamental and because the main ideas of this thesis, including the limit theorems, the approximating functionals and the algorithm can be made mutually coherent in this framework.

Another significant deviation we make from previously mentioned methods is our representation of the boundary. We propose to represent the boundaries by a function defined on the same domain as the image. We use the approximation scheme due to Ambrosio and Tortorelli [5] which is presented in Chapter 3. This approach has certain general advantages which we will discuss in Chapter 5. In our particular case it also facilitates the implementation of the smoothing mechanism mentioned earlier. In this approximation one replaces the set $\Gamma \subset \Omega$ with a function $v : \Omega \to [0,1]$. There are various representations of the approximating functional. The one we choose is the following.

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

The location of boundaries is in general given by $(1 - v^2)^n \simeq 0$. It was proved in [5] that the sequence of functional $\{E^n\}$ Γ*–converges* to E. We define this notion of convergence in Chapter 2. It's most significant property is that minimizers of E^n converge to minimizers of E (in an appropriate metric) as n tends to infinity. Also the minimal value of the E^n converges to the minimal value of E. Our computations employ a gradient descent to find local minimizers of E^n. This approach closely resembles the non–linear filtering approach. In our case however we have an explicit representation of the boundary.

1.4 Outline of the Thesis

Since the mathematical results logically precede the algorithm the next three chapters are devoted to them. In Chapter 2 we provide a brief introduction to some of the mathematical concepts which are used in later analysis and which by our estimate, are not widely known. Some of our work uses ideas from Geometric measure theory for which a monumental reference is Federer [12]. In particular we define various measures which we later use as generalizations of "length". We define what is meant by the *essential boundary* of a set. This object has many properties which make it useful from the point of view of the calculus of variations. It is closely related to the space of functions of bounded variation which have been extensively used in the study of minimal surface theory [18]. The most recent results on the fundamental questions associated the variational approach to image segmentation have been found within a subclass of the functions of bounded variation which have been called the *special functions of bounded variation*, SBV. The third section of Chapter 2 provides the essential properties of this space which are needed for results proved in later chapters. Finally in the last section of that chapter we outline the concept of Γ*–convergence*. This idea, due to E. De Giorgi and independently, to H. Attouch provides a rigorous way to formulate the approximation of one variational problem by another, with perhaps completely different mathematical variables. For the purposes of this thesis we intend to use the approximation mentioned in the preceding section in which the boundaries, which typically are represented by a one dimensional subset of the domain of definition of the image, are replaced by a function defined on the entire domain.

In Chapter 3 we have collected together some of the basic results which are relevant to the ideas developed in the thesis. The chapter is divided into three sections. In the first we review what is known about the fundamental question of existence of minimizers to the (continuous) variational problem. We state the known results and include a proof of our own which requires the boundaries to be constrained to have finitely many components. The second section provides a review of the results obtained by Mumford and Shah [29] via the calculus of variations for the minimizers of E which constrain the geometry of the boundaries. These results help to motivate our work which aims to circumvent those constraints. We view them as undesirable structural restrictions placed on solutions which result directly from the ad hoc formulation of the energy functional. In the final section we detail the Γ–convergent approximation mentioned above.

Chapter 4 contains 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 our algorithm. We suppose that our image is a corrupted version of an underlying piecewise constant or piecewise smooth function, depending on the particular problem formulation. We then show 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. Those results say that locally the minimizers have a certain 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. As we mentioned earlier these results also play the role of governing our algorithm for scale–independent segmentation.

In Chapter 5 we develop of the algorithm. In fact we propose several algorithms all of which are based on essentially the same idea. The main ideas are introduced in a paradigm which sketches the algorithm and states the essential insights the algorithms implement. These insights 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. With the paradigm stated we detail our particular implementation of it. This includes the choosing of the various parameters of the algorithm and a stability analysis. Following this we develop a computational model based on the Γ–convergent approximation to the variational formulation of the segmentation problem. Next, we consider potential discrete versions of the problem and develop a particular one based on our computational model. This chapter also contains our simulation results.

The closing chapter summarizes the main contributions of the thesis and points out some directions for further research.

Chapter 2

Mathematical Preliminaries

This chapter provides an introduction to and fixes our notation for what we consider to be some of the less familiar mathematical concepts which have been applied by the author and others to the variational formulation of the segmentation problem. In the first section we define the the Hausdorff and Minkowski measures. These will provide us with generalizations, to irregular sets, of the length term in the segmentation functional. We also introduce the Hausdorff metric which will provide us with a topology for boundaries or edges in images and which we will use to measure accuracy of boundary localization. The second section serves to define the notion of essential boundary. This construct, due to Federer [12], is for analytical purposes a very useful and tractable notion of boundary. It is closely related to the space of functions of bounded variation, BV. In the third section we define and present some basic properties of the space BV and a subspace SBV, the special functions of bounded variation. The space SBV(Ω) plays an important role in the study of the fundamental mathematical questions associated with the variational formulation. It is in the SBV setting that the most general existence results have been proved for the variational problem. Also, our asymptotic theorems for minimizers of E (see chapter 4) is proved in the SBV setting. Finally in the fourth section we present a notion of variational convergence called epi–convergence or Γ–convergence. This concept is used in the development of approximations to and computational models for the variational problem.

2.1 Metrics and Measures

In this section we introduce a variety of ideas useful in dealing with the 'boundaries' or 'edges' of an image. The 'image' will usually be a real valued function defined on a bounded open set $\Omega \subset \Re^2$, often Ω will be a rectangle. A *boundary* generally refers to a closed subset of Ω. However, sometimes the boundary may be restricted to have certain additional properties such as having a finite number of connected components. The following concepts can be applied to such objects.

The Hausdorff Metric

For $A \subset \Re^n$, the *ϵ-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 [34], $[A]_\epsilon$ is the dilation of A with the open ball of radius ϵ. A notion of distance between boundaries which we will often use is known as the Hausdorff metric. Denoted $d_H(\cdot,\cdot)$, it 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 $\Re^n$. An important property of this metric is that it induces a topology which makes the space of boundaries compact.

Theorem 2.1 For any infinite collection, $\mathcal{C}$, of non-empty closed subsets of any bounded closed set $\overline{\Omega}$ there exists a sequence $\{\Gamma_n\}$ of distinct sets from $\mathcal{C}$ and a non-empty closed set $\Gamma \subset \overline{\Omega}$ such that $\Gamma_n \to \Gamma$ in the Hausdorff metric.

Proof See [11], Theorem 3.16. □

Hausdorff Measure

The variational problems treated in this thesis generally penalize the total length of the boundary which ideally would be the union of a set of somewhat smooth curves. A *curve* $\Gamma \subset \Re^n$ is the image of a continuous injection $g : [0,1] \rightarrow \Re^n$. The *length* of a curve Γ 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 Γ is said to be *rectifiable* if $L(\Gamma) < \infty$.

Unfortunately some fundamental mathematical questions such as whether there exist minimizers of E have not been resolved with the boundary defined as a set of curves. Instead, for such results, potential boundaries are drawn from a wider class of sets and more general measures are applied to them. A measure on the space of boundaries which generalizes the usual notion of length is required. A variety of such measures for subsets of $\Re^n$ have been investigated (see [12] for many examples). Perhaps the most widely used and studied are the Hausdorff measures [11, 12, 31].

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

$$\omega_s = \frac{\Gamma(\frac{1}{2})^s}{\Gamma(\frac{s}{2} + 1)}$$

where $\Gamma(\cdot)$ is the usual Gamma function. For integer values of s, ω_s is the volume of the unit ball in $\Re^s$. For $s > 0$ and $\delta > 0$ define

$$\mathcal{H}^s_\delta(A) = 2^{-s}\omega_s \ \inf\{\sum_{i=1}^{\infty} \text{diam}(U_i)^s \ : A \subset \bigcup_{i=1}^{\infty} U_i, \ \text{diam}(U_i) \leq \delta\},$$

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

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

Note that the factor $2^{-s}\omega_s$ in the definition of $\mathcal{H}^s(\cdot)$ is included for proper normalization. For integer values of s Hausdorff measure gives the desired value on sets where the usual notions of length, area, and volume apply.

Many properties of Hausdorff measure can be found in [11, 12, 31]. The following theorem states that $\mathcal{H}^1$ is a generalization of length, as required.

Theorem 2.2 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. The one following it states an important lower-semicontinuity property. We refer to a compact connected set as *continuum.*

Theorem 2.3 If Γ is a continuum with $\mathcal{H}^1(\Gamma) < \infty$, then Γ consists of a countable union of rectifiable curves together with a set of $\mathcal{H}^1$-measure zero.

Proof See [11], Theorem 3.14. □

Theorem 2.4 If $\{\Gamma_n\}$ is a sequence of continuua in $\Re^n$ that converges (in Hausdorff metric) to a compact set Γ, then Γ is a continuum and $\mathcal{H}^1(\Gamma) \leq \liminf_{n\to\infty} \mathcal{H}^1(\Gamma_n)$.

Proof See [11], Theorem 3.18. □

In section 3.1.2 we extend this lower-semicontinuity result to sets with a finite number of connected components.

Minkowski Content

Another measure which can effectively replace the length is Minkowski content [12]. Let $|\cdot|$ denote Lebesgue measure in $\Re^n$. For any $A \subset \Re^n$, $0 \leq s \leq n$, and $\epsilon > 0$, define,

$$\mathcal{M}^s_\epsilon(A) = \frac{|[A]_\epsilon|}{\epsilon^{n-s}\omega_{n-s}}$$

Minkowski content is thus a neighborhood based definition. The algorithm developed in this thesis uses neighborhoods of sets of boundaries. Minkowski content will help use to characterize these neighborhoods. As in the definition of Hausdorff measure, the term ω_{n-s} is included for proper normalization. In general, $\lim_{\epsilon\to 0} \mathcal{M}^s_\epsilon(A)$ may not exist (for an example see [12], section 3.2.40). However, *lower* and *upper Minkowski contents* can be defined by

$$\mathcal{M}^s_*(A) = \liminf_{\epsilon \to 0} \mathcal{M}^s_\epsilon(A)$$

and

$$\mathcal{M}^{*s}(A) = \limsup_{\epsilon \to 0} \mathcal{M}^s_\epsilon(A)$$

respectively. If these two values agree then one refers to the common value as the *s-dimensional Minkowski content* of A and it is denoted simply as $\mathcal{M}^s(A)$.

The following theorem relates Minkowski content to Hausdorff measure. A subset Γ of $\Re^n$ is called *m-recifiable* if there exists a Lipshitzian function mapping a bounded subset of $\Re^m$ onto Γ.

Theorem 2.5 [12, Theorem 3.2.39] If Γ is a closed m-rectifiable subset of $\Re^n$ then $\mathcal{M}^m(A) = \mathcal{H}^m(\Gamma)$.

Another important estimate on Minkowski content for our purposes is the following,

Proposition 2.6 If $\Gamma \subset \Re^2$ is a recifiable curve then $|[\Gamma]_\epsilon| \leq 2\epsilon\mathcal{H}^1\Gamma + \pi\epsilon^2$ and so $\mathcal{M}(\Gamma) \leq \mathcal{H}^1(\Gamma) + \frac{1}{2}\pi\epsilon$.

Proof Since Γ is rectifiable, $\Gamma = \{\gamma(t) : 0 \leq t \leq 1\}$ where $\gamma : [0,1] \to \Re^2$ is rectifiable and $\mathcal{H}^1(\Gamma) = \sup\{\sum_{i=1}^m \|\gamma(t_i) - \gamma(t_{i-1})\| : 0 = t_0 < t_1 < \cdots < t_m = 1\}$. For $E = 1, 2, \ldots$ let $\{t_{ij}\}$ be a sequence of dissections such that $\max_i\{\|t_{ij} - t_{i-1,j}\|\} \to 0$ and $\mathcal{H}^1(\Gamma) = \lim_{j\to\infty} \sum_{i=1}^{m(j)} \|\gamma(t_{ij}) - \gamma(t_{i-1,j})\|$. Let $C_j = \cup_{i=1}^{m(j)} S_i$ where S_i is the straight line joining $\gamma(t_{i-1,j})$ and $\gamma(t_{ij})$. Then $|[S_{ij}]_\epsilon| = 2\epsilon\|\gamma(t_{ij}) - \gamma(t_{i-1,j})\| + \pi\epsilon^2$, and

$$\begin{aligned}
|\cup_{i=1}^k [S_{ij}]_\epsilon| &= |\cup_{i=1}^{k-1} [S_{ij}]_\epsilon| + |[S_{kj}]_\epsilon| - |[S_{kj}]_\epsilon \cap \bigcup_{i=1}^{k-1} [S_{ij}]_\epsilon| \\
&\leq |\cup_{i=1}^{k-1} [S_{ij}]_\epsilon| + |[S_{kj}]_\epsilon| - \pi\epsilon^2 \\
&= |\cup_{i=1}^{k-1} [S_{ij}]_\epsilon| + 2\epsilon\|\gamma(t_{kj}) - \gamma(t_{k-1,j})\|
\end{aligned}$$

By induction on i, we get

$$|[C_j]_\epsilon| \leq \sum_{i=1}^{m(j)} 2\epsilon\|\gamma(t_{ij}) - \gamma(t_{i-1,j})\| + \pi\epsilon^2$$

In [23, Theorem 6] it was shown that if $\Gamma_n \to \Gamma$ in Hausdorff metric then $|[\Gamma_n]_\epsilon| \to |[\Gamma]_\epsilon|$. Since $C_j \to \Gamma$ in Hausdorff metric we obtain,

$$|[\Gamma]_\epsilon = \lim_{j\to\infty} |[C_j]_\epsilon| \leq 2\epsilon\mathcal{H}^1(\Gamma) + \pi\epsilon^2.$$

□

2.2 Essential Boundaries

The problem of defining the perimeter of a set proved difficult from the point of view of the calculus of variations. The topological boundary does not in general possess sufficient mathematical properties to fulfill the usual requirements of the calculus. Federer [12] introduced a notion based on the idea of density. The *essential boundary* of a set is those points where the set has density other than zero or one. To be more precise, for a borel set $A \subset \Omega$ we set,

$$A_t = \{x \in \Omega : \lim_{\rho\to 0+} \frac{|B \cap B_\rho(x)|}{\omega_n \rho^n} = t\} \quad (t \in [0,1]).$$

where $w_n = |B_1|$. A_t is the set where A has density t. Federer defined the essential boundary $\partial^* A$ as

$$\partial^* A = \Omega \backslash (A_0 \cup A_1)$$

The essential boundary possesses the following property,

$$\partial^* A \supset A_{\frac{1}{2}} \text{ and } \mathcal{H}^{n-1}(\partial^* A \backslash A_{\frac{1}{2}}) = 0 \tag{2.2.1}$$

Also, the set $\partial^* A$ is countably rectifiable in the sense of Federer ([12], chapter 3),

$$\partial^* A \subset \bigcup_{n=1}^{\infty} \Gamma_n \cup N$$

where the Γ_n are C^1 hypersurfaces and $\mathcal{H}^{n-1}(N) = 0$.

A result which characterizes the essential boundary very nicely is the following. For bounded measurable sets A, if $\mathcal{H}^{n-1}(\partial^* A) < \infty$ then,

$$\begin{aligned}\mathcal{H}^{n-1}(\Omega \cap \partial^* A) \;&=\; \inf\,\{\, \liminf_{n\to\infty} \mathcal{H}^{n-1}(\Omega \cap \partial A_n) : \\ &\qquad A_n \to A \text{ locally in measure, } A_n \text{ polyhedral.}\}\end{aligned} \tag{2.2.2}$$

A measurable set $A \subset \Re^n$ satisfying $\mathcal{H}^{n-1}(K \cup \partial^* A) < \infty$ for all compact $K \subset \Re^n$ is referred to as a *Cacciopoli* set.

This concept of boundary will be helpful in the formulation of the piecewise constant version of limit theorem proved in chapter 4. Also it was used by Mumford and Shah in their proof of the existence of minimizers for E_0.

2.3 SBV Functions

Let J be an open interval in $\Re$ then $u : J \to \Re$ is a function of *bounded variation* in J if

$$V_J(u) = \sup\{\sum_{i=1}^{k-1} |u(t_{i+1}) - u(t_i)| : \inf J < t_1 < \ldots < t_k < \sup J\} < +\infty. \tag{2.3.1}$$

$V_J(u)$ is called the total variation of u in J. The space BV(J) is the space of Borel functions $u : J \to \Re$ such that

$$\text{ess-}V_J(u) = \inf\{V_J(v) : v = u \text{ almost everywhere }\} < +\infty.$$

In higher dimensions this definition can be generalized by slicing arguments [4]. Slicing arguments involve considering all lines which are perpendicular to an $n-1$ dimensional hyperplane. Membership in BV(Ω) can be characterized by the trace of a function on such one dimensional slices. The space BV(Ω) can be characterized in other ways. In particular the functions in BV(Ω) are those functions $u \in L^1(\Omega)$ such that Du, the distributional derivative of u is representable as a bounded Radon measure on Ω with values in $\Re^2$ [12] [18].

For each $x \in \Omega$ we can define the *approximate upper (and lower) limit* of u at x. The upper limit is the greatest lower bound of all $t \in [-\infty, \infty]$ such that $\{x \in \Omega :$

$u(x) > t\}$ has 0 density at x, i.e.,

$$u^+(x) = \inf\{t \in [-\infty, \infty] : \lim_{\rho \to 0+} \frac{|\{u > t\} \cap B_\rho(x)|}{\rho^n} = 0\}. \tag{2.3.2}$$

Similarly the approximate lower limit is,

$$u^-(x) = \sup\{t \in [-\infty, \infty] : \lim_{\rho \to 0+} \frac{|\{u < t\} \cap B_\rho(x)|}{\rho^n} = 0\}. \tag{2.3.3}$$

Points where $u^+ = u^-$ are points of approximate continuity for u. The remainder, which we denote $\mathcal{S}_u$, where $u^- < u^+$ is the jump set of u. If $u \in L^\infty(\Omega)$ then the points of approximate continuity are precisely the Lebesgue points of u, i.e.,

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

By the Lebesgue derivation theorem we conclude $|\mathcal{S}_u| = 0$.

It turns out that for $\mathcal{H}^{n-1}$-almost all $x \in \mathcal{S}_u$ one can define an approximate tangent to $\mathcal{S}_u$ while u^+ and u^- provide one sided limits. Given ν a unit vector in $\Re^n$, $z \in \Re$ we say that $z = u^+(x, \nu)$ if

$$\lim_{\rho \to 0+} \frac{|\{y \in B_\rho(x) : \langle y - x, \nu \rangle > 0, |u(y) - z| > \epsilon\}|}{\rho^n} = 0$$

for every $\epsilon > 0$. Similary one defines $u^-(x, \nu) = u^+(x, -\nu)$. For $\mathcal{H}^{n-1}$-almost all $x \in \mathcal{S}_u$ there is a unique ν such that $u^-(x) = u^-(x, \nu)$ and $u^+(x) = u^+(x, \nu)$. ν represents a normal to $\mathcal{S}_u$.

For any function $u \in \mathrm{BV}(\Omega)$ the measure Du can be decomposed as

$$Du = \nabla u \, dx + Ju + Cu.$$

$\nabla u \, dx$ is the part of Du which is absolutely continuous with respect to Lebesgue measure (which we denote by dx) and $\nabla u \in L^1(\Omega, \Re^2)$ is thus the corresponding Radon_Nikodym derivative. $Ju + Cu$ is singular with respect to Lebesgue measure. Ju is defined on any Borel set $B \in \Omega$ by,

$$Ju(B) = \int_{B \cap \mathcal{S}_u} (u^+ - u^-) \nu_u \, d\mathcal{H}^{n-1}$$

where $\nu_u(x)$ is the approximate normal to $\mathcal{S}_u$ at $x \in \mathcal{S}_u$. $Cu(B)$ is a bounded Radon measure on Ω with values in $\Re^2$. It is a fact that, $\mathcal{H}^{n-1}(B) < +\infty \Rightarrow Cu(B) = 0$ [4].

It is clear that Ju captures the jump of discontinuity set of u and $\nabla u\ dx$ the smooth part. Thus a reasonable formulation of the variational problem in this setting is find minimizers of,

$$E(u) = \int_\Omega (u-g)^2 + \int_\Omega |\nabla u|^2 + \mathcal{H}^{n-1}(\mathcal{S}_u) \tag{2.3.4}$$

for $u \in \mathrm{BV}(\Omega)$ (where Ω is n dimensional.) The difficulty which arises is that the functional E gives no control over Cu. In fact the Cantor–Vitelli function in one dimension satisfies $Du = Cu$. A consequence of this is that E is not coercive in $\mathrm{BV}(\Omega)$, that is E bounded sets are not compact.

We say $u \in \mathrm{SBV}(\Omega)$ if $u \in \mathrm{BV}(\Omega)$ and $Cu = 0$. SBV possesses some very useful properties. For example, as with $\mathrm{BV}(\Omega)$, membership in SBV can be determined by examining one–dimensional sections, and SBV is closed under L^1 limits. Furthermore the functional E is lower-semicontinuous in SBV with the L^1 topology. This issue will be discussed further in the chapter on existence results. As we have mentioned, some of our work is developed in the SBV setting and for this development we require some basic results. We collect a few of them here. We focus on the case $\Omega \subset\subset \Re^2$.

Let D be a relatively compact open subset of Ω with Lipshitz boundary such that $\mathcal{S}_u \cap \partial D$ has only a finite number of points. From the trace theorems for BV functions (see [18] Theorem 2.10) it follows that for $u \in \mathrm{BV}(\Omega)$,

$$\int_D \phi Du = -\int_D u \mathrm{div}\phi\, dx + \int_{\partial D} u\phi\nu \ \ d\mathcal{H}^1$$

for every bounded Borel vector field $\phi \in C^1(\overline{D}, \Re^2)$, where ν is the outward normal to D. Thus it is also true that if $u \in \mathrm{SBV}(\Omega)$ then,

$$-\int_D u\mathrm{div}\phi\, dx + \int_{\partial D} u\phi\nu\, d\mathcal{H}^1 = \int_D \phi\nabla u dx + \int_{D\cap\mathcal{S}_u} (u^+ - u^-)\phi\nu_u\, d\mathcal{H}^1. \tag{2.3.5}$$

Finally we present some rather deep results due to De Giorgi–Carriero–Leaci [10] which characterizes $\mathcal{S}_u$ in terms of the decay of functionals around points in Ω.

Theorem 2.7 [10, Theorem 3.6] Let $x \in \Omega$ and $u \in \mathrm{SBV}(\Omega)$. If,

$$\lim_{\rho\to 0} \rho^{-1}\left[\int_{B_\rho(x)} |\nabla u|^2 dy + \mathcal{H}^1(\mathcal{S}_u \cap B_\rho(x))\right] = 0$$

then $x \notin \mathcal{S}_u$.

The proof of this theorem is based on a generalization to SBV of the Poincaré–Wirtinger inequality. To state the next theorem we need to introduce some more notation. Let $u \in \mathrm{SBV}(\Omega)$. For every compact set $K \subset \Omega$ we set,

$$F(u,K) = \int_K |\nabla u|^2 + \mathcal{H}^1(\mathcal{S}_u \cap K)$$

and

$$\Phi(u,K) = \inf\{F(v,K) : v \in \mathrm{SBV}(\Omega), v = u \in \Omega \backslash K\}.$$

Obviously $\Phi \leq F$ and we define the *deviation from minimality* as

$$\Psi(u,K) = F(u,K) - \Phi(u,K).$$

We are now ready to state the advertised theorem.

Theorem 2.8 [10, Theorem 4.13] There exist universal constants $\xi, \gamma > 0$ such that if $u \in \mathrm{SBV}(\Omega)$, $B_\rho(x) \subset\subset \Omega$ for some $\rho > 0$, and each of the following three conditions hold;

$$\begin{aligned} F(u, \overline{B}_\rho(x)) &\leq \xi\rho, \\ \lim_{t\to 0^+} t^{-1}\Psi(u, \overline{B}_t(x)) &= 0, \\ \Psi(u, \overline{B}_t(x)) &\leq \gamma t, \ \forall t \leq \rho, \end{aligned}$$

then $\lim_{\rho\to 0} \rho^{-1} F(u, \overline{B}_\rho(x)) = 0$.

Proof See [10] or [4]. □

The proof of this theorem is based on the theorem quoted below. To state this theorem it is convenient to reintroduce α into the notation. Thus temporarily we set,

$$F(u,\alpha,K) = \int_K |\nabla u|^2 + \alpha\mathcal{H}^1(\mathcal{S}_u \cap K)$$

and

$$\Phi(u,\alpha,K) = \inf\{F(v,\alpha,K) : v \in \mathrm{SBV}(\Omega), v = u \in \Omega \backslash K\}.$$

Theorem 2.9 For any $\delta \in (0,1)$ there exist two universal constants ξ and θ such that if $\rho > 0, \overline{B}_\rho(x) \in \Omega, u \in \mathrm{SBV}(\Omega)$ with,

$$\begin{aligned} F(u,\alpha,\overline{B}_\rho(x)) &\leq \xi\rho, \\ \Psi(u,\alpha,\overline{B}_\rho(x)) &\leq \theta F(u,\alpha,\overline{B}_\rho(x)) \end{aligned}$$

then,

$$F(u,\alpha,\overline{B}_{\frac{\rho}{2}}(x)) \leq (\frac{1}{2})^{2-\delta} F(u,\alpha,\overline{B}_\rho(x))$$

The theorem is proved by contradiction. Assuming the theorem is false it is possible to find $\delta \in (0,1)$, an $\alpha > 0$, and sequences $\xi_n, \theta_n, \rho_n, x_n, u_n$ such that $\xi_n \downarrow 0$, $\theta \downarrow 0$, $\overline{B}_{\rho_n}(x_n) \subset \Omega$,

$$\begin{aligned} F(u_n,\alpha,\overline{B}_{\rho_n}(x_n)) &\leq \xi_n\rho, \\ \Psi(u_n,\alpha,\overline{B}_{\rho_n}(x_n)) &\leq \theta_n F(u,\alpha,\overline{B}_{\rho_n}(x_n)) \end{aligned}$$

and,

$$F(u_n,\alpha,\overline{B}_{\frac{\rho_n}{2}}(x_n)) > (\frac{1}{2})^{2-\delta} F(u_n,\alpha,\overline{B}_{\rho_n}(x_n)).$$

By rescaling and translating one obtains a sequence v_n of functions in $\mathrm{SBV}(B_1)$ such that $F(v_n,\frac{\alpha}{\xi_n},\overline{B}_1) = 1$, $\Psi(v_n,\frac{\alpha}{\xi_n},\overline{B}_1) \leq \theta_n$ and

$$F(v_n,\frac{\alpha}{\xi_n},\overline{B}_{\frac{1}{2}}) \geq (\frac{1}{2})^{2-\delta}$$

Since $\frac{\alpha}{\xi_n} \uparrow \infty$ and the deviation from minimality in B_1 tends to 0, in the limit the functions v_n should behave like harmonic functions. But if v is an harmonic function then,

$$\int_{B_{\frac{1}{2}}} |\nabla v|^2 dx \leq (\frac{1}{2})^2 \int_{B_1} |\nabla v|^2 dx$$

and it is by this that a contradiction is found. Although we have sketched the proof in the two dimensional setting the theorem has been proved in $\Re^n$.

2.4 Γ–Convergence

Another concept which plays an important role in our study of the segmentation problem is the notion of Γ-*convergence* due to De Giorgi. The same concept was developed independently in France under the name *epi-convergence* by H. Attouch [6].

This concept concerns variational convergence, i.e. the approximation of one variational problem by another. In this section we provide a definition of Γ-convergence and state some of its basic properties.

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 $x \in S$,

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

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

Proposition 2.10 (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 $t \in [0, \infty)$ 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.

Ambrosio and Tortorelli [5] found a sequence of functionals which Γ–converge (in an appropriate setting) to the variational segmentation functional. In this approximation the boundaries are replaced by a function. This formulation then admits a finite element discrete approximation. Our proposed segmentation algorithm uses these approximate formulations in an essential way. Greater detail concerning them is provided in subsequent chapters. Other applications of Γ–convergence are discussed in [4].

Chapter 3

Some Analytical Results

A considerable amount of work has been done to analyze the variational problem. The fact that the variational problem is amenable to analysis, particularly in the continuous formulation was one of the reasons for considering it as a means to image segmentation. In this chapter we review the present state of knowledge on some of the basic analytical problems associated with the variational formulation. In the first section we consider the most fundamental of these problems, namely the existence of solutions. First we provide a statement of an existence result in a class of regular Γ which has been proven for the functional E_0. Following this we introduce a weak version of the variational problem in which Γ is required to be closed and $f \in W^{1,2}(\Omega)$. A proof is given for a slightly modified version of the problem in which the number of connected components of the boundary is required to be finite. The proof is given in a form general enough to include the "weak plate" (see the introduction) formulations of the segmentation problem. Finally we present a different and technically much harder formulation of the problem. Cast in the SBV setting, this formulation has been studied by Ambrosio [2, 3, 4], and De Giorgi–Carriero–Leaci [10]. With a remarkable result on the SBV minimizers to the variational formulation of the segmentation problem, De Giorgi and coworkers [10] were able to show the existence of minimizers of E for the weak formulation referred to above. The second section reviews the results of the calculus of variations which constrain the form of minimizers of E. These results were given in [29]. Finally in the last section we present a parametrized sequence of approximations which converge to the variational formulation in the sense of Γ–convergence. This result is due to L. Ambrosio and V. Tortorelli [5].

3.1 Existence Results

3.1.1 Piecewise Constant Case

As we have mentioned 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. However for the functional,

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

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

Theorem 3.1 [27] Let Ω 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 Γ 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 Γ 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 in [27] using direct, constructive methods. Finally, another proof using Γ restricted to be unions of line segments and then taking limits as the segment lengths tend to zero was achieved by Y. Wang [36].

Existence results for minimizers of E have now been found for various weak versions of the problem. The natural formulation of a weak version of the problem is to define,

$$E(f, \Gamma) = \beta \int_\Omega (g - f)^2 + \int_{\Omega \backslash \Gamma} \|\nabla f\|^2 + \mathcal{H}^1(\Gamma) \qquad (3.1.1)$$

with Γ being a relatively closed subset of Ω 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 but the first result obtained is presented in the next section. It requires the addition of another term to E which forces the number of connected components to be finite, so it is a slight modification of this formulation.

3.1.2 Finite Component Existence Theorem

In this section we prove the existence result mentioned above. The cost functional we will be considering has as a special case one of the following form,

$$E(f,\Gamma) = \int_\Omega (g-f)^2 + \int_{\Omega\setminus\Gamma} \|\nabla f\|^2 + \nu(\Gamma) \tag{3.1.2}$$

g, as usual, we interpret as a gray-level image which has been observed and lies in $L_\infty(\Omega)$. The domain of our problem, Ω, is a nonempty bounded open set in $\Re^2$. For convenience we require Ω to be convex (see proof of theorem 3.3) but this constraint is not essential. "Boundaries", which we will denote by Γ are nonempty closed sets in $\Re^2$ such that $\Gamma \subset \overline{\Omega}$. For 3.1.2 the function f belongs to the Sobolev space $W^{1,2}(\Omega\setminus\Gamma)$, this will be generalized later. The functional "ν" assigns some cost to Γ. We will define ν to be one dimensional Hausdorff measure plus some term to control the number of connected components of Γ. We provide an appropriate topology for the solution space and obtain compactness and lower-semicontinuity results yielding the existence of minimizing solutions for E.

Boundaries

In this section we formally define and show relevant properties of the cost associated with boundaries. We require only that the boundary Γ be a closed subset of $\overline{\Omega}$ and that $\mathcal{H}^1(\Gamma) < \infty$. We recall that $d_H(\cdot,\cdot)$ is a (complete) metric on the space of boundaries as defined here. When we speak of convergence of boundaries we mean with respect to the topology induced by this metric.

The cost associated with the boundaries has the following form.

$$\nu(\Gamma) = \mathcal{H}^1(\Gamma) + F(\#(\Gamma)) \tag{3.1.3}$$

F is any nondecreasing function defined on the nonnegative integers which satisfies $\lim_{n\to\infty} F(n) = \infty$. One acceptable version of F would be a function which is zero for $n \leq K$ and ∞ otherwise; this would simply bound the number of connected components of Γ.

The following results establish the lower-semicontinuity properties required to demonstrate the existence of minimizers of E. For convenience we introduce the notation,

$$r(A,B) = \sup_{x \in A} \inf_{y \in B} \|x-y\|$$

and note $r(A,B) \leq d_H(A,B)$.

Lemma 3.2 The functional $\#(\cdot)$ is a lower-semicontinuous function on the space of boundaries.

Proof: Suppose $\Gamma_n \rightarrow \Gamma$ and $\#(\Gamma) = c < \infty$. Then there exists an open cover of Γ consisting of c disjoint open sets $G_1, G_2, \ldots, G_c$ such that $\Gamma \cap G_i \neq \emptyset$, $\forall i$. Γ is closed so $\exists \delta > 0$ such that $\forall i$, $\mathrm{r}(\Gamma \cap G_i, \Re^2 \backslash G_i) > \delta$. Since $\mathrm{r}(\Gamma, \Gamma_n) \rightarrow 0$, for n sufficiently large $\Gamma_n \subset \cup_i G_i$ and for each i, $\Gamma_n \cap G_i \neq \emptyset$. Thus $\liminf_{n \rightarrow \infty} \#(\Gamma_n) \geq c$. If $\#(\Gamma) = \infty$ then we can repeat this argument for any c and the result follows.

Theorem 3.3 The functional ν is lower-semicontinuous on the space of boundaries i.e.

$$\nu(\Gamma) \leq \liminf_{n \rightarrow \infty} \nu(\Gamma_n) \tag{3.1.4}$$

whenever $\Gamma_n \rightarrow \Gamma$.

Proof: Assume (without loss of generality) that for all n $\nu(\Gamma_n) \leq K$. It follows that $\#(\Gamma_n)$ is uniformly bounded and by Lemma 3.2, $\#(\Gamma) \leq M < \infty$. Since the connected components of Γ are thus separated pairwise by some finite distance the result follows once we show it for connected Γ.

Assume Γ is connected. Let $\delta_n = \mathrm{d}_H(\Gamma_n, \Gamma)$. Suppose Γ_n has more than one connected component and let C be one connected component of Γ_n. If for some $\epsilon > 0$, $\mathrm{dist}(C, \Gamma_n \backslash C) = 2(\delta_n + \epsilon)$, then $\{x : \mathrm{dist}(x, C) < \delta_n + \epsilon\}$ and $\{x : \mathrm{dist}(x, \Gamma_n \backslash C) < \delta_n + \epsilon\}$ are two disjoint open sets both containing points of Γ and whose union covers Γ. This contradicts the connectedness of Γ. Thus we can find $x \in C$ and $y \in \Gamma_n \backslash C$ such that $\|x - y\| \leq 2\delta_n$. Consider the straight line segment from x to y. It connects

C to some other connected component of Γ_n. Since C was an arbitrary connected component of Γ_n we can find a similar straight line segment from each connected component of Γ_n joining it to some other component. Now if we add all the line segments to Γ_n, (we can do this because we have assumed the domain is convex,) the number of connected components is reduced to $M/2$ or fewer, the Hausdorff measure will increase by at most $2M\delta_n$ and we will have $\mathrm{d}_H(\Gamma_n, \Gamma) \leq 2\delta_n$. Let p be the smallest integer such that $2^p \geq M$, then by repeating the above argument p times we get a modified, connected Γ_n such that its Hausdorff measure is at most $(2pM)\delta_n$ larger than before and $\mathrm{d}_H(\Gamma_n, \Gamma) \leq 2^p\delta_n$. Thus the modified Γ_n still converge to Γ and since they are connected we can apply theorem 2.4 to get, in terms of the original sequence

$$\mathcal{H}^1(\Gamma) \leq \liminf_{n\to\infty} \mathcal{H}^1(\Gamma_n) + 2pM\delta_n = \liminf_{n\to\infty} \mathcal{H}^1(\Gamma_n). \tag{3.1.5}$$

Lemma 3.2 implies lower semicontinuity of F and together with the above result we get lower semicontinuity of ν.

Existence Proof

In this section we will treat the question of the existence of a minimizing pair (f, Γ) for E. We have already developed some results for the cost associated strictly with the boundary so in this section we will be focusing on the function f. Since it may be desirable to introduce other costs associated with the boundary, we will state assumptions required on the boundaries in order to treat the remainder of the problem rather than quote results from the last section. We mention here however that these assumptions are satisfied by the definitions given in the preceding section; A1 is satisfied by definition and A2 follows from theorem 2.1 and theorem 3.3. Also, we will generalize the functional E. We will use the following set of assumptions on the space of boundaries.

A1 The space of boundaries is contained in the set of nonempty closed sets in $\Re^2$.

A2 With respect to the topology induced by the Hausdorff metric on the space of boundaries $\nu(\cdot)$ is a nonnegative lower semicontinuous, coercive functional. (I.e. the sets $\{\Gamma : \nu(\Gamma) \leq t\}$ are σ-compact for all $t \geq 0$.)

We now generalize the functional E somewhat, in anticipation of other applications. Henceforth E is defined by,

$$E(f,\gamma) = \int_{\Omega\backslash\Gamma} \Phi(g, f, D^{\alpha_1} f, D^{\alpha_2} f, \ldots, D^{\alpha_s} f) + \nu(\Gamma)$$

As before $g \in L_\infty(\Omega)$, s is a positive integer. Each α_i is a fixed multi-index, using the notation of [33]. f belongs to the subspace of functions in $L_{p_0}(\Omega\backslash\Gamma)$ whose distributional derivative $D^{\alpha_i} f$ exists as an $L_{p_i}(\Omega\backslash\Gamma)$ function, where each p_i satisfies $1 \le p_i < \infty$ for all $1 \le i \le s$. We will denote this space of functions by $\mathcal{D}(\Omega\backslash\Gamma)$. The following describes the assumptions on Φ.

A3 Φ is a nonnegative real function on $\Re^{2+s}$ such that for any fixed domain $\Omega' \subset \Omega$ and fixed $g \in L_\infty(\Omega)$ the functional $\int_{\Omega'} \Phi(g, f, v_1, v_2, \ldots, v_s)$ is a lower semicontinuous, coercive functional on $L_{p_0}(\Omega') \times L_{p_1}(\Omega') \times \ldots \times L_{p_s}(\Omega')$ with respect to the weak (product) topology. Furthermore $\int_\Omega \Phi(g, 0, 0, \ldots, 0) < \infty$.

We note that $(g-f)^2 + v_1^2 + v_2^2$ is such a function with $p_0 = p_1 = p_2 = 2$. The formulation presented in the introduction satisfies these conditions with $\mathcal{D}(\Omega\backslash\Gamma) = W^{1,2}(\Omega\backslash\Gamma)$.

We are now ready to introduce a notion of convergence on sequences of pairs $\{(f_n, \Gamma_n)\}$. The convergence of a sequence $\{(f_n, \Gamma_n)\}$ to (f, Γ) will imply $\Gamma_n \to \Gamma$ in the topology induced by the Hausdorff metric. Now, given any function $w_n \in L_p(\Omega\backslash\Gamma_n)$ let $\hat{w_n} \in L_p(\Omega\backslash\Gamma)$ be defined by extending w_n to Ω, setting it to zero on Γ_n and then restricting it to $\Omega\backslash\Gamma$. By $(f_n, \Gamma_n) \to (f, \Gamma)$ we mean $\Gamma_n \to \Gamma$ in the topology induced by the Hausdorff metric, $\widehat{f_n} \to f$ weakly in $L_{p_0}(\Omega\backslash\Gamma)$ and $\widehat{D^{\alpha_i} f_n} \to D^{\alpha_i} f$ weakly in $L_{p_i}(\Omega\backslash\Gamma)$ for each $1 \le i \le s$.

Lemma 3.4 Under assumptions A1, A2 and A3, for any E bounded sequence $\{(f_n, \Gamma_n)\}$, we can extract a subsequence (also denoted $\{(f_n, \Gamma_n)\}$) such that for some boundary Γ and some $f \in \mathcal{D}(\Omega\backslash\Gamma)$, we have $(f_n, \Gamma_n) \to (f, \Gamma)$.

Proof:

Assume the conditions of the Lemma and suppose we are given an E bounded sequence. We can assume there is some Γ such that $\Gamma_n \to \Gamma$ since otherwise by assumption A2 we can first extract a subsequence and find a boundary with this property. Since the sequence is E bounded we can conclude from A3 that the sequence $\{\int_{\Omega\backslash\Gamma}\Phi(g,\widehat{f_n},\widehat{D^{\alpha_1}f_n},\ldots,\widehat{D^{\alpha_s}f_n})\}$ is bounded. Hence, by A3, we can find functions $f\in L_{p_0}(\Omega\backslash\Gamma), v_1\in L_{p_1}(\Omega\backslash\Gamma),\ldots,v_s\in L_{p_s}(\Omega\backslash\Gamma)$ and a subsequence (which we still denote the same way) such that, $\widehat{f_n}\to f$ weakly in $L_{p_0}(\Omega\backslash\Gamma)$ and $\widehat{D^{\alpha_i}f_n}\to v_i$ weakly in $L_{p_i}(\Omega\backslash\Gamma)$ for each $1\leq i\leq s$. We claim that $f\in\mathcal{D}$ and $D^{\alpha_i}f=v_i$.

Let g be any test function in $\Omega\backslash\Gamma$, i.e. $g\in C_o^\infty(\Omega\backslash\Gamma)$. For convenience we also define $g=0$ on $\Omega\cap\Gamma$. Consider the subsequence extracted above. Since $\text{dist}(\text{supp}(g),\Gamma)>0$ (using A1) it follows that for all n sufficiently large $g|_{\Omega\backslash\Gamma_n}\in C_o^\infty(\Omega\backslash\Gamma_n)$ and $\hat{f}_n|_{\text{supp}(g)}=f_n|_{\text{supp}(g)}$. Thus along the subsequence we have,

$$\begin{aligned}\int_{\Omega\backslash\Gamma}v_ig = & \lim_{n\to\infty}\int_{\Omega\backslash\Gamma}\widehat{D^{\alpha_i}f_n}g & = \lim_{n\to\infty}\int_{\Omega\backslash\Gamma_n}D^{\alpha_i}f_ng \\ = & -\lim_{n\to\infty}\int_{\Omega\backslash\Gamma_n}f_nD^{\alpha_i}g & = -\lim_{n\to\infty}\int_{\Omega\backslash\Gamma_n}\widehat{f_n}D^{\alpha_i}g \\ = & -\int_{\Omega\backslash\Gamma}fD^{\alpha_i}g & \end{aligned}$$

We conclude from this that $D^{\alpha_i}f=v_i$ and hence $f\in\mathcal{D}(\Omega\backslash\Gamma)$.

Lemma 3.5 Let $\{(f_n,\Gamma_n)\}$ be any E bounded sequence such that $(f_n,\Gamma_n)\to(f,\Gamma)$, then under assumptions A1 and A2,

$$E(f,\Gamma)\leq\liminf_{n\to\infty}E(f_n,\Gamma_n) \tag{3.1.6}$$

Proof: Let Γ^ϵ be a closed ϵ neighborhood of Γ, i.e. a closed neighborhood of Γ such that $r(\Gamma^\epsilon,\Gamma)\leq\epsilon$ and define,

$$E_\epsilon(f,\Gamma')=\int_{\Omega\backslash\Gamma^\epsilon\cup\Gamma'}\Phi(g,f,D^{\alpha_1}f,\ldots,D^{\alpha_s}f)+\nu(\Gamma') \tag{3.1.7}$$

For n sufficiently large ($\geq N$ say), $\Gamma_n\subset\Gamma^\epsilon$ and since $\Gamma\subset\Gamma^\epsilon$ we get $\widehat{D^{\alpha_i}f_n}|_{\Omega\backslash(\Gamma^\epsilon\cup\Gamma_n)}=D^{\alpha_i}f_n|_{\Omega\backslash\Gamma^\epsilon}$. Hence the sequence $\{D^{\alpha_i}f_n|_{\Omega\backslash(\Gamma^\epsilon\cup\Gamma_n)}\}_{n\geq N}$ converges weakly to $D^{\alpha_i}f|_{\Omega\backslash\Gamma^\epsilon}$ in $L_{p_i}(\Omega\backslash\Gamma^\epsilon)$ and similarly $\{f_n|_{\Omega\backslash(\Gamma^\epsilon\cup\Gamma_n)}\}_{n\geq N}$ converges weakly to $f|_{\Omega\backslash\Gamma^\epsilon}$ in $L_{p_0}(\Omega\backslash\Gamma^\epsilon)$. We can now write

$$\begin{aligned}\liminf_{n\to\infty}E_\epsilon(f_n,\Gamma_n) & \geq \liminf_{n\to\infty}\int_{\Omega\backslash\Gamma^\epsilon\cup\Gamma_n}\Phi(g,f_n,D^{\alpha_1}f_n,\ldots,D^{\alpha_s}f_n)+\liminf_{n\to\infty}\nu(\Gamma_n) \\ & \geq \int_{\Omega\backslash\Gamma^\epsilon}\Phi(g,f,D^{\alpha_1}f,\ldots,D^{\alpha_s}f)+\nu(\Gamma) \\ & = E_\epsilon(f,\Gamma)\end{aligned} \tag{3.1.8}$$

where the second inequality follows from A_2 and lower semicontinuity of $\int \Phi$ in the weak topology on $\mathcal{D}(\Omega \backslash \Gamma^\epsilon)$. From the nonnegativity of Φ and the closedness of Γ we conclude $\sup_{\epsilon>0} E_\epsilon(\cdot) = E(\cdot)$ and hence

$$\liminf_{n\to\infty} E(f_n, \Gamma_n) \geq E(f, \Gamma) \tag{3.1.9}$$

Theorem 3.6 Under assumptions A1 and A2 (and in particular letting ν be defined as in the preceding section), there exists a minimizing pair, (f, Γ) for the functional E.

Proof: Apply Lemma 3.4 to a minimizing sequence, then apply Lemma 3.5.

It may be possible to show for some functionals that the number of connected components is necessarily finite for optimality even when the term $F(\#\Gamma)$ is absent from the definition of ν. If this is the case then the main results can be proved with this term removed.

The various proofs can be modified to allow the space of boundaries to be closed in Ω rather than $\Re^2$ once it is shown that $\mathcal{H}^1(\Gamma \cap \partial\Omega) = 0$ for the present formulation. This can be accomplished for special domains (such as rectangles) be a reflection argument.

3.1.3 Existence in SBV

The class of admissible f, Γ pairs can be enlarged beyond that required by the "weak formulation" by formulating the problem in SBV(Ω). In this case the functional appears as below,

$$E(f) = \beta \int_\Omega (g-f)^2 + \int_\Omega \|\nabla f\|^2 + \mathcal{H}^1(\mathcal{S}_f)$$

where $\nabla f\, dx$ is the part of Df which is absolutely continuous with respect to Lebesgue measure and $\mathcal{S}_f$ is the jump set of f (see chapter 2 for an introduction to SBV). L. Ambrosio proved a compactness theorem and a lower-semicontinuity theorem for the space SBV(Ω) which allows the assertion of existence of minimizers to this formulation of the variational segmentation problem. We will be using these theorems in the proofs of our asymptotic results in chapter 4. Thus for later reference we paraphrase these theorems here.

Theorem 3.7 [4, Theorem 3.1] Let $\{u_n\} \subset \mathrm{SBV}(\Omega) \cap L^\infty(\Omega)$ be a sequence such that

$$\limsup_{n\to\infty} \left\{ \|u_n\|_\infty + \int_\Omega |\nabla u_n|^2 \, dx + \mathcal{H}^{n-1}(\mathcal{S}_{u_n}) \right\} < \infty$$

Then, there exists a subsequence u_{n_k} converging in $L^1_{\text{loc}}(\Omega)$ to $u \in \mathrm{SBV}(\Omega)$. Moreover,

$$\begin{aligned} Ju_n &\to Jg && \text{weakly as radon measures} \\ \nabla u_n &\to \nabla u && \text{weakly in } L^1(\Omega; \Re^n). \end{aligned}$$

We mention in passing that if it can be shown that $u_n \to u$ in $L^1_{\text{loc}}(\Omega)$ by other means then the weak convergence results apply to the original sequence.

To complete the proof existence of $\mathrm{SBV}(\Omega)$ minimizers of E the following has been proved.

Theorem 3.8 [4, Theorem 4.2] If $u_n \to u$ in $L^1_{\text{loc}}(\Omega)$ with $\|u_n\|_\infty \le C < \infty$ then,

$$E(u) \le \liminf_{n\to\infty} E(u_n)$$

To make the connection between the "weak" formulation mentioned earlier and the SBV formulation the first step is to note that the SBV formulation is more general. The following proposition makes this assertion.

Proposition 3.9 [4, Proposition 3.3] Let $\Gamma \subset \Omega$ be a closed set such that $\mathcal{H}^{n-1}(\Gamma) < \infty$, and let $u \in W^{1,1}(\Omega \backslash \Gamma) \cap L^\infty(\Omega)$. Then, $u \in \mathrm{SBV}(\Omega)$ and $\mathrm{SBV}(\Omega) \subset \Gamma \cup N$ with $\mathcal{H}^{n-1}(N) = 0$.

A consequence of this proposition is that the minimum achieved under the SBV formulation is less than or equal to the infimum of the "weak" formulation. However, De Giorgi and coworkers in [10] proved an equivalence between this formulation and the weak formulation. This was achieved through a regularity theorem for SBV minimizers. We state the theorem here not only to complete this overview of the existence results but also because we intend to use the result later.

Theorem 3.10 [4, Theorem 5.1] Let $u \in \mathrm{SBV}(\Omega)$ be a minimizer of E. Then,

(i)
$$u \in L^\infty(\Omega), \;\; \|u\|_\infty \le \|g\|_\infty;$$

(ii)
$$u \in W^{2,p}(\Omega \backslash \overline{\mathcal{S}}_u)\,\forall p \in [1,\infty) \text{ and } \Delta u = \beta(u-g) \text{ in } \Omega \backslash \overline{\mathcal{S}}_u;$$

(iii)
$$\text{the function } \tilde{u}(x) = u^+(x) = u^-(x) \text{ belongs to } C^1(\Omega \backslash \overline{\mathcal{S}}_u);$$

(iv)
$$\mathcal{H}^{n-1}(\Omega \cap \overline{\mathcal{S}}_u \backslash \mathcal{S}_u) = 0.$$

The most difficult and interesting part of this theorem is the last statement. The proof uses the two theorems quoted in section 2.3. The most important result which is needed for the proof beyond what has already been mentioned in section 2.3 is that for SBV minimizers of E the set

$$\Omega_0 = \{x \in \Omega : \limsup_{\rho \to 0} \rho^{-1} \left[\int_{B_\rho(x)} |\nabla u|^2 dy + \mathcal{H}^1(\mathcal{S}_u \cap B_\rho(x)) \right] = 0\}$$

is open. This is established by showing that is $x \in \Omega_0$ then the conditions of theorem 2.8 as satisfied at x and in some neighborhood of x. But all points where these conditions are satisfied are in Ω_0 by theorem 2.8. Thus Ω_0 is open. To see why (iv) follows let $\Gamma = \Omega \backslash \Omega_0$:

$$\Gamma = \cup_{\delta > 0} \Gamma_\delta$$

where

$$\Gamma_\delta = \{x \in \Omega : \limsup_{\rho \to 0} \rho^{-1} \left[\int_{B_\rho(x)} |\nabla u|^2 dy + \mathcal{H}^1(\mathcal{S}_u \cap B_\rho(x)) \right] \ge \delta\}$$

Γ_δ has zero Lebesgue measure since,

$$\Gamma_\delta \subset \{x \in \Omega : \limsup_{\rho \to 0} \rho^{-2} \int_{B_\rho(x)} |\nabla u|^2 dy = \infty\}$$

A general result for Hausdorff measures [10] implies,

$$\int_B |\nabla u|^2 dy + \mathcal{H}^1(\mathcal{S}_u \cap B) \ge \delta \mathcal{H}^1(\Gamma_\delta \cap B)$$

for all $\delta > 0$ and Borel sets B. Thus by setting $B = \Gamma_\delta \backslash \mathcal{S}_u$ we obtain,

$$\mathcal{H}^1(\Gamma_\delta \backslash \mathcal{S}_u) = 0$$

and since δ is arbitrary $\mathcal{H}^1(\Gamma\backslash\mathcal{S}_u) = 0$. Finally since Γ is relatively closed in Ω we get $\Omega \cap \overline{\mathcal{S}}_u\backslash\mathcal{S}_u \subset \Gamma\backslash\mathcal{S}_u$ and (iv) is proved. (We point out that the theorem was originally proved in $\Re^n$ and not just in $\Re^2$.)

3.2 Variational Results

The calculus of variations has been applied primarily in [29] to characterize solutions to the variational problem. We summarize a few of the results here since our work is partially motivated be the desire to circumvent some of these constraints. We consider first the function f. In theorem 3.10 it is already stated that (in the weak sense) minimal f satisfy,

$$\Delta f = \beta(f - g) \text{ in } \Omega\backslash\Gamma$$

If in addition some regularity assumptions are made on the domain i.e. on Γ and $\partial\Omega$ then f must satisfy Neumann boundary conditions. Thus once the boundaries are determined we have an explicit way to calculate f. Most of the difficulties both in analysis and computation lie with the boundary term. Carrying out the analysis of the calculus of variations generally requires making certain assumptions on the regularity of minimizing Γ but once the assumptions are made many interesting results can be derived. The following constraints on Γ's which minimize E were proved by Mumford and Shah in [29]. They are illustrated in Figure 3.1.

- If Γ 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 Γ is composed of $C^{1,1}$ arcs then they meet $\partial\Omega$ only at an angle of 90^o.

- If Γ is composed of $C^{1,1}$ arcs then it never occurs that two arcs meet at an angle other than 180^o.

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

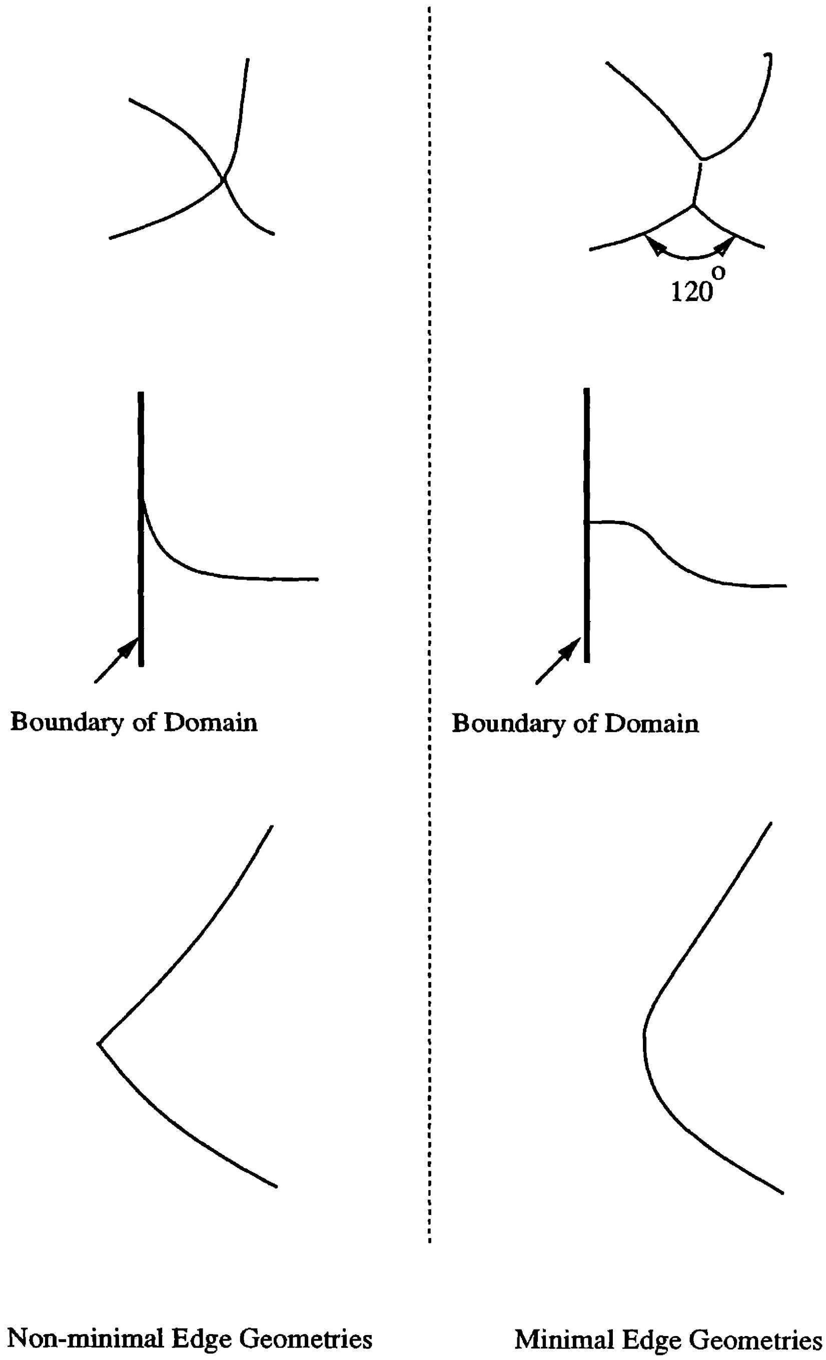


Figure 3.1: Calculus of Variations Results

These results help to characterize what minimizers of E (or E_0) should look like but they do so only on a local fashion. That is, they only say something about solutions on the level of microscopic detail. One of the contributions of our work is a set of theorems which show at least in some asymptotic sense that the solutions found by minimizing E and E_0 may be quite reasonable at the global level. The results show that as β or α tend to appropriate limits the minimizers of E and E_0 return solutions "close to" what is "appropriate" where appropriate is defined as the discontinuity set of the image.

3.3 A Γ–Convergent Approximation

In chapter 2 we sketched the concept of Γ–convergence. In this section we present such an approximation to the variational problem. It was proved in [5] that the sequence of functionals $\{E^n\}$, defined below, Γ–converges to E. We describe some of the ideas in the proof which can be found in [5].

In this approximation one replaces the set $\Gamma \subset \Omega$ with a function $v : \Omega \to [0, 1]$. There are various representations of the approximating functional. The one we choose is the following.

$$E^n(f,v) = \beta \int_\Omega (f-g)^2 + \int_\Omega (1-v^2)^n |\nabla f|^2 + \alpha (\int_\Omega (1-v^2)^n |\nabla v|^2 + \frac{n^2 v^2}{16}) \quad (3.3.1)$$

It is evident from the form of the functional that the discontinuity set of f is modeled by the set $(1-v^2)^n \simeq 0$. The smoothing term, $(1-v^2)^n|\nabla f|^2$ allows f to have large gradients wherever $(1-v^2)^n$ is near zero with low cost. The last integral in 3.3.1 prevents this from happening over a significant area, and in fact forces it in the limit to be confined to a one–dimensional set with the integral converging to the length (Minkowski content) of that set.

The definition of Γ–convergence requires the functionals be defined on a metric space. For this problem the appropriate space is $(f, v) \in \mathcal{B}(\Omega) \times \mathcal{B}(\Omega)$ with the (metrizable) topology of convergence in measure. The functionals E^n can not be shown to have minimizers on this space so a sub-domain is defined in which the requisite compactness and lower–semicontinuity results can be shown. This domain can be defined via a change of variables. We set,

$$\varphi(t) = \int_0^t (1-s^2)^{\frac{n}{2}}\, ds$$

and

$$\psi(s,t) = s(1-t^2)^{\frac{n}{2}+1}\, ds$$

Now, $|\nabla\varphi\circ v|^2 = |(n+2)v(1-v^2)^{\frac{n}{2}}\,\nabla v|^2 \leq (n+2)^2v^2 + |(1-v^2)^{\frac{n}{2}}\nabla v|^2$ by the chain rule and the Schwartz inequality. Also, $\nabla\psi(f,v) = (1-v^2)^{\frac{n+2}{2}}\,\nabla f + (n+2)fv(1-v^2)^{\frac{n}{2}}\,\nabla v$. Thus a natural domain of definition for functional E^n turns out to be,

$$\begin{aligned} D_n \quad = \quad & \{(f,v)\in\mathcal{B}(\Omega)\times\mathcal{B}(\Omega) : \varphi\circ v\in W^{1,2}(\Omega), \\ & \text{and } \phi(N\wedge u\vee -N, v)\in W^{1,2}(\Omega)\ \forall N\in\aleph\} \end{aligned}$$

The definition of E^n is extended to all of $\mathcal{B}(\Omega)\times\mathcal{B}(\Omega)$ by setting it equal to infinity outside of D_n. The importance of this construction is that the existence of minimizers of E^n can now be shown. The definition of E must be modified to include the function v. It is defined by setting $E(f,v) = E(f)$ where $v=0$ and setting it equal to infinity otherwise. With these definitions the existence of minimizers to E^n and Γ–convergence of E^n to E was shown in [5].

The essential insight into how the length term is approximated can be realized by considering the function,

$$c_n = (1-v_n^2)^{\frac{n}{2}+1}$$

The following inequality is fundamental,

$$\begin{aligned} \frac{1}{2}\int_\Omega |\nabla c_n| \quad &= \quad (\frac{n}{2}+1)\int_\Omega v_n(1-v_n^2)^{\frac{n}{2}}|\nabla v_n|\,dx \\ &\leq \quad \int_\Omega \frac{(n+2)^2}{16}v^2 + (1-v_n^2)^n|\nabla v_n|^2\,dx \end{aligned} \tag{3.3.2}$$

Consider the one–dimensional case. If $h(x)$ is any absolutely continuous function that is equal to 1 at $x=0$ and $x=b>0$, is equal to 0 at a with $0<a<b$, is monotonically decreasing on $(0,a)$ and monotonically increasing on (a,b) then $\frac{1}{2}\int_0^b |\nabla h(x)|\,dx = 1$. If the function c_n is 0 at the break points of f, increasing as a function of distance from

the break points, and the inequality 3.3.2 can be made into an equality by suitable choice of v, then we see that the term $\int_\Omega \frac{n^2}{16}v^2 + (1-v_n^2)^n|\nabla v_n|^2\,dx$ should closely approximate the number of break points.

For notational convenience we define for each measurable $A \subset \Omega$ the functional,

$$F^n(f,v,A) = \int_A (1-v^2)^n|\nabla f|^2 + \alpha\left(\int_\Omega (1-v^2)^n|\nabla v|^2 + \frac{n^2v^2}{16}\right)$$

All of the difficulty with respect to the proof of the convergence lies with the two terms in F^n. The missing fidelity term represents an easily handled perturbation of the problem restricted to these terms. Note that F^n also has a set variable. Similarly we define,

$$F(f,A) = \int_A (1-v^2)^n|\nabla f|^2 + \alpha\mathcal{H}^1(A \cap S_f)$$

for each f in SBV(Ω).

The proof of Γ–convergence is not restricted to the two dimensional case. However, again for simplicity of notation we will restrict ourselves to this case. We first describe how for every sequence $\{(f_n, v_n)\} \subset D_n$ which converges to $(f, 0)$ in measure the inequality,

$$\liminf_{n\to\infty} F_n(f_n, v_n) \geq F(f)$$

is established. It is a somewhat surprising fact that the problem can be reduced to the one–dimensional case. The central fact used to prove this reduction is,

$$F(f,\Omega) = \sup\left\{\sum_{i=1}^{\infty}\int_{A_i}|\langle\nabla f,\nu_i\rangle|^2\,dx + \int_{S_f\cap A_i}|\langle\nu_f,\nu_i\rangle|^2\,d\mathcal{H}^1\right\}$$

where $\{A_i\}$ is any sequence of pairwise disjoint open subsets of Ω and $v_i \in S^1$ for each $i \in \aleph$. For each A_i we can evaluate, $\int_{A_i}|\langle\nabla f,\nu_i\rangle|^2\,dx + \int_{S_f\cap A_i}|\langle\nu_f,\nu_i\rangle|^2\,d\mathcal{H}^1$ by considering slices in the direction ν_i through A_i.

The proof for the one–dimensional case has two steps. If $f \in W^{1,2}(B_\rho(x))$ for some $\rho > 0$ then it can be shown that,

$$\liminf_{n\to\infty} F^n(f_n, v_n, B_\rho(x)) \geq \int_{B_\rho(x)}|\nabla f|^2\,dx \tag{3.3.3}$$

The functions c_n converge to 1 almost everywhere but not necessarily uniformly. The trick used to circumvent this difficulty is to get a uniform bound on the number of connected components of the set $A_n^t = \{y \in B_\rho(x) : c_n(y) \geq t\}$ for some suitable t satisfying $\delta < t < \sigma$, for any $0 < \delta < \sigma < 1$. This t is found using the co–area formula. The Kuratowski limit (i.e. the limit with respect to the Hausdorff metric) of these sets (or a suitable subsequence) is then found to be a finite number of points which we denote P. It then follows that,

$$\liminf_{n\to\infty} \int_K (1-v_n^2)^n |\nabla f_n|^2 \, dx \geq \delta \int_K |\nabla f|^2 \, dx$$

for every $K \subset\subset B_\rho(x)\backslash P$. Letting $K \to B_\rho(x)\backslash P$ and $\delta \to 1$ one obtains 3.3.3.

The second half of the Γ–convergence conditions requires finding for each $f \in$ SBV(Ω) a sequence $\{(f_n, v_n) \in D_n\}$ such that $v_n \to 0$, $f_n \to f$ and,

$$\limsup_{n\to\infty} E^n(f_n, v_n) \leq E(f)$$

This is accomplished under the assumption,

$$\mathcal{H}^1(\mathcal{S}_f) = M^1(\mathcal{S}_f). \tag{3.3.4}$$

Now it turns out that minimizers of E actually satisfy this condition (see [27] [5]), so for our purposes there is no loss of generality under this assumption. The sequence constructed for the proof sets $f_n = f$ so we concentrate solely on v_n. The construction of v_n is achieved by defining $v_n = \tilde{v}_n \circ \tau$ where $\tau(x) = \text{dist}(x, \mathcal{S}_f)$ where $\tilde{v}_n$ is a function of a single real variable.

To allow for the discontinuities in f, v_n is set equal to 1 in $[\mathcal{S}_f]_{b_n}$ where the positive sequence b_n is chosen to satisfy $\lim_{n\to\infty} n^2 b_n = 0$ so that $\lim_{n\to\infty} \int_{[\mathcal{S}_f]_{b_n}} n^2 v_n^2 = 0$. In other words $\tilde{v}_n(t) = 1$ for $t \in [0, b_n]$. Outside of this neighborhood v_n will decay quickly to some small positive constant $\eta_n = \frac{2}{n}\sqrt{\int_0^1 (1-s^2)^n \, ds}$, which satisfies $\lim_{n\to\infty} n^2 \eta_n^2 = 0$ and (hence) $\lim_{n\to\infty}(1-\eta^2)^n = 1$. This decay will occur over some finite range a_n such that $\lim_{n\to\infty} a_n = 0$. These conditions imply that $\lim_{n\to\infty} \int_\Omega (1 - v_n^2)^n |\nabla f|^2 = \int_\Omega |\nabla f|^2 = 0$ and, $\lim_{n\to\infty} \int_{[\mathcal{S}_f]_{b_n}} \cup \Omega \backslash [\mathcal{S}_f]_{a_n+b_n} (1-v_n^2)^n |\nabla v_n|^2 + \frac{n^2 v_n^2}{16} = 0$. Thus the only remaining concern is the convergence of $\int_{[\mathcal{S}_f]_{a_n+b_n} \backslash [\mathcal{S}_f]_{b_n}} (1-v_n^2)^n |\nabla v_n|^2 + \frac{n^2 v_n^2}{16}$ which depends on the behavior of $\tilde{v}_n(t)$ for the range $t \in [b_n, b_n + a_n]$. Using the co-area formula this remaining contribution can be written in the form,

$$\int_{b_n}^{b_n+a_n} [|\nabla \tilde{v}_n(t)|^2 (1-\tilde{v}_n(t))^2 + \frac{n^2}{16} \tilde{v}_n^2(t)] \mathcal{H}^1(\{y : \tau(y) = t\}) \, dt \tag{3.3.5}$$

If $\mathcal{H}^1(\{y : \tau(y) = t\})$ were constant then the optimal solution could be calculated explicitly. Since this is not the case the exact solution is not available. However, it turns out that using the solution given under the assumption that $\mathcal{H}^1(\{y : \tau(y) = t\})$ is constant is sufficient to prove the bound. To this end we define the function θ as the solution to the differential equation,

$$\nabla\theta = \frac{n\theta}{4(1-\theta^2)^{\frac{n}{2}}}, \quad \theta(0) = \eta_n$$

The scalar a_n is now defined as the maximal existence interval of the solution. Since θ is monotonically increasing on $(0, a_n)$ we obtain,

$$a_n = 4\int_{\eta_n}^{1} \frac{(1-s^2)^{\frac{n}{2}}}{ns}\,ds \leq \frac{4}{n\eta_n}\int_0^1 (1-s^2)^{\frac{n}{2}} = 2\sqrt{\int_0^1 (1-s^2)^{\frac{n}{2}}\,ds}$$

Finally we define $\tilde{v}_n(t) = \theta_n(b_n + a_n - t)$ for $t \in [b_n, b_n + a_n]$. With this definition equation 3.3.5 reduces to,

$$\int_{b_n}^{b_n+a_n} [\frac{n^2}{8}\tilde{v}_n^2(t)]\mathcal{H}^1(\{y : \tau(y) = t\})\,dt \tag{3.3.6}$$

If we define $A(t) = |[S_f]_t|$ then $\nabla A = \mathcal{H}^1(\{y : \tau(y) = t\})$ almost everywhere. Also by assumption 3.3.4 we have $A(t) \leq 2t(\mathcal{H}^1(S_f) + w_n)$ for all $t \in [b_n, b_n + a_n]$ with $\lim_{n\to\infty} w_n = 0$. These facts allow one to integrate by parts the expression 3.3.6 two times, finally obtaining the desired result.

The manipulations above capture the essential form of minimal v. We wish to determine the (local) width of the regions $\{(1 - v_n^2)^n < t\}$. Let x_t be the solution to $\theta(x_t) = \sqrt{1 - t^{\frac{1}{n}}}$, we wish to determine $w_t = a_n - x_t$ since this is approximately one half of the width of the region $\{(1 - v_n^2)^n < t\}$. It is clearly true that θ is a monotonically increasing function on $[0, a_n)$, thus,

$$\begin{aligned} w_t &= \int_{\sqrt{1-t^{\frac{1}{n}}}}^{1} \frac{1}{\nabla\theta}\,d\theta \\ &= \int_{\sqrt{1-t^{\frac{1}{n}}}}^{1} \frac{4(1-\theta^2)^{\frac{n}{2}}}{n\theta}\,d\theta \end{aligned}$$

Now, if we make the change of variables $r = \frac{n}{2}\theta^2$ we obtain,

$$w_t = \frac{2}{n}\int_{\frac{n}{2}(1-t^{\frac{1}{n}})}^{\frac{n}{2}} \frac{(1-\frac{2r}{n})^{\frac{n}{2}}}{r}\,dr \tag{3.3.7}$$

Thus we obtain,

$$\int_{\frac{n}{2}(1-t^{\frac{1}{n}})}^{\frac{n}{4}} \frac{\exp -r}{r} \exp -\frac{2r^2}{n}\, dr \leq \frac{nw_t}{2} \leq \int_{\frac{n}{2}(1-t^{\frac{1}{n}})}^{\frac{n}{4}} \frac{\exp -r}{r}\, dr + 2^{-\frac{n}{2}} \tag{3.3.8}$$

It is a fact that $\lim_{n\to\infty} \frac{n}{2}(1-t^{\frac{1}{n}}) = -\ln\sqrt{t}$ and thus we get,

$$\lim_{n\to\infty} \frac{nw_t}{2} = \int_{-\ln\sqrt{t}}^{\infty} \frac{\exp -r}{r}\, dr$$

Actually, as indicated by 3.3.8 the rate of convergence of nw_t is quite fast since $\frac{\exp -r}{r}$ decays quickly. Simulations indicate that truly optimal v may not tend to 1 at the jump set of f as n tends to ∞. In fact one can have $v \to 0$ and $(1-v^2)^n \to 0$. However, the estimate of the width of neighborhoods of the boundaries is still good since most of the contribution to the integral 3.3.7 lies in range of small θ. These results allow us to conveniently find neighborhoods of minimal Γ. These neighborhoods play an important role in our segmentation algorithm.

We close this section by mentioning that an alternative Γ–convergence sequence of functionals can be defined by replacing $(1-v^2)^n$ with e^{-nv^2}. We claim that with this substitution the sequence of functions F^n still Γ–converge to F. The proof of this fact requires a reworking of the proof we have sketched above. This alternative form may be useful for computational purposes which we discuss in Chapters 5 and 6.

Chapter 4

On The Recovery of Discontinuities

4.1 Introduction

In this chapter we present the proofs of the limit theorems to which we refer throughout the rest of the thesis. The theorems are concerned with what happens to solutions of the variational formulation of the segmentation problem as $\beta \to \infty$. We prove that if the image is ideal i.e. a piecewise smooth or picewise constant function then the optimal boundaries Γ converge to the discontinuity set of image with respect to the Hausdorff metric. Furthermore we show that the results still hold if the image is corrupted by smearing and additive noise provided the smearing effect and the magnitude of the noise decay sufficiently quickly as β tends to infinity. We treat the piecewise constant case (i.e. minimizing E_0) and piecewise smooth case (i.e. minimizing E) separately. We also consider a hybrid case in which we study piecewise smooth segmentations where the image is essentially a piecewise constant function. The piecewise constant case admits a much more constructive proof using relatively elementary methods. These techniques may have applications in more explicit situations than those for which the theorems are proved. The proof for the piecewise smooth case on the other hand relies explicitly on the theorem 2.8 and since the universal constants used in the theorem are unknown we cannot quantify the convergence as explicitly as in the piecewise constant case.

There are two somewhat distinct motives for proving these theorems. The first is to prove an asymptotic fidelity result for the variational formulation which provides a counterpoint to the calculus of variations results quoted in chapter 3. The calculus of variations results constrain the boundary portion of solutions of the variational problem into certain geometries which do not conform with our expectations of "boundaries of objects". In particular they exclude the possibility of several desirable structures such as t–junctions. The second purpose is to yield information on how to govern the algorithms developed in chapters 5 which ideally find accurate boundaries even when looking only for large scale objects.

4.2 The Piecewise Constant Case

The presentation of the proof proceeds as follows. In the first section we state our assumptions and explicitly recall the particular variational problem we are considering concluding with a statement of the theorem to be proved. In the subsequent section we present some basic results which are then used in the proof of the main result, which occurs in the last section.

4.2.1 Problem Formulation

The Variational Problem

Let $\Omega \subset \Re^2$ be an open rectangle. We will be examining minimizers of

$$E_0(f,\Gamma) = \beta \sum_{k=1}^{\infty} \int_{\Omega_k} (g-f)^2 + \mathcal{H}^1(\Gamma)$$

where $\beta > 0$ is a real parameter, the $\Omega_k \in \Omega$ are disjoint open connected sets with $\Gamma = \Omega \backslash \cup_k \Omega_k$, and f is a function constant on each Ω_k. It is easy to see that minimality of E_0 requires $f = \frac{1}{|\Omega_k|} \int_{\Omega_k} g$ in Ω_k so minimizing solutions are determined by Γ and we will often refer to the solution Γ meaning the pair f, Γ. Also, we will be varying the parameter β and will use Γ_β to indicate an optimal solution for a particular value of β. For given g and β we define

$$E_0^*(g,\beta) = \inf\{E_0(f,\Gamma)\}$$

The goal of this section is to show that for g which are approximately piecewise constant the minimizers of E_0 return boundaries which approximate the "true" boundaries of g. The result is asymptotic in nature stating that for β sufficiently large and g sufficiently close to a piecewise constant function the Hausdorff distance (defined below) between the "true" boundaries of g and Γ is arbitrarily small. The techniques used here are elementary, admitting constructive analysis useful perhaps for computation and extension to non-asymptotic results. The assumptions are not the most general possible but to weaken them requires the introduction of more sophisticated techniques into the proof; also, they are certainly general enough for vision applications. In the proof of the piecewise smooth case which can be found later in this chapter, techniques suitable for proving a slightly more general theorem are used.

In chapter 3 we stated a theorem concerning the existence of minimizers of E_0 when Ω is an open rectangle (theorem 3.1). The theorem can be proved for more general domains but we do not wish to address this issue here and we will therefore simply assume that our domain is a rectangle. When we refer to a minimizer of E_0 we mean a set such as defined in theorem 3.1 i.e. $\Gamma \subset \Omega$ is composed of a finite number of $C^{1,1}$ curves joined only at their end points.

4.2.2 Assumptions on the Domain

The assumptions we require on the domain do not go beyond those needed for theorem 3.1, the existence theorem. For convenience we will therefore assume that our domain is an open rectangle. For the results of this section we will need the following isoperimetric inequality: there is a constant $\zeta > 0$ such that if $A \in \Omega$ is a Cacciopoli set then,

$$\mathcal{H}^1(\gamma \cap \Omega) \geq \zeta \min\{|A|^{\frac{1}{2}}, |\Omega \backslash A|^{\frac{1}{2}}\}.$$

We remark that it is enough that this inequality be satisfied when A is a polygon for it to hold for all Cacciopoli sets.

Suppose γ is a connected component of some minimizer Γ of E_0 which is some positive distance from $\partial\Omega$. Let O be the connected component of $\Re^2 \backslash \gamma$ containing $\Re^2 \backslash \Omega$. By *the set bounded by* γ we mean $F = \Re^2 \backslash O = \Omega \backslash O$. If γ is not separated from the boundary of Ω by a positive distance then $\overline{\gamma} \cap \partial\Omega$ is some finite set of points $p_1, \ldots, p_m$. Since the boundary of Ω is a Jordon curve we can assume the points are

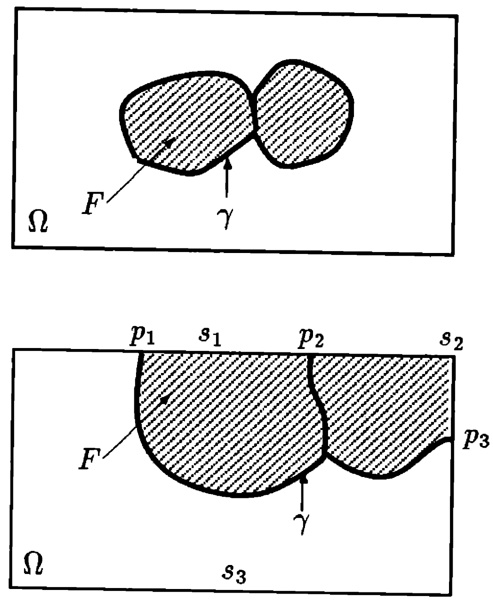


Figure 4.1: Sets Bounded by a Curve

ordered along the boundary. Thus $\partial\Omega\backslash\{p_1,\ldots,p_m\}$ consists of m segments of the boundary which we denote $s_1,\ldots,s_m$. For any $i \in \{1,\ldots,m\}$ we define the set O_i as the connected component of $\Re^2\backslash\{\gamma \cup (\partial\Omega\backslash s_i)\}$ containing $\Re^2\backslash\overline{\Omega}$. Set $F_i = \Omega\backslash O_i$. We can now define the set bounded by γ, F, to be an F_i of minimal area. (We choose it arbitrarily if it is not unique.) These definitions are illustrated in Figure 4.1. If $\mathcal{H}^1(\gamma) < \zeta\sqrt{\frac{|\Omega|}{2}}$ then we can conclude from the isoperimetric inequality that it is unique. In either case we have the isoperimetric inequality implies that,

$$\mathcal{H}^1(\gamma) \geq \zeta|F|^{\frac{1}{2}}. \tag{4.2.1}$$

Assumptions on the Image

The demands of the proof of the limit theorem requires that we make certain assumptions on the data g. The case we are interested in is one in which the image is a corrupted version of a piecewise constant L^∞ function g_c. We will define a set which we interpret as the natural candidate for a set of boundaries in the image.

We assume that Ω can be decomposed into a countable number of disjoint Cacciopoli sets A_j, such that on each A_j, g_c is constant. We define the boundary $\mathcal{B}_g$ to be $\Omega \cap \bigcup_j \partial^* A_j$. We assume $\mathcal{H}^1(\mathcal{B}_g) < \infty$. Without loss of generality we may assume that if $\mathcal{H}^1(\partial^* A_j \cap \partial^* A_i) > 0$ for $i \neq j$ then $g_c(A_j) \neq g_c(A_i)$ (if this fails replace A_j and A_i with their union.) The set $\{g_c(A_j)\}$ is countable and bounded; we denote it $\{a_i\}$ and define $R_i = \bigcup_{\{j: g_c(A_j)=a_i\}} A_j$. R_i is thus essentially the portion of the domain on which the image takes the value a_i.

Assumption 4.1 $\mathcal{H}^1(\mathcal{B}_g) < \infty$ and $\mathcal{H}^1(\overline{\mathcal{B}_g} \backslash \mathcal{B}_g) = 0$.

The first part of this assumption we stated earlier. The second part of the assumption is a mild regularity constraint and is actually used for a only small portion of the results. Furthermore, all of the results still hold without this assumption but with it the proofs become more elementary. Without loss of generality we assume each connected component of $\Omega \backslash \overline{\mathcal{B}_g}$ is contained in some single A_j, or that $(A_j)_1 \subset A_j$. ($(A_j)_1$ we recall is the set of points where A_j has density 1.) This is just a technical convenience to avoid having to make 'almost everywhere' statements throughout the proof of the limit theorem.

The observed image g which appears in E_0, will be a corrupted version of g_c. We are allowing for some smearing of the image and additive noise. To simplify the problem of controlling the effect of the smearing we make the following assumption;

Assumption 4.2 There is a constant $c_b < \infty$ such that $|[\mathcal{B}_g]_r \cap \Omega| < c_b r$.

This assumption can be dropped if we do not need to allow for smearing. Alternatively the decay in r can be weakened. The main reason for allowing for smearing is not to require the image to have actual jumps. The 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, which follows from theorem 2.5 and the fact that a continuum $\Gamma \in \Re^2$ satisfying $\mathcal{H}^1(\Gamma) < \infty$ is 1–rectifiable (see [12] or [11].)

Proposition 4.1 [12] Let Γ be a continuum in $\Re^2$ with $\mathcal{H}^1(\Gamma) < \infty$ then

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

The Noise Model

Let $\mathcal{S}_r$ be 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 $\Phi \in \mathcal{S}_r$ iff Φ has the property

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

An example of such a Φ would be a smoothing operator defined using a mollifier with support lying inside the ball of radius r, but nonlinear perturbations are also allowed. Rather than prove results for a single image our results will hold for all images belonging to some class which we will now define. We assume that any image g has a representation of the form,

$$g = \Phi(g_c) + \vartheta w \tag{4.2.2}$$

for some $\Phi \in \mathcal{S}_r$ and $w \in L^\infty$ with $\|w\|_\infty \leq 1$ and ϑ a real scalar. We will be allowing β to tend to infinity and we will need to make corresponding assumptions on r and ϑ. To this end we suppose that we are given any two functions, $h_r : (0, \infty) \to [0, \infty)$ and $h_\vartheta : (0, \infty) \to [0, \infty)$, and positive constants $c_r, c_\vartheta < \infty$ such that,

$$\begin{aligned} h_r(\beta) &\leq c_r\beta^{-1} \\ h_\vartheta(\beta) &\leq c_\vartheta\beta^{-\frac{1}{2}} \\ \lim_{\beta\to\infty} \beta h_r(\beta) &= 0 \\ \lim_{\beta\to\infty} \beta^{\frac{1}{2}} h_\vartheta(\beta) &= 0 \end{aligned}$$

We say $g \in \Upsilon(\beta)$ if and only if g satisfies 4.2.2 for some Φ_r, w and ϑ with $r \leq h_r(\beta)$ and $\vartheta \leq h_\vartheta(\beta)$. We will assume in general that we always have $g \in \Upsilon(\beta)$. Actually only the last two of the conditions on h_r and h_ϑ stated above are needed for the limit theorem but the first two help to make the analysis more general in some instances. Further, for convenience we assume that $\vartheta \leq 1$ and define $K = 2 + \text{ess sup } g_c - \text{ess inf } g_c$. ($K$ bounds the gap between the maximum and minimum of g.)

We can now state the limit theorem which is to be proved for the piecewise constant formulation.

Theorem 4.2 Under our stated assumptions 4.1 and 4.2, as $\beta \to \infty$ $\{\Gamma_\beta\}$ converges to $\overline{\mathcal{B}_g}$ with respect to the Hausdorff metric, and $\mathcal{H}^1(\Gamma_\beta) \to \mathcal{H}^1(\mathcal{B}_g)$. We mean by this that for any $\epsilon > 0$ there exists $\beta' > \infty$ such that if $\beta > \beta'$ and Γ_β is a minimizer of E_0 for some $g \in \Upsilon(\beta)$, then $d_H(\Gamma_\beta, \mathcal{B}_g) < \epsilon$ and $|\mathcal{H}^1(\Gamma_\beta) - \mathcal{H}^1(\mathcal{B}_g)| < \epsilon$. Furthermore $\sqrt{\beta}(f - g_c)$ converges to 0 in $L^2(\Omega)$.

4.2.3 Preliminary Results

In this section we prove some inequalities which will be of importance for the development of the main results, which occurs in the next section.

Lemma 4.3 Let $\{C_n\}$ be a sequence of closed subsets of Ω such that each is composed of at most $N < \infty$ (N is arbitrary) connected components of a minimizer of E_0 for some g and some β, then there exists a subsequence (which we denote the same way) and a closed set $C \in \overline{\Omega}$ such that $d_H(C_n, C) \to 0$ and $\mathcal{H}^1(C) \leq \liminf_{n\to\infty} \mathcal{H}^1(C_n)$.

Proof Because of the conditions on minimizers of E_0 we have $\mathcal{H}^1(\overline{C_n}) = \mathcal{H}^1(C_n)$ (where the closure is taken in $\Re^2$). The number of connected components of $\overline{C_n}$ is bounded above by the number of connected components of C_n. By applying theorem 2.1 we first extract a convergent subsequence of $\overline{C_n}$ with the limit C. In lemma 3.2 and theorem 3.3 it was shown under these conditions that $\mathcal{H}^1(C) \leq \liminf_{n\to\infty} \mathcal{H}^1(\overline{C_n})$. Noting $d_H(\overline{C_n}, C) = d_H(C_n, C)$ we get the desired result. □

This lemma allows us to capture limits of minimizers of E_0. The following proposition will be used later to get bounds on $f - g_c$.

Proposition 4.4 Given a countable set $\{a_i : i = 0, 1, \ldots\} \subset \Re$, a nonnegative l_1 sequence $\{r_i : i = 1, 2, \ldots\}$ and constants $c_1, c_2 > 0$, there exists a nondecreasing function $h : (0, \infty) \to [0, \infty)$ satisfying $\lim_{t\to 0^+} \frac{h(t)}{\sqrt{t}} = 0$ such that for any sequence $\{p_i : i = 0, 1, \ldots\}$ satisfying,

$$p_0 \geq c_1 \text{ and } r_i \geq p_i \geq 0 \text{ for } i > 0;$$

$$\sum_{i=0}^{\infty} p_i = 1 \text{ and}$$

$$\sum_{i=0}^{\infty} p_i(a_i - \hat{a})^2 < c_2 t \text{ where } \hat{a} = \sum_{i=0}^{\infty} p_i a_i,$$

$$\text{we have } |\hat{a} - a_0| < h(t)$$

Proof We define the constant $b = \sum_{i=1}^{\infty} r_i$. We assume $b > 0$ (the result is trivial otherwise). Define $h_1 : (0, \infty] \to [0, b]$ by,

$$h_1(t) = \sum_{i:0<|a_i - a_0|<t} r_i.$$

Clearly $h_1(t)$ is nondecreasing in t. We that claim h_1 is continuous from the left and $\lim_{t \to 0^+} h_1(t) = 0$. For any $\epsilon > 0$, $\exists N < \infty$ such that $\sum_{i=N}^{\infty} r_i < \epsilon$. For $t \leq \min\{|a_i - a_0| : 0 < i < N\}$ we have $h_1(t) < \epsilon$ proving the second part of the claim. Given $t > 0$ let $\delta = \min\{t - |a_i - a_0| : 0 < i < N \text{ and } t - |a_i - a_0| > 0\}$; for $t' \in (t - \delta, t)$ we have $h_1(t) - h_1(t') < \epsilon$, proving the first part of the claim.

Define $h_2 : (0, \infty) \to [b^{-\frac{1}{2}}, \infty)$ by

$$h_2(t) = \sup\{c : h_1(x\sqrt{t}) \leq \frac{1}{x^2}, \forall x < c\}$$

Since h_1 is nondecreasing and bounded above by b, h_2 is nonincreasing and bounded below by $b^{-\frac{1}{2}}$. h_2 is finite for finite t since $\lim_{x \to \infty} h_1(x) = b$. For any $N < \infty$, $\exists \eta > 0$ such that $\epsilon \leq \eta \Rightarrow h_1(\epsilon) \leq \frac{1}{N^2}$. Thus $N\sqrt{t} \leq \eta \Rightarrow h_1(N\sqrt{t}) \leq \frac{1}{N^2}$ and since h_1 is nondecreasing while $\frac{1}{x^2}$ is decreasing we have $t \leq (\frac{\eta}{N})^2 \Rightarrow h_2(t) \geq N$. We conclude $\lim_{t \to 0^+} h_2(t) = \infty$. Also since h_1 is continuous from the left we have $0 < x \leq h_2(t) \Rightarrow h_1(x\sqrt{t}) \leq \frac{1}{h_2^2(t)}$.

Define $h_3 : (0, \infty) \to [b^{-\frac{1}{2}}, \infty)$, by

$$h_3(t) = \max(b^{-\frac{1}{2}}, h_2(t) - \sqrt{\frac{c_2}{c_1}})$$

Consider the case $\hat{a} \geq a_0$. We have

$$\sum_{\{i:a_i \leq \hat{a}\}} p_i(\hat{a} - a_i) = \sum_{\{i:0<a_i-\hat{a}<h_3(t)\sqrt{t}\}} p_i(a_i - \hat{a}) + \sum_{\{i:h_3(t)\sqrt{t} \leq a_i - \hat{a}\}} p_i(a_i - \hat{a})$$

$$\leq \sum_{\{i:0<|a_i-a_0|<h_2(t)\sqrt{t}\}} p_i|a_i-\hat{a}| + \frac{1}{h_3(t)\sqrt{t}}\sum_{i=0}^{\infty} p_i(a_i-\hat{a})^2$$
$$\leq h_1(h_2(t)\sqrt{t})h_2(t)\sqrt{t} + \frac{c_2 t}{h_3(t)\sqrt{t}}$$
$$\leq \left(\frac{1}{h_2(t)} + \frac{c_2}{h_3(t)}\right)\sqrt{t} \qquad (4.2.3)$$

For the first inequality we used the obvious bound; $|\hat{a}-a_0| < \sqrt{\frac{c_2}{c_1}t}$. Define $h(t) = \frac{1}{c_1}(\frac{1}{h_2(t)} + \frac{c_2}{h_3(t)})\sqrt{t}$. That h is nondecreasing follows from the fact that h_2 (and hence h_3) is nonincreasing. Also since $\lim_{t\to 0+} h_2(t) = \infty$, and hence $\lim_{t\to 0+} h_3(t) = \infty$, it follows $\lim_{t\to 0+} \frac{h(t)}{\sqrt{t}} = 0$. Now, $p_0(\hat{a}-a_0) \leq \sum_{\{i:a_i\leq \hat{a}\}} p_i(\hat{a}-a_i)$ and from 4.2.3 we get $|a_0-\hat{a}| \leq h(t)$. The case $\hat{a} \leq a_0$ can be treated similarly yielding the same result. This completes the proof of the proposition. □

The set $\{a_i\}$ will represent the set of values of the image g_c. If $\{a_i\}$ has finite cardinality then slightly stronger results can be obtained using the same line of proof.

Lemma 4.5 There is a constant $c_{E_0} < \infty$ and a function $U : (0,\infty) \to [0,\infty)$ satisfying $U(\beta) \leq c_{E_0}$ and $\lim_{\beta\to\infty} U(\beta) = \mathcal{H}^1(\mathcal{B}_g)$ such that $\sup_{g\in\Upsilon(\beta)}\{E_0^*(g,\beta)\} \leq U(\beta)$.

Proof For any measurable set $A \subset \Omega$ we have the following,

$$\int_A (g-g_c)^2 = \int_A \chi_{[\mathcal{B}_g]_r}(g-g_c)^2 + \int_A (1-\chi_{[\mathcal{B}_g]_r})(\vartheta w)^2$$
$$\leq K^2 c_b r + \vartheta^2|A|$$

where we invoke assumption 4.2. Thus for $g \in \Upsilon(\beta)$ we now obtain,

$$E_0^*(g,\beta) \leq E_0(g_c,\overline{\mathcal{B}_g}) \leq \beta\left(K^2 c_b h_r(\beta) + h_\vartheta^2(\beta)|\Omega|\right) + \mathcal{H}^1(\mathcal{B}_g)$$

The result now follows from our assumptions on h_r and h_ϑ. □

For each Ω_k in a segmentation we define the following constants,

$$p_i^k \triangleq \frac{|\Omega_k \cap R_i|}{|\Omega_k|}$$

Note that $\sum_{i=1}^{\infty} p_i^k = 1$.

Lemma 4.6 Given $\xi > 0$ and $i \geq 0$ there exists a function $H : (0,\infty) \to [0,\infty)$ satisfying $\lim_{\beta\to\infty} \sqrt{\beta}H(\beta) = 0$ such that if Γ_β is a minimizer of E_0 for some $g \in \Upsilon(\beta)$ and Ω_k is a connected component of $\Omega\backslash\Gamma_\beta$ satisfying $|\Omega_k \cup R_i| \geq \xi$, then

$$|f(\Omega_k) - a_i| \leq H(\beta)$$

Proof For convenience of notation we set $i = 0$ and re-enumerate the other a_i starting from 1. Let $\hat{a} = \sum_{i=0}^{\infty} p_i^k a_i = \frac{1}{|\Omega_k|}\int_{\Omega_k} g_c$. We have,

$$|f(\Omega_k) - a_0| \leq |f(\Omega_k) - \hat{a}| + |\hat{a} - a_0|$$

We can bound the first term as follows,

$$\begin{aligned} |f(\Omega_k) - \hat{a}| &= \left|\frac{\vartheta}{|\Omega_k|}\int_{\Omega_k} w + \frac{1}{|\Omega_k|}\int_{\Omega_k}(\Phi_r g_c - g_c)\right| \\ &\leq h_\vartheta(\beta) + \left|\frac{1}{|\Omega_k|}\int_{[B_g]_r}(\Phi_r g_c - g_c)\right| \\ &\leq h_\vartheta(\beta) + \frac{Kc_b}{\xi}h_r(\beta). \end{aligned}$$

since $|\Omega_k| > \xi$. To bound the second term we first note,

$$\sum_{i=0}^{\infty} p_i^k(\hat{a} - a_i)^2 = \int_{\Omega_k}(\hat{a} - g_c)^2.$$

and since $\hat{a} = \frac{1}{|\Omega_k|}\int_{\Omega_k} g_c$ we obtain,

$$\begin{aligned} \frac{1}{2}\int_{\Omega_k}(\hat{a} - g_c)^2 &\leq \frac{1}{2}\int_{\Omega_k}(f - g_c)^2 \\ \frac{1}{2}\int_{\Omega_k}(\hat{a} - g_c)^2 &\leq \int_{\Omega_k}(f - g)^2 + \int_{\Omega_k}(g - g_c)^2 \\ &\leq \beta^{-1}U(\beta) + K^2 c_b h_r(\beta) + |\Omega_k| h_\vartheta^2(\beta) \end{aligned}$$

where U is the function from lemma 4.5. Applying our assumptions and lemma 4.5 we see that there is a constant $c > 0$ such that

$$\sum_{i=0}^{\infty} p_i^k(\hat{a} - a_i)^2 \leq c\beta^{-1}$$

Define $r_i = \min\{1, \frac{|R_i|}{\xi}\}$. Clearly $p_i^k \leq r_i$ and we have $\sum_{i=1}^{\infty} r_i < \infty$. Note also that $p_0^k \geq \frac{\xi}{|\Omega|}$. We can now apply proposition 4.4 to conclude there exists a function $h : (0,\infty) \to [0,\infty)$ satisfying $\lim_{\beta\to\infty}\sqrt{\beta}h(\beta) = 0$ such that $|\hat{a} - a_0| \leq h(\beta)$. Set $H = h + h_\vartheta + \frac{Kc_b}{\xi}h_r$ and the result follows. □

Corollary *Under the conditions of lemma 4.6 we have*

$$\int_{\Omega_k}(a_i - g)^2 - \int_{\Omega_k}(f - g)^2 \leq |\Omega_k| H^2(\beta)$$

Proof

$$(a_i - g)^2 - (f - g)^2 = (a_i - f)^2 + 2(a_i - f)(f - g)$$

Since $a_i - f$ is constant in Ω_k and $f(\Omega_k) = \frac{1}{|\Omega_k|}\int_{\Omega_k} g$ we get,

$$\int_{\Omega_k}(a_i - g)^2 - \int_{\Omega_k}(f_i - g)^2 = \int_{\Omega_k}(a_i - f)^2.$$

Now apply lemma 4.6.

□

4.2.4 Main Results

The goal of this section is to prove theorem 4.2. The proof of this theorem is quite long so we have broken it up into several steps. We assume throughout that $r \leq h_r(\beta)$ and $\vartheta \leq h_\vartheta(\beta)$.

Lemma 4.7 For any $\epsilon, \gamma > 0$, there exists a constant $\beta' < \infty$ such that if $\beta > \beta'$ and Γ_β is a minimizer of E_0 for some $g \in \Upsilon(\beta)$, then $\tilde{\Gamma}_\beta \subset [\mathcal{B}_g]_\epsilon$ where $\tilde{\Gamma}_\beta$ is the union of all the connected components of Γ_β having $\mathcal{H}^1$ measure greater than γ.

Proof Without loss of generality we assume $\gamma \leq \frac{\epsilon}{4}$. Assume the lemma is false. Then, we can find a sequence $\tilde{\Gamma}_{\beta_n}$ with $\beta_n \uparrow \infty$ such that $\tilde{\Gamma}_{\beta_n} \not\subset [\mathcal{B}_g]_\epsilon$ for each n. In general $\mathcal{H}^1(\Gamma_\beta) \leq E_0^*(g, \beta) \leq c_{E_0}$, by lemma 4.5. Thus the number of connected components of $\tilde{\Gamma}_{\beta_n}$ is bounded above by $\frac{c_{E_0}}{\gamma}$. Applying lemma 4.3 we can assume that the $\tilde{\Gamma}_{\beta_n}$ converge with respect to the Hausdorff metric to a closed set $\tilde{\Gamma} \subset \overline{\Omega}$ satisfying $\mathcal{H}^1(\tilde{\Gamma}) < \infty$. It follows that there is an $x \in \tilde{\Gamma}$ such that $\text{dist}(x, \mathcal{B}_g) \geq \epsilon$. By translation we can assume $x = 0$ and we henceforth drop it from the notation. Since all connected components of $\tilde{\Gamma}_{\beta_n}$ have $\mathcal{H}^1$ measure greater than or equal to γ and since the $\mathcal{H}^1$ measure of a continuua is bounded below by the diameter of the continuua we obtain, $\liminf_{n\to\infty} \mathcal{H}^1(\tilde{\Gamma}_{\beta_n} \cap B_\gamma) \geq \gamma$ and hence

$$\liminf_{n\to\infty} \mathcal{H}^1(\Gamma_{\beta_n} \cap B_{\epsilon/4}) \geq \gamma \tag{4.2.4}$$

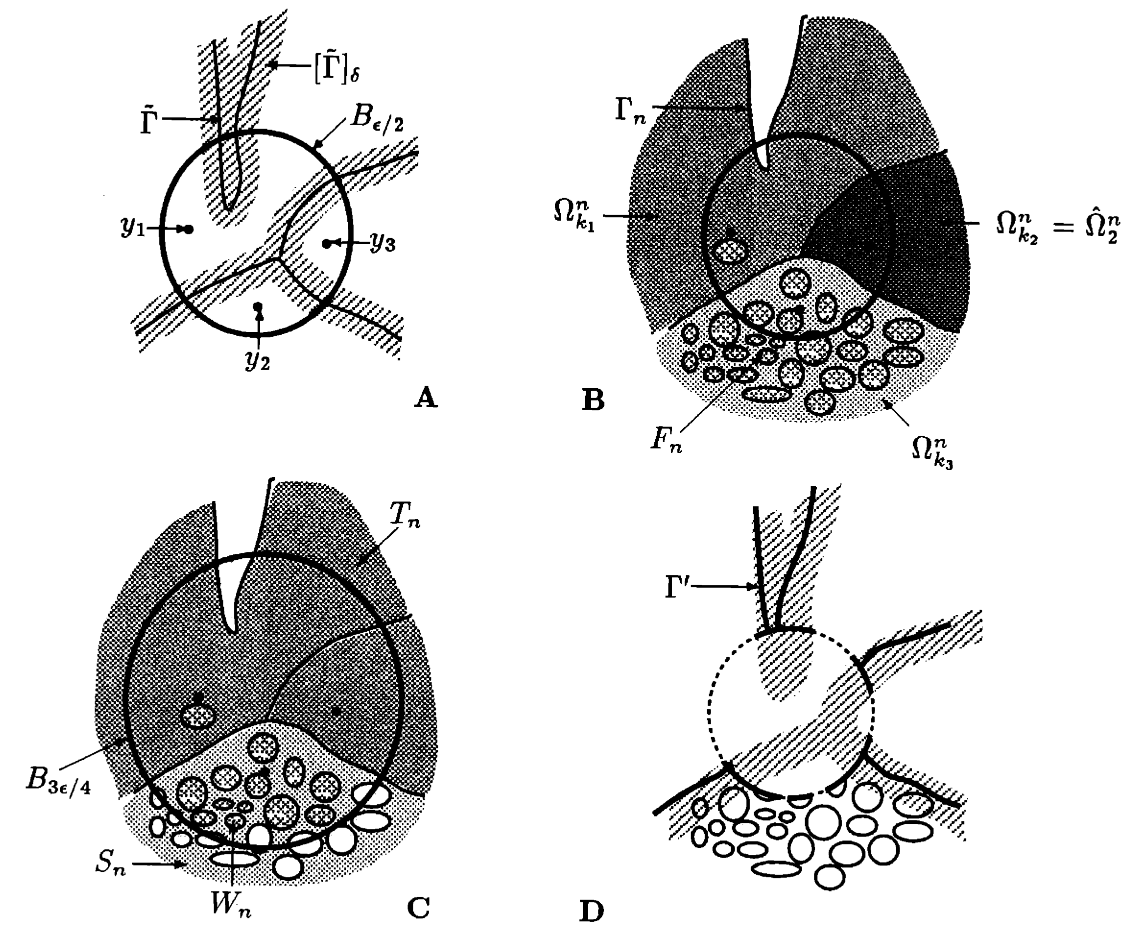


Figure 4.2: Illustration for Lemma 4.7

In the remainder of the proof of the lemma we show that by removing some Γ from B_ϵ that we can reduce the energy, thus obtaining a contradiction.

We can choose δ, $0 < \delta < \epsilon/2$ sufficiently small so that

$$|\overline{[\tilde{\Gamma}]_\delta} \cap B_{\epsilon/2}| < \gamma \frac{\epsilon}{2^5} \tag{4.2.5}$$

It is necessarily the case that $0 = x \in \overline{A_j}$ for some A_j. $A_j \backslash \tilde{\Gamma}$ is the union countable number of disjoint connected open sets. Only finitely many of these sets are not contained in $[\tilde{\Gamma}]_\delta$. Let m be the number of these sets that have nonempty intersection with $B_{\epsilon/2}$. We can find points $y_1, \ldots y_m$ such that y_i lies in the interior of the ith component and $\min_{i\in 1,\ldots,M} \operatorname{dist}(y_i, \tilde{\Gamma}) \geq \delta$. See Figure 4.2 **A** for an illustration. For n sufficiently large $\tilde{\Gamma}_{\beta_n} \subset [\tilde{\Gamma}]_\delta$ and we can then define $\tilde{\Omega}_i^n$ to be the connected component of $\Omega\backslash\tilde{\Gamma}_{\beta_n}$ which contains y_i. (They may be identical for different i .) Each $\tilde{\Omega}_i^n$ may contain connected components of Γ_{β_n} having $\mathcal{H}^1$ measure less than or equal to γ. Let F_n be the union of the sets bounded by these components (see the section entitled Assumptions on the Domain for an explanation of this terminology and Figure 4.2 **B** for an illustration). It follows that $\tilde{\Omega}_i^n\backslash F_n$ is some $\Omega_k(n)$ which we denote $\Omega_{k_i}^n$. By reordering the $\Omega_{k_i}^n$ with respect to i and choosing an appropriate subsequence (which we still index by n) we can find $m' \leq m$ and $\xi > 0$ such that;

$$\lim_{n\to\infty} |\Omega_{k_i}^n \cap R_l| = 0 \quad \text{for} \quad i \in 1,\ldots,m'$$
$$\liminf_{n\to\infty} |\Omega_{k_i}^n \cap R_l| \geq \xi \quad \text{for} \quad i \in m'+1,\ldots,m$$

where l is given by $g_c(A_j) = a_l$, (and $A_j \in R_l$). (Note that we can not have $|\hat{\Omega}_i^n \cap R_l| \to 0$ so $|\Omega_{k_i}^n \cap R_l| \to 0$ can only occur if $\hat{\Omega}_{k_i}^n$ is filled up with small sets bound by components of Γ having $\mathcal{H}^1$ measure smaller than γ.) Define S_n and T_n by,

$$S_n = \cup_{i=1}^{m'} \overline{\Omega_{k_i}^n}$$

$$T_n = \cup_{i=m'+1}^{m} \Omega_{k_i}^n$$

Note that $|\overline{\Omega_{k_i}^n}| = |\Omega_{k_i}^n|$ so

$$\lim_{n\to\infty} |S_n \cap R_l| = 0 \tag{4.2.6}$$

These definitions are illustrated in Figure 4.2 **C**. Let H be the function whose existence is asserted in lemma 4.6 with ξ as above and $i = l$. There exists N such that if $n > N$ then the following are all satisfied;

$$|(B_{\epsilon/2}\backslash B_{\epsilon/4}) \cap ([\tilde{\Gamma}]_\delta \cup S_n)| \quad < \quad \gamma\frac{\epsilon}{2^4} \tag{4.2.7}$$
$$\mathcal{H}^1(\Gamma_{\beta_n} \cap B_{\epsilon/4}) \quad > \quad \frac{3}{4}\gamma \tag{4.2.8}$$
$$\tilde{\Gamma}_{\beta_n} \quad \subset \quad [\tilde{\Gamma}]_\delta \tag{4.2.9}$$
$$\int_{T_n} (a_i - g)^2 - (f_n - g)^2 \quad \leq \quad |T_n| H^2(\beta) \tag{4.2.10}$$
$$h_r(\beta_n) \quad < \quad \frac{\epsilon}{4} \tag{4.2.11}$$

Equation 4.2.7 follows from 4.2.5 and 4.2.6. The inequality 4.2.8 follows from 4.2.4 and 4.2.9 follows by definition of $\tilde{\Gamma}$. Finally 4.2.10 follows from the corollary to lemma 4.6.

Consider any $n > N$. Since $\mathcal{H}^1(\Gamma_{\beta_n} \cap \partial B_\rho) > 0$ for at most countably many $\rho \in [\epsilon/4, \epsilon/2]$ we can conclude from 4.2.7 that for some $\rho_n \in [\epsilon/4, \epsilon/2]$

$$\mathcal{H}^1(\partial B_{\rho_n} \cap (\overline{[\tilde{\Gamma}]_\delta} \cup S_n \cup \Gamma_{\beta_n})) < \frac{\gamma}{4} \tag{4.2.12}$$

Note by 4.2.9 that any connected component of Γ_{β_n} which has nonempty intersection with $\partial B_{\rho_n} \backslash [\tilde{\Gamma}]_\delta$ has $\mathcal{H}^1$ measure and hence diameter less than or equal to γ which itself is bounded by $\epsilon/4$ and since $\rho_n \leq \epsilon/2$ such a connected component must lie entirely within $\Omega \cap \overline{B_{\frac{3}{4}\epsilon}}$, which in turn lies in A_j. Let $W_n = \cup\{\Omega_k(n) : \Omega_k(n) \subset B_{\frac{3}{4}\epsilon}\}$. This definition is depicted in Figure 4.2 **C**. We have

$$\overline{B_{\rho_n}} \subset T_n \cup S_n \cup W_n \cup \Gamma_{\beta_n} \cup \overline{[\tilde{\Gamma}]_\delta}$$

Now, define

$$\Gamma'_{\beta_n} = (\Gamma_{\beta_n} \backslash B_{\rho_n}) \cup (\partial B_{\rho_n} \cap (\overline{[\tilde{\Gamma}]_\delta} \cup S_n \cup \Gamma_{\beta_n}))$$

The construction of Γ' is illustrated in Figure 4.2 **D**. Let

$$f'_n(x) = \begin{cases} a_i & x \in T_n \cup W_n \cup B_{\rho_n} \\ f_n & \text{elsewhere} \end{cases}$$

It follows that f'_n is constant on each connected component of $\Omega \backslash \Gamma'_{\beta_n}$.

From 4.2.12 and 4.2.8 we have

$$\mathcal{H}^1(\Gamma_{\beta_n}) - \mathcal{H}^1(\Gamma') \geq \frac{\gamma}{2}$$

and from 4.2.10 and 4.2.11 and the fact $\Omega \cap B_{\frac{3}{4}\epsilon} \subset R_l$ we get

$$\begin{aligned} \int_\Omega (f'_n - g)^2 - \int_\Omega (f_n - g)^2 &= \int_{T_n \cup W_n \cup B_{\rho_n}} (f'_n - g)^2 - (f_n - g)^2 \\ &\leq \int_{T_n} (a_i - g)^2 - (f_n - g)^2 + \int_{B_{\frac{3}{4}\epsilon} \backslash T_n} (a_i - g)^2 \\ &\leq |T_n| H^2(\beta_n) + \pi(\frac{3}{4}\epsilon)^2 h_\vartheta^2(\beta_n) \end{aligned}$$

where in the last step we have used equation 4.2.10. We can now write,

$$E_0(f'_n, \Gamma'_n) - E_0(\Gamma_{\beta_n}) \leq -\frac{\gamma}{2} + \beta_n \left(|T_n| H^2(\beta_n) + \pi(\frac{3}{4}\epsilon)^2 h_\vartheta^2(\beta_n) \right)$$

which is negative for n sufficiently large (and hence β_n sufficiently large) contradicting the optimality of Γ_{β_n}. We conclude $\bar{\Gamma}_{\beta_n} \cap \Omega \backslash [\mathcal{B}_g]_\epsilon = \emptyset$ for all n sufficiently large. □

Theorem 4.8 For any $\epsilon > 0$ there exists a constant $\beta' < \infty$ such that if $\beta > \beta'$ then $\Gamma_\beta \subset [\mathcal{B}_g]_\epsilon$ for any Γ_β which minimizes E_0 with $g \in \Upsilon(\beta)$.

Proof Assume the theorem is false. There exists some $\epsilon > 0$ and a sequence of minimizing Γ_{β_n} with $\beta_n \uparrow \infty$ such that $\Gamma_{\beta_n} \not\subset [\mathcal{B}_g]_\epsilon \neq \emptyset$ for each n. Since only finitely many A_j can satisfy $A_j \not\subset [\mathcal{B}_g]_\epsilon$ there exists some single A_j and a subsequence (which we denote the same way) such that $\Gamma_{\beta_n} \cap A_j \backslash [\mathcal{B}_g]_\epsilon \neq \emptyset$ for each n. Let G represent an arbitrary connected component of $A_j \backslash [\mathcal{B}_g]_\epsilon$. It follows that G is a subset of some connected component of $A_j \backslash [\mathcal{B}_g]_{\epsilon/2}$ which we denote $\hat{G}$. It follows that $|\hat{G}| \geq \frac{\pi\epsilon^2}{16}$ and hence there are only finitely many distinct $\hat{G}$. (Note that for G containing points at distance greater than ϵ from $\partial\Omega$ we have $|\hat{G}| \geq \frac{\pi\epsilon^2}{4}$, we potentially loose a factor of 4 when G is essentially a corner of the rectangle Ω.) We can assume therefore that there is some single $\hat{G}$ such that $\Gamma_{\beta_n} \cap \hat{G} \neq \emptyset$ for all n. Let $\{C_i^n\}$ be the set of connected components of Γ_{β_n} satisfying $C_i^n \cap \hat{G} \neq \emptyset$ and let $\hat{\Gamma}_{\beta_n} = \Gamma_{\beta_n} \backslash \cup_i C_i^n$. We will denote by $\hat{\Omega}_k^n$ the connected component of $\Omega \backslash \hat{\Gamma}_{\beta_n}$ which is a superset of $\hat{G}$. Some subset of the Ω_k^n lying in $\hat{\Omega}_k^n$, whose union we denote by O_n, are the sets bounded by the C_i^n. It follows from the isoperimetric inequality that $|O_n| \leq \frac{1}{\zeta^2}(\max_i \mathcal{H}^1(C_i^n)) \sum_i \mathcal{H}^1(C_i^n)$. From lemma 4.7 we conclude

$$\lim_{n\to\infty} \max_i (\mathcal{H}^1(C_i^n)) = 0. \tag{4.2.13}$$

and now, since the sum is bounded we can conclude that for n large enough $|O_n| \leq \frac{\pi\epsilon^2}{32}$. Hence there is some $\Omega_k^n \subset \hat{\Omega}_k^n$ satisfying $|\Omega_k^n \cap A_j| \geq \frac{\pi\epsilon^2}{8}$. Let H be the function from lemma 4.6 with $\xi = \frac{\pi\epsilon^2}{32}$ and i defined by $A_j \subset R_i$. Assuming n is large enough so that $h_r(\beta_n) < \epsilon - \max_i \mathcal{H}^1(C_i^n)$ we have,

$$E_0(\hat{\Gamma}_{\beta_n}) - E_0(\Gamma) \;\leq\; \beta_n |O_n| (|f_n(\Omega_k) - a_i|^2 + \vartheta^2) - \sum_i \mathcal{H}^1(C_i^n)$$

$$\leq \left(\frac{1}{\zeta^2}\max_i \mathcal{H}^1(C_i^n)\beta_n(H^2(\beta_n) - h_\vartheta^2(\beta_n)) - 1\right)\sum_i \mathcal{H}^1(C_i^n)$$

Since the term in square brackets is negative for n sufficiently large while the sum is positive we get a contradiction of the optimality of Γ. This completes the proof of the theorem.

Lemma 4.9 For all $\epsilon > 0$, $\exists\beta' < \infty$ such that if $\beta > \beta'$ then $\mathcal{B}_g \subset [\Gamma_\beta]_\epsilon$.

Proof Suppose the lemma is false. There then exists a sequence of Γ_{β_n}, $\beta_n \uparrow \infty$, an $x \in \mathcal{B}_g$ and a $\rho > 0$ such that $B_\rho(x) \cap \Gamma_{\beta_n} = \emptyset$ for all n. We can find at least two values i_1, i_2 such that for some $\xi > 0$,

$$\min_{l=1,2} |B_\rho(x) \cap R_{i_l}| = \xi > 0$$

Let $\delta a = |a_{i_1} - a_{i_2}|$. Clearly f_n is constant in $B_\rho(x)$ and for at least one of the i_l we have $f(x) - a_{i_l} \geq \frac{\delta a}{2}$. But from this we conclude

$$\begin{aligned} E_0(\Gamma_{\beta_n}) &\geq \beta_n\left(\int_{B_\rho(x)}(g_c - f_n)^2 - \int_{B_\rho(x)}(g - g_c)^2\right) \\ &\geq \beta_n\left(\xi(\frac{\delta a}{2})^2 - Kc_b h_r(\beta_n) - |B_\rho(x)|h_\vartheta^2(\beta_n)\right) \end{aligned}$$

which contradicts the bound $E_0(\Gamma_{\beta_n}) \leq U(\beta_n) \leq c_{E_0}$ given in lemma 4.5, when β_n is sufficiently large.

Lemma 4.10 For any $\epsilon > 0$, $\exists\beta' < \infty$ such that if $\beta > \beta'$ then

$$\mathcal{H}^1(\Gamma_\beta) > \mathcal{H}^1(\mathcal{B}_g) - \epsilon.$$

Proof The symmetric difference between $\overline{\mathcal{B}_g}$ and $\cup_{i\neq j}\partial^* A_i \cap \partial^* A_j$ is an $\mathcal{H}^1$ negligible because of our assumptions and the property of essential boundaries stated in equation 2.2.1. Each $\overline{\partial^* A_i \cup \partial^* A_j}$ can be written as a countable union of rectifiable curves meeting only at their end points together with a $\mathcal{H}^1$ negligible set by theorem 2.3. Thus in general we can write $\overline{\mathcal{B}_g} = N \cup \bigcup_{i=1}^{\infty} E_i$ where N has negligible $\mathcal{H}^1$ measure and the E_i's are rectifiable curves joined only at their end points such that for each we can find A_{j_1}, A_{j_2} such that $\delta a_i = |g(A_{j_1}) - g(A_{j_2})| > 0$ and $E_i \subset \overline{\partial^* A_{j_1} \cap \partial^* A_{j_2}}$. Define $h(E_i) = \min(|A_{j_1}|, |A_{j_2}|)$.

Suppose the lemma is false. Then there exists a sequence of minimizers Γ_{β_n} such that $\mathcal{H}^1(\Gamma_{\beta_n}) \leq \mathcal{H}^1(\mathcal{B}_g) - \epsilon$. We can find an $M < \infty$ so that $\sum_{i=1}^M \mathcal{H}^1(E_i) > \mathcal{H}^1(\mathcal{B}_g) - \epsilon/2$. Let $2\xi = \min_{i \in 1,\ldots,M} h(E_i)$ and let $\tilde{\Gamma}_{\beta_n} = \Gamma_{\beta_n} \backslash \cup_i C_i^n$ where the C_i^n are the connected components of Γ_{β_n} satisfying $\mathcal{H}^1(C_i^n) \leq \xi\zeta^2/2c_E$ where c_E is the constant from lemma 4.5. By theorem 2.1 we can find a subsequence of the $\tilde{\Gamma}_{\beta_n}$ (still indexed by n) which converges in the Hausdorff metric to some $\tilde{\Gamma}_l$; we claim $\cup_{i=1}^M E_i \subset \tilde{\Gamma}_l$. Assume this is not the case. Then $\exists x \in E_i$ for some $i \leq M$ such that $\text{dist}(x, \tilde{\Gamma}_l) > 0$. We can find compact connected sets K_1, K_2 such that $K_1 \subset A_{j_1}$, $K_2 \subset A_{j_2}$ and $|K_1|, |K_2| > \xi$. From the isoperimetric inequality we conclude that the total area of the sets bounded by the C_i^n is less than or equal to $\frac{1}{\zeta^2}(\max_i \mathcal{H}^1(C_i^n)) \sum_i \mathcal{H}^1(C_i^n) \leq \xi/2$, since the sum is bounded by c_E. It follows that for n sufficiently large there is a single $\Omega_k(n)$ such that

$$|\Omega_k(n) \cap K_i| \geq \frac{\xi}{2},\ i = 1, 2$$

and hence

$$E_0(\Gamma_{\beta_n}) \geq \beta_n \int_{\Omega_k(n)} (g - f)^2 \geq \beta_n \left(\frac{\xi}{4} (\frac{\delta a_i}{2})^2 - K c_b h_r(\beta_n) - h_g^2(\beta_n) |\Omega_k(n)| \right)$$

which, for n large enough, contradicts the optimality of Γ_{β_n}. Thus $\cup_{i=1}^M E_i \subset \tilde{\Gamma}_l$ and by lemma 4.3 we have along the subsequence,

$$\liminf_{n \to \infty} \mathcal{H}^1(\tilde{\Gamma}_{\beta_n}) \geq \mathcal{H}^1(\mathcal{B}_g) - \epsilon/2$$

and hence

$$\liminf_{n \to \infty} \mathcal{H}^1(\Gamma_{\beta_n}) \geq \mathcal{H}^1(\mathcal{B}_g) - \epsilon/2$$

which gives us a contradiction. QED.

Lemma 4.11 For any $\epsilon > 0$, $\exists \beta' < \infty$ such that if $\beta > \beta'$ then,

$$\mathcal{H}^1(\Gamma_\beta) < \mathcal{H}^1(\mathcal{B}_g) + \epsilon$$

Proof This lemma is a simple consequence of lemma 4.5 which yields $\mathcal{H}^1(\Gamma_\beta) \leq \beta E_0^*(g,\beta) \leq \beta U(\beta) \leq \mathcal{H}^1(\mathcal{B}_g) + \epsilon$ for all β sufficiently large.

Proof of Theorem 4.2: Theorem 4.8 and lemma 4.9 establish $d_H(\Gamma_\beta, \overline{\mathcal{B}_g}) < \epsilon$ while lemmas 4.10 and 4.11 prove $|\mathcal{H}^1(\Gamma_\beta) - \mathcal{H}^1(\mathcal{B}_g)| < \epsilon$ for all $\beta < \beta'$ for some $\beta' < \infty$.

We remark that in the course of the proof we have shown that $|E_0^*(g,\beta) - \mathcal{H}^1(\mathcal{B}_g)| < \epsilon$ as well as $|\mathcal{H}^1(\Gamma_\beta) - \mathcal{H}^1(\mathcal{B}_g)| < \epsilon$; we conclude from this that $\beta \int_\Omega (f-g)^2 < 2\epsilon$. Since $\int_\Omega (g - g_c)^2 \leq U(\beta) < \epsilon$ (see lemma 4.5) we also have $\beta \int_\Omega (f - g_c)^2 < 6\epsilon$.

4.2.5 Weakening the Noise Constraints

An obviously relevant question concerning the limit theorem is: how tight are the estimates on the noise and smearing decay rates? In this section we show that our additive noise estimates are tight. The smearing decay rate is a little more unclear, we can relax this rate and still retain convergence of the boundaries in Hausdorff metric. What is lost is the convergence of the cost.

A Counter–example

In this section we show that our decay requirements on the additive noise are tight in the sense for any $c > 0$ and arbitrary rectangular domain we can find a piece–wise constant function g_c such that if we allow additive L^∞ noise with a norm bounded above by $c\beta^{-1}$ rather than some $h_\theta(\beta)$, as in the limit theorem, then the optimal boundaries Γ_β need not converge to the discontinuity set of g_c in hausdorff metric as β tends to infinity. Consider the function g_c as illustrated in Figure 4.3. There is some constant $a > 0$ such that the area of the region over which $g_c = 1$ is greater than a for all $\delta > 0$. We choose $\delta > 0$ small enough to satisfy $2\delta < c^2 a$.

Now we let the observed image be as in Figure 4.3 where $\eta = c\beta^{-\frac{1}{2}}$. Suppose we have a sequence of solutions to the piecewise constant variational problem, $\{\Gamma_{\beta_n}\}$ converging to the discontinuity set of g_c in Hausdorff metric with $\beta_n \to \infty$. Let f_n be the associated minimizing function for the given Γ_{β_n}. It follows that

$$\lim_{n\to\infty} \beta_n \int_\Omega (f_n - g)^2 \geq c^2 a$$

and

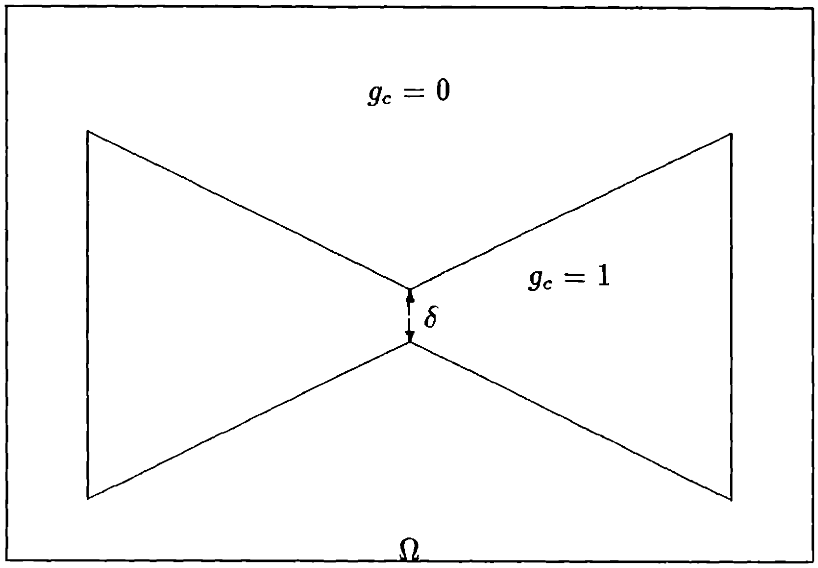
$$\liminf_{n\to\infty} \mathcal{H}^1(\Gamma_n) \geq \mathcal{H}^1(\mathcal{B}_g)$$

Let γ be the dashed line of length δ in Figure 4.3. Now consider $\Gamma' = \mathcal{B}_g \cup \gamma$, it satisfies $E_0(\Gamma') = \mathcal{H}^1(\mathcal{B}_g) + \delta$ for all n. Since we have $\delta < c^2 a$ by construction, we get a contradiction. Thus for this example the sequence $\{\Gamma_{\beta_n}\}$ cannot be a sequence of minimizers of E_0.

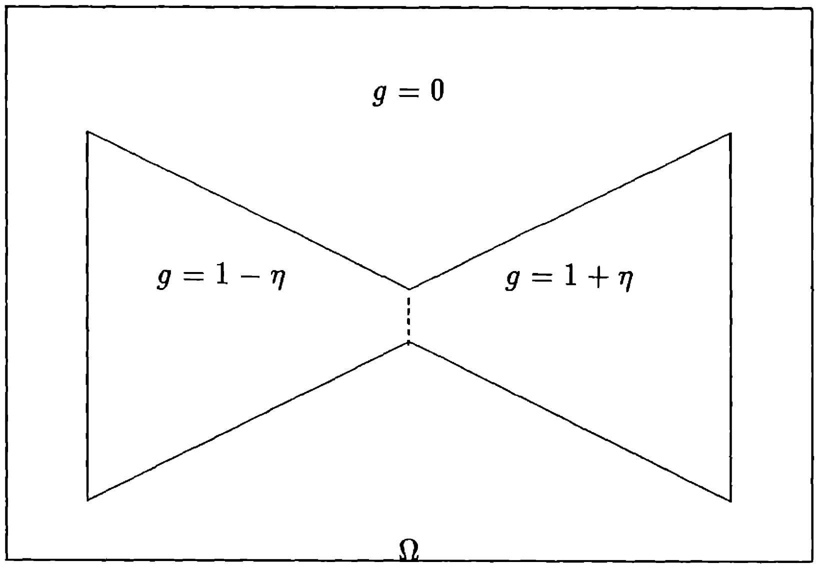
The constraint on the smearing is more than required for convergence of the boundaries in Hausdorff metric. The particular rate we have given is really only essential to make the cost converge. For convergence of the boundaries in Hausdorff metric it is sufficient that there exist a constant c such that $r \leq c\beta^{-1}$. With such a constraint it is not true in general that $\mathcal{H}^1(\Gamma_\beta) \to \mathcal{H}^1(\mathcal{B}_g)$. If we allow $\beta r \to \infty$ with $r \to 0$ and $\beta \to \infty$ then it may still be true that $\Gamma_\beta \to \mathcal{B}_g$ in Hausdorff metric. However in this case the cost can blow up so the analysis does not go through.

4.3 The Piecewise Smooth Case

The goal of this section is to prove an asymptotic fidelity result for the piecewise smooth variational formulation of the segmentation problem much as was shown in the previous section for the piecewise constant case. The plan is much the same as in the piecewise constant case although the nature of the proof is entirely different. We first state our assumptions and the theorem to be proved. Since the proof includes several detailed technical arguments we provide a sketch of the proof before actually presenting it. The proof itself is divided into two sections consisting of preliminary and main results.



The Function g_c



The Function g

Figure 4.3: The Counter Example

4.3.1 Problem Formulation

The Variational Problem

For the purposes of this section we assume that we have posed the variational problem in the SBV setting. Thus the functional whose minimizers we will be considering is,

$$E(f,\beta) = \beta \int_\Omega (g-f)^2 + \int_\Omega |\nabla f|^2 + \mathcal{H}^1(\mathcal{S}_f).$$

Theorem 3.10 provides a regularity result for minimizers of E which will play an essential in the proof of the limit theorem. In particular we recall $\mathcal{H}^1(\Omega \cap \overline{\mathcal{S}}_f \backslash \mathcal{S}_f) = 0$ and $f \in C^1(\Omega \backslash \overline{\mathcal{S}}_f)$. We have written β as an explicit parameter since we intend to vary it. Without loss of generality, we have set the parameter $\alpha = 1$.

Assumptions on the Domain

We will be assuming that our domain is a rectangle. We do this primarily to allow a reflection argument to be used to show that the boundary of the domain does not cause the introduction of spurious boundaries (see theorem 4.24.)

Assumptions on the Image

We will need some mild assumptions on the regularity of the image in order to achieve the desired result. We summarize them below.

Assumption 1 $g_u \in L^\infty(\Omega) \cap \text{SBV}(\Omega)$, $\int_\Omega |\nabla g_u|^2 + \mathcal{H}^1(\mathcal{S}_{g_u}) < \infty$ and $\mathcal{S}_{g_u}$ has no isolated points i.e. if $x \in \mathcal{S}_{g_u}$ then $\forall \rho > 0$, $\mathcal{H}^1(\mathcal{S}_{g_u} \cap B_\rho(x)) > 0$.

Assumption 2 If $A \subset \Omega$ is an open set satisfying $\text{dist}(A, \mathcal{S}_{g_u}) > 0$ 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 $|g_u(x) - g_u(y)| \le L|x-y|$. We refer to L as the Lipschitz constant associated with A.

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

The Noise Model

In the piecewise smooth case we define our noise a little less explicitly then what was done in the piecewise constant case and consequently the results are a little more general. We show below how under some additional assumptions we can replace these assumptions with those used for the piecewise constant case. As before we denote the class of images we allow by $\Upsilon(\beta)$. The following are our assumptions.

$$\lim_{\beta\to\infty} \sup_{g\in\Upsilon(\beta)} \beta \int_\Omega (g-g_u)^2 = 0, \tag{4.3.1}$$

and,

$$\forall\epsilon > 0,\ \lim_{\beta\to\infty} \sup_{g\in\Upsilon(\beta)} \|(g-g_u)(1-\chi_{[\mathcal{S}_{g_u}]_\epsilon})\|_\infty = 0 \tag{4.3.2}$$

We can make essentially the same noise assumptions that were made in the piecewise constant case if we assume that the Lipschitz constants referred to above can be uniformly bounded on $\Omega\backslash\overline{\mathcal{S}}_{g_u}$ and that the Minkowski content of $\mathcal{S}_{g_u}$ is finite.

Suppose $c_b r \geq [\mathcal{S}_{g_u}]_r$ for all $r > 0$ and the uniform Lipschitz constant for g_u on $\Omega\backslash\overline{\mathcal{S}}_{g_u}$ is L. Suppose also, as in the piecewise constant case, that g has a representation of the form,

$$g = \Phi(g_u) + \vartheta w \tag{4.3.3}$$

for some $\Phi \in \mathcal{S}_r$ and $w \in L^\infty$ with $\|w\|_\infty \leq 1$ and ϑ 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

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

Define $\Upsilon(\beta)$ to be those functions g which can be written in the form 4.3.3 for some Φ_r, w and ϑ with $r \leq h_r(\beta)$ and $\vartheta \leq h_\vartheta(\beta)$. It now follows that with this definition of $\Upsilon(\beta)$ the assumptions 4.3.1 and 4.3.2 are satisfied since

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

We can now state the limit theorem to be proved. It is essentially the same as in the piecewise constant case.

Theorem 4.12 Under our stated assumptions, as $\beta \to \infty$ $\{\mathcal{S}_f(\beta)\}$ converges to $\mathcal{S}_{g_u}$ with respect to the Hausdorff metric, and $\mathcal{H}^1(\mathcal{S}_f(\beta)) \to \mathcal{H}^1(\mathcal{S}_{g_u})$. We mean by this that for any $\epsilon > 0$ there exists $\beta' < \infty$ such that if $\beta > \beta'$ and f is a minimizer of E for some $g \in \Upsilon(\beta)$, then $d_H(\mathcal{S}_f, \mathcal{S}_{g_u}) < \epsilon$ and $|\mathcal{H}^1(\mathcal{S}_f) - \mathcal{H}^1(\mathcal{S}_{g_u})| < \epsilon$. Furthermore $\sqrt{\beta}(f - g_u)$ converges to 0 in $L^2(\Omega)$.

The proof of the theorem is quite involved so before presenting the technical arguments we provide a sketch of the main ideas. The first few results establish convergence of minimizers of E to g_u in various senses somewhat weaker than that stated by the theorem. The lemmas 4.13, 4.14 and 4.15 establish that if f_n is a sequence of minimizers of E with $g = g_n \in \Upsilon(\beta_n)$ then f_n converges to g_u in $L^1(\Omega)$, $Jf_n \to Jg_u$ weakly as radon measures and $\nabla f_n \to \nabla g_u$ weakly in $L^1(\Omega; \Re^2)$ and strongly in $L^2(\Omega; \Re^2)$. It is also shown that $\mathcal{H}^1(\mathcal{S}_{f_n}) \to \mathcal{H}^1(\mathcal{S}_{g_u})$ and that for any set $A \subset \Omega$ which is positively separated from $\mathcal{S}_{g_u}$, $\lim_{n\to+\infty} \mathcal{H}^1(\mathcal{S}_{f_n} \cap A) = 0$. These results alone are enough to assure that for β large enough $\mathcal{S}_{g_u} \subset [\mathcal{S}_{f_n}]_\epsilon$ but we defer the statement of this fact until the end of the proof. The opposite containment i.e. $[\mathcal{S}_{f_n}]_\epsilon \subset \mathcal{S}_{g_u}$ does not follow directly and it is the proving of this statement which constitutes most of the difficulty of the proof.

The results of De Giorgi, Carriero, Leaci [10] which were quoted in section 2.3 provide conditions under which one can assert for a given $x \in \Omega$ and $f \in \mathrm{SBV}(\Omega)$ that $x \notin \mathcal{S}_f$. Our goal is to show that for β sufficiently large we can for each $x \in \Omega \backslash [\mathcal{S}_{g_u}]_\epsilon$ find a ρ such that the following three conditions,

$$\begin{aligned} F(u, \overline{B}_\rho(x)) &\leq \xi\rho, \\ \lim_{t\to 0+} t^{-1}\Psi(u, \overline{B}_t(x)) &= 0, \\ \Psi(u, \overline{B}_t(x)) &\leq \eta t, \ \forall t \leq \rho, \end{aligned}$$

are satisfied.

We recall some notation,

$$F(f, \overline{B}_\rho(x)) = \int_{B_\rho(x)} |\nabla f|^2 + \mathcal{H}^1(\mathcal{S}_f \cap \overline{B}_\rho(x)).$$

and,

$$\Psi(f,\overline{B}_t(x)) = \int_{B_t(x)} |\nabla v^t|^2 + \mathcal{H}^1(\mathcal{S}_{v^t} \cap \overline{B}_t(x)) - \int_{B_t(x)} |\nabla f|^2 + \mathcal{H}^1(\mathcal{S}_f \cap \overline{B}_t(x))$$

where v^t is an extension of f into $B_t(x)$ which minimizes $\int_{B_t(x)} |\nabla v^t|^2 + \mathcal{H}^1(\mathcal{S}_{v^t} \cap \overline{B}_t(x))$. Since f minimizes E we know that,

$$\beta \int_{B_t(x)} (v^t - g)^2 + F(v^t, \overline{B}_t(x)) \geq \beta \int_{B_t(x)} (f-g)^2 + F(f, \overline{B}_t(x))$$

thus a uniform bound on $v^t - f$, h say, can provide a bound on $\Psi(u, \overline{B}_t(x))$ of the form $\beta\pi h t^2$. Since t is bounded above by ρ if we can choose ρ as a function of β and show that $h\rho$ decays faster than $\frac{1}{\beta}$ then the third condition can be met. The second condition is met rather easily and the first condition also presents not great difficulties.

The remainder of this sketch is devoted to describing how we can achieve the desired estimates on $v^t - f$. v^t can be bounded by its boundary conditions i.e. by a bound on f restricted to $\partial B_t(x)$. Thus our goal now becomes to achieve estimates on f restricted to $B_\rho(x)$ sufficient to meet the demands of the conditions on v^t. As was mentioned we intend to let ρ tend to zero as β tends to infinity. To get strong bounds on the range of f on $\partial B_\rho(x)$ it is extremely helpful to have $\overline{\mathcal{S}}_f \cap \partial B_\rho = \emptyset$. We accomplish this essentially through lemma 4.21.

In lemma 4.21 we consider small balls of radius ρ around a point $x \in K \subset \Omega \backslash \overline{\mathcal{S}}_{g_u}$ for some compact K. We let ρ be a function of β of the form $\rho = \beta^{-\gamma}$ for some positive constant γ. We use the notation $J(u,\beta,\rho,x) = \beta \int_{B_\rho(x)} u^2 + F(u, \overline{B}_\rho(x))$ and establish the result,

$$\sup_{x\in K} J(f - g_s, \beta, \beta^{-\gamma}, x) \leq c\beta^{-2\gamma}$$

where g_s is a smoothed version of g_u. There are two key observations we wish to make concerning the proof of this lemma. The first is that the proof involves redefining f in a ball $B_\rho(x)$ where $\rho \simeq \beta^{-\gamma}$ and then using the fact that f itself minimizes E to obtain estimates. In order to redefine f in a useful way we use two ideas. At points disjoint from $\overline{\mathcal{S}}_f$, f satisfies $\Delta f = \beta(f-g)$. Very roughly speaking solutions to equations of this form look like smoothed versions of g where the smoothing is done over a ball of radius $\beta^{-\frac{1}{2}}$. Now we have assumed that g tracks g_u reasonably closely so a reasonable candidate for f might be a smoothed version of g_u found by convolving g_u with a mollifier with support in a ball of radius $\beta^{-\frac{1}{2}}$. This is precisely how we define g_s. Lemma 4.20 provides some estimates on a g_s which provide us with the tools to estimate f. The redefined f is formed by continuously transforming f into g_s inside the ball B_ρ. To get good estimates on the energy associated with the new f we need to estimate f on $\partial B_\rho(x)$. Since we eventually want to compare the energy contribution in the ball of the original f with the redefined f it is clear that what is needed is some way of bounding the ratio of the contribution to the energy occurring from the boundary of the ball to that occurring from the interior of the ball. Proposition 4.18 provides us with a means of choosing the radius of the ball to guarantee that this ratio is somewhat controlled. Following this proposition we present the particular form of the redined f we have in mind and lemma 4.19 then states the relationship between the contribution to the energy incurred from the interior of the ball for the redefined f to the corresponding contribution incurred from the boundary of the ball.

The second key idea in the proof of lemma 4.21 concerns how we obtain strong estimates on $\nabla w = \nabla(f - g_s)$. We achieve this partly through noting $\nabla w = Dw - Jw$. Bounds on Dw are relatively easy to obtain using integration by parts. Thus having bounds on Jw can yield bounds on ∇w. We know that the $\mathcal{H}^1$ measure of the support set of Jw in B_ρ is tending to zero as β tends to infinity. This is insufficient because we intend to compare Jw with this measure (the details are in the proof.) The extra information which gives us a strong estimate is the result in lemma 4.17 which states that for any compact $K \subset \Omega$ disjoint from $\overline{S}_{g_u} \cup \partial\Omega$ if g_n, f_n, β_n are sequences such that $\beta_n \uparrow +\infty$ and $E(f_n, \beta_n) = E^*(\beta_n)$ with $g = g_n \in \Upsilon(\beta_n)$ then,

$$\lim_{n \to +\infty} \sup_{x \in K} |f_n^+(x) - f_n^-(x)| = 0.$$

The final stage of the proof is essentially carried out in lemma 4.23 i.e. in this lemma desired bounds on v^t are established. The remaining results tie up some loose ends such as extending the results by a reflection argument to include points x which may lie arbitrarily close to $\partial\Omega$.

4.3.2 Preliminary Results

The first few results in this section are largely consequences of the compactness and lower-semicontinuity theorem for SBV(Ω) functions due to L. Ambrosio which we quoted section 3.1.3.

Let $E^*(\beta)$ denote the minimal value of $E(f,\beta)$. By simply substituting g_u for f we get the following bound,

$$E^*(\beta) \leq \beta \int_\Omega (g - g_u)^2 + \int_\Omega |\nabla g_u|^2 + \mathcal{H}^1(\mathcal{S}_{g_u}). \tag{4.3.4}$$

Lemma 4.13 If g_n, f_n, β_n are sequences such that $\beta_n \uparrow +\infty$ and $E(f_n, \beta_n) = E^*(\beta_n)$ with $g = g_n \in \Upsilon(\beta_n)$, then,

$$\begin{aligned} f_n &\to g_u && \text{in } L^1(\Omega) \\ Jf_n &\to Jg_u && \text{weakly as radon measures} \\ \nabla f_n &\to \nabla g_u && \text{weakly in } L^1(\Omega;\Re^2). \end{aligned}$$

Proof From the triangle inequality and equation 4.3.4 we obtain, $\beta_n \int_\Omega (f_n - g_u)^2 \leq \beta_n \int_\Omega (g_n - g_u)^2 + E^*(\beta_n) \leq 2\beta \int_\Omega (g_n - g_u)^2 + \int_\Omega |\nabla g_u|^2 + \mathcal{H}^1(\mathcal{S}_{g_u})$. Since $\beta_n \to \infty$ it follows from assumption 4.3.1 that f_n converges to g_u in $L^2(\Omega)$. Noting $|\Omega| < \infty$ we conclude f_n converges to g_u in $L^1(\Omega)$. The other statements are now an obvious consequence of the SBV(Ω) compactness theorem due to L. Ambrosio, which we paraphrased in theorem 3.7. □

Lemma 4.14 If g_n, f_n, β_n are sequences such that $\beta_n \uparrow +\infty$ and $E(f_n, \beta_n) = E^*(\beta_n)$ with $g = g_n \in \Upsilon(\beta_n)$, then,

$$\begin{aligned} \lim_{n\to+\infty} \int_\Omega |\nabla f_n|^2 &= \int_\Omega |\nabla g_u|^2 \\ \lim_{n\to+\infty} \mathcal{H}^1(\mathcal{S}_{f_n}) &= \mathcal{H}^1(\mathcal{S}_{g_u}) \end{aligned}$$

Proof By 4.3.4 and assumption 4.3.1 we have

$$\limsup_{n\to\infty} \int_\Omega |\nabla f_n|^2 + \mathcal{H}^1(\mathcal{S}_{f_n}) \leq \int_\Omega |\nabla g_u|^2 + \mathcal{H}^1(\mathcal{S}_{g_u}). \tag{4.3.5}$$

A lower-semicontinuity result due to L. Ambrosio, which we paraphrased in the preceding chapter as theorem 3.8 yields,

$$\int_\Omega |\nabla g_u|^2 + \mathcal{H}^1(\mathcal{S}_{g_u}) \le \liminf_{n\to\infty} \int_\Omega |\nabla f_n|^2 + \mathcal{H}^1(\mathcal{S}_{f_n}) \tag{4.3.6}$$

An examination of the proof reveals that each term is lower-semicontinuous separately, i.e. $\int_\Omega |\nabla g_u|^2 \le \liminf_{n\to+\infty} \int_\Omega |\nabla f_n|^2$ and $\mathcal{H}^1(\mathcal{S}_{g_u}) \le \liminf_{n\to+\infty} \mathcal{H}^1(\mathcal{S}_{f_n})$. (This can also easily be seen by rescaling f_n and g_u and noting that 4.3.5 and 4.3.6 still hold.)

□

Corollary If $A \subset \Omega$ is any borel set such that $\text{dist}(A, \overline{\mathcal{S}_{g_u}} \cup \partial\Omega) > 0$ then,

$$\lim_{n\to+\infty} \mathcal{H}^1(\mathcal{S}_{f_n} \cap A) = 0.$$

Proof For some $\epsilon > 0$, $A \cap [\mathcal{S}_{g_u}]_\epsilon = \emptyset$. From lemma 4.13 we conclude $f_n \to g_u$ in $L^1([\mathcal{S}_{g_u}]_\epsilon)$. We can now apply essentially the same argument as in lemma 4.14 to conclude, $\mathcal{H}^1(\mathcal{S}_{g_u}) \le \liminf_{n\to\infty} \mathcal{H}^1(\mathcal{S}_{f_n}) \cap [\mathcal{S}_{g_u}]_\epsilon)$. But the result of lemma 4.14 states $\mathcal{H}^1(\mathcal{S}_{g_u} = \lim_{n\to\infty} \mathcal{H}^1(\mathcal{S}_{f_n})$ so it follows that $\lim_{n\to\infty} \mathcal{H}^1(\mathcal{S}_{f_n} \cap A) = 0$. □

This result apparently takes us quite close to the desired containment $\mathcal{S}_f \in [\mathcal{S}_{g_u}]_\epsilon$ for β sufficiently large. However, closing the gap, i.e. actually achieving this result is quite difficult and represents the primary accomplishment of the remainder of this section.

We append the additional notation $e = f - g_u$.

Lemma 4.15 If g_n, f_n, β_n are sequences such that $\beta_n \uparrow +\infty$ and $E(f_n, \beta_n) = E^*(\beta_n)$ with $g = g_n \in \Upsilon(\beta_n)$, then,

$$\lim_{n\to+\infty} \beta_n \int_\Omega e_n^2 + \int_\Omega |\nabla e_n|^2 = 0.$$

Proof From lemma 4.14 and inequality 4.3.4 we conclude $\lim_{n\to\infty} E(f_n, \beta_n) = \int_\Omega |\nabla g_u|^2 + \mathcal{H}^1(\mathcal{S}_{g_u})$ and hence $\lim_{n\to\infty} \beta_n \int_\Omega (f - g_n)^2 = 0$. Now,

$$\beta_n \int_\Omega (e_n)^2 \le 2(\beta_n \int_\Omega (f_n - g_n)^2 + \beta_n \int_\Omega (g_n - g_u)^2)$$

Since $\lim_{n\to\infty}\beta_n \int_\Omega (g_n - g_u)^2 = 0$ according to assumption 4.3.1 we obtain,

$$\lim_{n\to\infty}\beta_n \int_\Omega (e_n)^2 = 0.$$

Now,

$$\int_\Omega |\nabla e_n|^2 = \int_\Omega |\nabla f_n|^2 - \int_\Omega \nabla f_n \cdot \nabla g_u + \int_\Omega |\nabla g_u|^2$$

and since by lemma 4.14 we have $\lim_{n\to\infty}\int_\Omega |\nabla f_n|^2 = \int_\Omega |\nabla g_u|^2$, the desired result follows if we can show $\lim_{n\to\infty}\int_\Omega \nabla f_n \cdot \nabla g_u = \int_\Omega |\nabla g_u|^2$. We have by assumption $\int_\Omega |\nabla g_u|^2 < \infty$ so if we define $\chi_N = \{|\nabla g_u|^2 \le N\}$ then by the monotone convergence theorem we have $\lim_{N\to\infty}\int_\Omega |\nabla g_u|^2(1-\chi_N) = 0$. Thus for any $\epsilon > 0$ we can choose N sufficiently large so that $2(\int_\Omega |\nabla g_u|^2)^{\frac{1}{2}}(\int_\Omega |\nabla g_u|^2(1-\chi_N))^{\frac{1}{2}} < \epsilon$. Now,

$$\begin{aligned}|\int_\Omega \nabla f_n \cdot \nabla g_u - \int_\Omega |\nabla g_u|^2| &\le |\int_\Omega \nabla f_n \cdot \nabla g_u \chi_N - \int_\Omega |\nabla g_u|^2 \chi_n| \\ &\quad + |\int_\Omega \nabla f_n \cdot \nabla g_u(1-\chi_N)| + \int_\Omega |\nabla g_u|^2(1-\chi_N)\end{aligned}$$

Since $\nabla f_n \to \nabla g_u$ weakly in $L^1(\Omega, \Re^2)$ we have $\lim_{n\to+\infty}\int_\Omega \nabla f_n \cdot \nabla g_u \chi_N = \int_\Omega |\nabla g_u|^2 \chi_N$. Using the Schwartz inequality we obtain,

$$\limsup_{n\to\infty} |\int_\Omega \nabla f_n \cdot \nabla g_u(1-\chi_N)| \le (\int_\Omega |\nabla f_n|^2)^{\frac{1}{2}}(\int_\Omega |\nabla g_u|^2(1-\chi_N))^{\frac{1}{2}}.$$

Again we use lemma 4.14 which states that $\lim_{n\to\infty}\int_\Omega |\nabla f_n|^2 = \int_\Omega |\nabla g_u|^2$ thus,

$$\begin{aligned}\limsup_{n\to+\infty} |\int_\Omega \nabla f_n \cdot \nabla g_u - \int_\Omega |\nabla g_u|^2| &\le (\int_\Omega |\nabla g_u|^2|)^{\frac{1}{2}}(\int_\Omega |\nabla g_u|^2(1-\chi_N))^{\frac{1}{2}} + \\ &\quad + \int_\Omega |\nabla g_u|^2(1-\chi_N) \\ &\le 2(\int_\Omega |\nabla g_u|^2|)^{\frac{1}{2}}(\int_\Omega |\nabla g_u|^2(1-\chi_N))^{\frac{1}{2}} \\ &\le \epsilon\end{aligned}$$

Since ϵ is arbitrary the proof is complete. □

At several points in the results to follow it will be necessary to obtain a uniform bound on the trace of an SBV function on some circle $C \subset \Omega$. In all cases the circle will be disjoint from the closure of the jump set of the function and the function under consideration will be at least Lipschitz continuous on C. The following proposition is essentially a Sobolev inequality and will provide us with such a bound. It's proof is elementary so we include it for sake of completeness.

Proposition 4.16 If $u \in C^{0,1}(\partial B_\rho)$, $4\pi\rho \geq \beta^{-\frac{1}{2}}$ and $\beta \int_{\partial B_\rho} u^2 \, d\mathcal{H}^1 + \int_{\partial B_\rho} \dot{u}^2 \, d\mathcal{H}^1 \leq p$ then

$$\max_{y \in \partial B_\rho} |u(y)| \leq \sqrt{2p}\beta^{-\frac{1}{4}} \tag{4.3.7}$$

Proof Let $\overline{u} = \max_{y \in \partial B_\rho(x)} |w|$. Since $\beta \int_{\partial B_\rho} u^2 \, d\mathcal{H}^1 \leq p$ it follows that $\beta(\frac{\overline{u}}{2})^2 \cdot \mathcal{H}^1\{y \in \partial B_\rho(x) : |w| > \frac{\overline{u}}{2}\} \leq p$. If $|u(x)| \geq \frac{\overline{u}}{2}$ for all $x \in \partial B_\rho$ then we get $\beta(\frac{\overline{u}}{2})^2 2\pi\rho \leq p$. Assuming $4\pi\rho \geq \beta^{-\frac{1}{2}}$ we obtain 4.3.7. Now, if there is an $x \in \partial B_\rho$ such that $|u(x)| \leq \frac{\overline{u}}{2}$ then it follows that $\int_{\partial B_\rho(x)} \dot{u}^2 \geq 4(\frac{\overline{u}}{2})^4 \beta p^{-1}$ and again 4.3.7 is fulfilled. □

The next lemma will give us some control over the measure Jf. We show that the jump height of f at points of $\mathcal{S}_f$ positively separated from $\mathcal{S}_{g_u}$ (assuming they exist) must tend to zero as β tends to infinity.

Lemma 4.17 Let $K \in \Omega$ be any compact set disjoint from $\overline{\mathcal{S}}_{g_u} \cup \partial\Omega$ and let g_n, f_n, β_n be sequences such that $\beta_n \uparrow +\infty$ and $E(f_n, \beta_n) = E^*(\beta_n)$ with $g = g_n \in \Upsilon(\beta_n)$. It then follows that

$$\lim_{n \to +\infty} \sup_{x \in K} |f_n^+(x) - f_n^-(x)| = 0.$$

Proof Let $\delta = \frac{1}{2}\text{dist}(K, (\mathcal{S}_{g_u} \cup \partial\Omega))$ and let L be the Lipschitz constant associated with g_u on $[K]_\delta$. Given $\epsilon > 0$ let $\rho = \min(\frac{1}{4}L^{-1}\epsilon, \delta)$. Define $\eta_n = \sup_{x \in K}(\beta_n \int_{B_\rho(x)} e_n^2 + \int_{B_\rho(x)} |\nabla e_n|^2)$ where as before $e = f - g_u$. Lemma 4.15 asserts that $\lim_{n \to +\infty} \eta_n = 0$. This and the corollary to lemma 4.14 imply $\exists N$ such that if $n \geq N$ then the following are all satisfied,

$$\sup_{x \in K} \mathcal{H}^1(B_\rho(x) \cap \mathcal{S}_{f_n}) \leq \frac{\rho}{4} \tag{4.3.8}$$

$$\beta_n^{-\frac{1}{4}}(\frac{8\eta_n}{\rho})^{\frac{1}{2}} \leq \frac{\epsilon}{4} \tag{4.3.9}$$

$$\sup_{x \in K} \|(g - g_u)\chi_{B_{\rho(x)}}\|_\infty \leq \frac{\epsilon}{4} \tag{4.3.10}$$

Since f_n satisfies $\mathcal{H}^1(\overline{\mathcal{S}}_{f_n} \backslash \mathcal{S}_{f_n}) = 0$ (according to theorem 3.10,) we conclude that for each $x \in K$ and $n \geq N$ there exists $\rho_n(x) \in (\frac{\rho}{2}, \rho)$ such that $\overline{\mathcal{S}}_{f_n} \cap B_{\rho_n(x)}(x) = \emptyset$ and $\beta_n \int_{\partial B_{\rho_n(x)}(x)} e_n^2 + \int_{\partial B_{\rho_n(x)}(x)} |\nabla e_n|^2 \leq \frac{4}{\rho}\eta_n$. Now, since $f_n \in C^1(\Omega \backslash \overline{\mathcal{S}}_{f_n})$ and g_u is a Lipschitz function on $[K]_\delta$ we conclude e_n is a Lipschitz function on $\partial B_{\rho_n(x)}(x)$. We can now apply proposition 4.16 (assuming n is large enough so that $2\pi\rho \geq \beta_n^{-\frac{1}{2}}$) to obtain,

$$\sup_{x\in K}\max_{y\in\partial B_{\rho_n(x)}(x)} |e_n(y)| \leq \beta_n^{-\frac{1}{4}}\left(\frac{8\eta_n}{\rho}\right)^{\frac{1}{2}}$$

and from 4.3.9 we get,

$$\sup_{x\in K}\max_{y\in\partial B_{\rho_n(x)}(x)} |e_n(y)| \leq \frac{\epsilon}{4}$$

Let

$$t_n = \max\left(\sup_{y\in B_{\rho_n(x)}(x)} g(y), \max_{y\in\partial B_{\rho_n(x)}(x)} f_n(y)\right).$$

Suppose now that for some $y \in B_{\rho_n(x)}(x)\backslash\overline{S}_{f_n}$ we have $f_n(y) > t_n$. Define,

$$\tilde{f}_n = \begin{cases} f_n & x \in \Omega\backslash B_{\rho_n(x)}(x) \\ f_n \vee t_n & x \in \overline{B}_{\rho_n(x)}(x) \end{cases}$$

It follows that $|\nabla\tilde{f}_n| \leq |\nabla f_n|$ almost everywhere and $\mathcal{H}^1(\mathcal{S}_{\tilde{f}_n}) \leq \mathcal{H}^1(\mathcal{S}_{f_n})$. However since $f_n \in C^1(\Omega\backslash\overline{\mathcal{S}}_{f_n})$ it is also true that $\int_\Omega(\tilde{f}_n - g_n)^2 < \int_\Omega(f_n - g_n)^2$ which contradicts f_n being a minimizer of E. Thus it is necessarily the case that,

$$\begin{aligned}\sup_{y\in B_{\rho_n(x)}(x)} f_n(y) &\leq \max\left(\sup_{y\in B_{\rho_n(x)}(x)} g(y), \max_{y\in\partial B_{\rho_n(x)}(x)} f_n(y)\right) \\ &\leq \sup_{y\in B_{\rho_n(x)}(x)} g_u(y) + \frac{\epsilon}{4}\end{aligned}$$

and similarly,

$$\begin{aligned}\inf_{y\in B_{\rho_n(x)}(x)} f_n(y) &\geq \min\left(\inf_{y\in B_{\rho_n(x)}(x)} g(y), \min_{y\in\partial B_{\rho_n(x)}(x)} f_n(y)\right) \\ &\geq \inf_{y\in B_{\rho_n(x)}(x)} g_u(y) - \frac{\epsilon}{4}\end{aligned}$$

From this we obtain,

$$\begin{aligned}\sup_{y\in B_{\rho_n(x)}(x)}(f_n^+(y) - f_n^-(y)) &\leq \sup_{y\in B_{\rho_n(x)}(x)} g_u(y) - \inf_{y\in B_{\rho_n(x)}(x)} g_u(y) + \frac{\epsilon}{2} \\ &\leq 2Lr + \frac{\epsilon}{2} \\ &\leq \epsilon\end{aligned}$$

The lemma now follows from the arbitrariness of ϵ. □

We now introduce some further notation. For $\overline{B}_\rho(x) \in \Omega$ and $u \in \mathrm{SBV}(\Omega)$ we define,

$$J(u,\beta,\rho,x) = \beta \int_{B_\rho(x)} u^2 + \int_{B_\rho(x)} |\nabla u|^2 + \mathcal{H}^1(\overline{\mathcal{S}}_u \cap B_\rho(x))$$

and, wherever it exists,

$$J'(u,\beta,\rho,x) = \beta \int_{\partial B_\rho(x)} u^2 \, d\mathcal{H}^1 + \int_{\partial B_\rho(x)} |\nabla u|^2 \, d\mathcal{H}^1 + \mathcal{H}^0(\overline{\mathcal{S}}_u \cap \partial B_\rho(x))$$

The proposition to follow provides us with a means of determining a ρ for a given $x \in \Omega$ at which the ratio $\frac{J'(u,\beta,\rho,x)}{J(u,\beta,\rho_1,x)}$ can be bounded in some sense. This will be important when we redefine minimizers of E in the interior of a ball by extending the boundary values into the interior.

Proposition 4.18 Suppose we are given $u \in \mathrm{SBV}(\overline{B}_{\rho_2}(x))$. Let $0 < \rho_1 < \rho_2$ and assume $J(u,\beta,\rho_1,x) > 0$. Then there exists $\rho \in (\rho_1, \rho_2)$ such that

$$J'(u,\beta,\rho,x) \leq 2(\rho_2 - \rho_1)^{-1} \ln\left(\frac{J(u,\beta,\rho_2,x)}{J(u,\beta,\rho_1,x)}\right) J(u,\beta,\rho,x)$$

Proof Define

$$t = \inf_{\rho \in (\rho_1,\rho_2)} \frac{J'(u,\beta,\rho,x)}{J(u,\beta,\rho_1,x)}$$

and

$$\hat{J}(\rho) = J(u,\beta,\rho,x) + \beta \int_{\rho'}^{\rho} J'(u,\beta,r,x) \, d\mathcal{H}^1(r).$$

$\hat{J}(\rho)$ is a nondecreasing absolutely continuous function of ρ and $\frac{\partial}{\partial\rho}\hat{J}(\rho) = J'(u,\beta,\rho,x)$ for almost all $\rho \in (\rho_1,\rho_2)$. Thus for almost all $\rho \in (\rho_1,\rho_2)$ we have $\frac{\partial}{\partial\rho}\hat{J}(\rho) \geq tJ(u,\beta,\rho,x)$. Now, we will establish the relation

$$J(u,\beta,\rho,x) \geq \hat{J}(\rho) \tag{4.3.11}$$

but first we show how this implies the desired result. For almost all $\rho \in (\rho_1,\rho_2)$ we have,

$$\begin{aligned} \frac{\partial}{\partial\rho}(e^{-(\rho-\rho_1)t}\hat{J}(\rho)) &= e^{-(\rho-\rho_1)t}\frac{\partial}{\partial\rho}\hat{J}(\rho) - t(e^{-(\rho-\rho_1)t}\hat{J}(\rho)) \\ &\leq 0 \end{aligned}$$

by the definition of t. Thus we obtain $\hat{J}(\rho_2) \geq \hat{J}(\rho_1)\exp(\rho_2 - \rho_1)t$. We note $\hat{J}(\rho_1) = J(u,\beta,\rho_1,x)$ and using 4.3.11 we have $\hat{J}(\rho_2) \leq J(u,\beta,\rho_2,x)$. The lemma now follows by choosing ρ such that $\frac{J'(u,\beta,\rho,x)}{J(u,\beta,\rho_1,x)} < 2t$.

To prove 4.3.11 we first note that,

$$\begin{aligned} \beta \int_{B_\rho(x)} u^2 + \int_{B_\rho(x)} |\nabla u|^2 &= \int_{\rho_1}^{\rho} \left[\beta \int_{\partial B_\rho(x)} u^2 \, d\mathcal{H}^1 + \int_{\partial B_\rho(x)} |\nabla u|^2 \, d\mathcal{H}^1 \right] d\mathcal{H}^1 \\ &\quad + \beta \int_{B_{\rho_1}(x)} u^2 + \int_{B_{\rho_1}(x)} |\nabla u|^2 \end{aligned}$$

by the Tonelli-Fubini theorem. Thus all we need establish is,

$$\mathcal{H}^1(\overline{\mathcal{S}}_u \cap (B_\rho(x) \backslash B_{\rho_1}(x))) \leq \int_{\rho_1}^{\rho} \mathcal{H}^0(\overline{\mathcal{S}}_u \cap \partial B_r(x)) dr \tag{4.3.12}$$

To simplify the notation we will assume x is the origin, and denote $\overline{\mathcal{S}}_u \cap (B_\rho \backslash B_{\rho_1})$ by Γ. For any $\epsilon, \delta > 0$ we can find collection of sets $\{U_i\}$ such that $\Gamma \subset \cup_i U_i$, $\text{diam}(U_i) < \delta$ and $\mathcal{H}^1_\delta(\Gamma) + \epsilon \geq \sum_i \text{diam}(U_i)$. Let $\chi_{i,r}$ be the indicator function of the condition $U_i \cap \partial B_r \neq \emptyset$. It follows by definition that $\mathcal{H}^0_\delta(\Gamma \cap \partial B_r) \leq \sum_i \chi_{i,r}$. Now $\int_{\rho_1}^{\rho} \chi_{i,r} \, dr \leq \text{diam}(U_i)$ hence,

$$\begin{aligned} \int_{\rho_1}^{\rho} \mathcal{H}^0_\delta(\Gamma \cap \partial B_r) &\leq \sum_i \text{diam}(U_i) \\ &\leq \mathcal{H}^1_\delta(\Gamma) + \epsilon \\ &\leq \mathcal{H}^1(\Gamma) + \epsilon \end{aligned}$$

Since ϵ is arbitrary we have in fact,

$$\int_{\rho_1}^{\rho} \mathcal{H}^0_\delta(\Gamma \cap \partial B_r) \leq \mathcal{H}^1(\Gamma)$$

Now, consider any sequence $\delta_n \downarrow 0$. For each r the sequence, $\mathcal{H}^0_{\delta_n}(\Gamma \cap \partial B_r)$ is monotonically increasing to $\mathcal{H}^0(\Gamma \cap \partial B_r)$ thus by the monotone convergence theorem we have,

$$\lim_{n \to \infty} \int_{\rho_1}^{\rho} \mathcal{H}^0_{\delta_n}(\Gamma \cap \partial B_r) \, dr = \int_{\rho_1}^{\rho} \mathcal{H}^0(\Gamma \cap \partial B_r) \, dr$$

which completes the proof. □

In order to get some bounds on the contribution to E occurring in certain subsets of Ω we will redefine f in various balls in Ω. To facilitate this we will introduce some more notation.

Suppose $u \in \mathrm{SBV}(\Omega)$ and $\overline{B_\rho}(x) \subset \Omega$. We will introduce polar coordinates r, θ centered at x. For $0 < \rho' < \rho$ we define,

$$\begin{aligned} \Phi(\rho,\rho',r,\theta) &= 1 \wedge (\frac{r-\rho'}{\rho-\rho'} \vee 0) \\ \hat{u}(\rho,r,\theta) &= \begin{cases} u(r,\theta) & (r,\theta) \in \Omega \backslash \overline{B_\rho} \\ u(\rho,\theta) & \text{otherwise} \end{cases} \\ \tilde{u}(\rho,\rho',r,\theta) &= \Phi(\rho,\rho',r,\theta)\hat{u}(\rho,r,\theta) \end{aligned} \tag{4.3.13}$$

Figure 4.4 illustrates this definition.

Lemma 4.19 Let $B_\rho(x) \subset\subset \Omega$ and let $u \in \mathrm{SBV}(\Omega)$ satisfy $u \in C^{0,1}(\Omega \backslash \overline{\mathcal{S}}_u)$. Then with $\tilde{u}$ defined as above we have,

$$J(\tilde{u},\beta,\rho,x) \leq \left([\beta(\rho-\rho')]^{-1} + (\rho-\rho')\right) J'(u,\beta,\rho,x)$$

Proof The following inequality is easily derived,

$$\int_{B_\rho(x)} \Phi^2 \hat{u}^2 \leq \frac{1}{2}(\rho-\rho') \int_{\partial B_\rho(x)} u^2 \, d\mathcal{H}^1$$

Note that $\nabla \hat{u} \cdot \nabla \Phi = 0$ so,

$$\int_{B_\rho(x)} |\nabla(\Phi\hat{u})|^2 = \int_{B_\rho(x)} \Phi^2 |\nabla \hat{u}|^2 + \int_{B_\rho(x)} |\nabla\Phi|^2 \hat{u}^2$$

Some straightforward algebra now verifies,

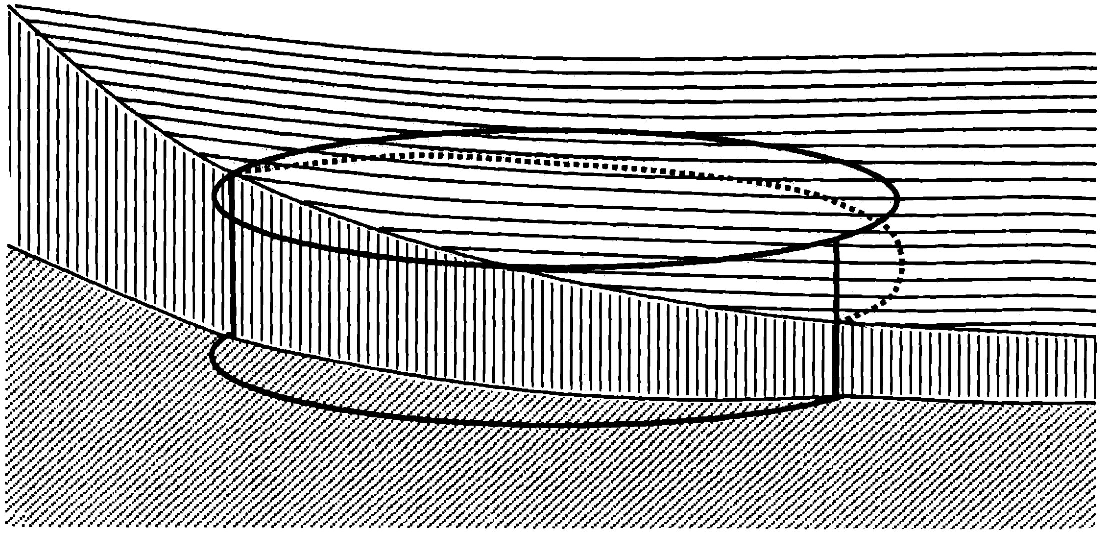
$$\int_{B_\rho(x)} |\nabla(\Phi\hat{u})|^2 \leq \frac{1}{2}(\rho-\rho') \int_{\partial B_\rho(x)} |\nabla u|^2 \, d\mathcal{H}^1 + \frac{1}{\rho-\rho'} \int_{\partial B_\rho(x)} u^2 \, d\mathcal{H}^1$$

Finally, because of the regularity assumption on u it follows that if u_1 is the restriction of u to $\partial B_\rho(x)$ then as a member of $\mathrm{SBV}(\partial B_\rho(x))$ the function u_1 satisfies $\mathcal{S}_{u_1} \subset \overline{\mathcal{S}}_u$. Thus we obtain,

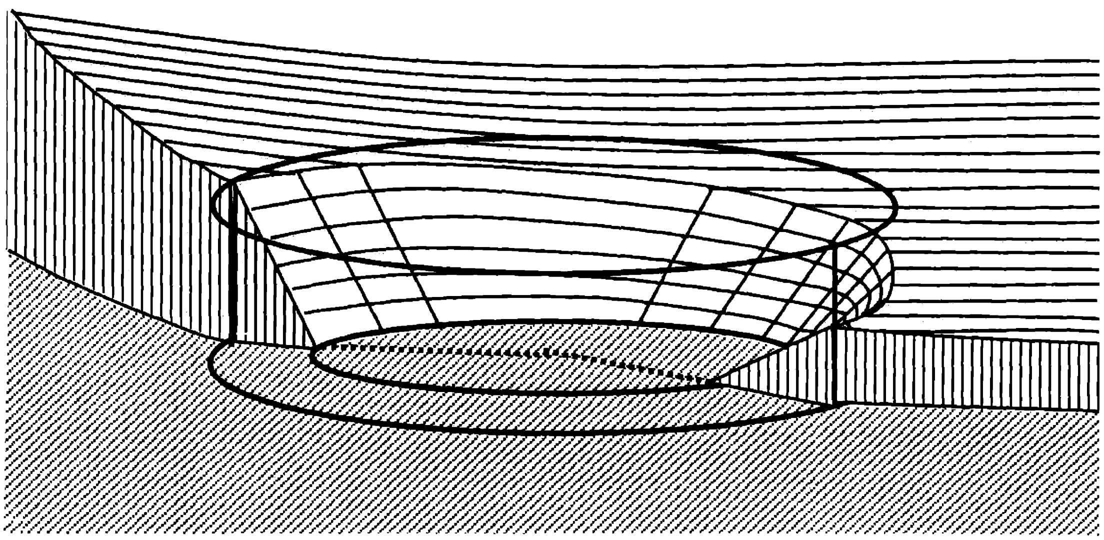
$$\mathcal{H}^1(\mathcal{S}_{\tilde{u}} \cap B_\rho(x)) \leq (\rho-\rho')\mathcal{H}^0(\partial B_\rho(x) \cap \overline{\mathcal{S}}_u)$$

Together these inequalities constitute the proof the lemma. □

We will construct smoothed versions of g_u. Wherever $x \notin \overline{\mathcal{S}}_f$ we have $\Delta f = \beta(f-g)$ thus, roughly speaking, f is a smoothed version of g where the support of



u, and the ball B_ρ



$\hat{u}$, and the balls B_ρ and B_{ρ_1}

Figure 4.4: Illustration of Construction of $\hat{u}$ from u

the smoothing occurs over a region of radius $\beta^{-\frac{1}{2}}$. We will compare the optimal f to such a smoothed version of g_u. However, it is more convenient to use a mollifier to do the smoothing than to consider the solution to a p.d.e. Let $\eta \in C_0^\infty(B_1)$ be a positive, symmetric function satisfying $\int_{B_1} \eta = 1$. If $x \in \Omega$ and $0 < \beta^{-\frac{1}{2}} < \text{dist}(x, \partial\Omega)$ then we can define,

$$g_s(x) = \int_{B_{\beta^{-\frac{1}{2}}}(x)} \beta\, \eta(\beta^{\frac{1}{2}}(x-y)) g_u(y)\, dy$$

By definition η is uniformly continuous. We will denote the modulus of continuity of η by c_η; i.e. $|\eta(x) - \eta(y)| \le c_\eta |x-y|$ for all $x, y \in \Re^2$.

Lemma 4.20 Let $g_u \in \text{SBV}(\Omega)$ satisfy our assumptions (assumption 2 in particular,) and $K \subset \Omega$ be a compact set such that $K \cap \overline{S}_{g_u} = \emptyset$. Now define $\delta = \frac{1}{2}\text{dist}(K, S_{g_u} \cup \partial\Omega)$ and denote by L the Lipschitz constant associated with $[K]_\delta$. If $\beta^{-\frac{1}{2}} < \delta$ then the following estimates hold;

$$\begin{aligned}
\sup_{y \in K} |g_s(y) - g_u(y)| &\le L\beta^{-\frac{1}{2}} \\
\sup_{y \in K} |\nabla g_s(y)| &\le L \\
\sup_{y \in K} |\Delta g_s(y)| &\le \sqrt{2}\pi c_\eta L \beta^{\frac{1}{2}}
\end{aligned}$$

Proof Let $x \in K$ then,

$$\begin{aligned}
|g_s(x) - g_u(x)| &\le \beta \Big| \int_{B_{\beta^{-\frac{1}{2}}}(x)} \eta(\beta^{\frac{1}{2}}(x-y)) g_u(y) dy - g_u(x) \Big| \\
&\le \beta \Big| \int_{B_{\beta^{-\frac{1}{2}}}(x)} \eta(\beta^{\frac{1}{2}}(x-y))(g_u(y) - g_u(x)) dy \Big| \\
&\le \beta \Big| \int_{B_{\beta^{-\frac{1}{2}}}(x)} \eta(\beta^{\frac{1}{2}}(x-y)) L \beta^{-\frac{1}{2}} \Big| \\
&\le L\beta^{-\frac{1}{2}}
\end{aligned}$$

proving the first statement.

Since g_u is Lipschitz in $[K]_\delta$, ∇g_u exists almost everywhere (in the strong sense) on $[K]_\delta$ and satisfies $|\nabla g_u| \le L$. Now,

$$\nabla g_s(x) = \beta \int_{B_{\beta^{-\frac{1}{2}}}(x)} \nabla_x \eta(\beta^{\frac{1}{2}}(x-y))\, g_u(y) dy$$

$$= \beta \int_{B_{\beta^{-\frac{1}{2}}}(x)} \eta(\beta^{\frac{1}{2}}(x-y)) \, \nabla_y g_u(y) dy$$

and the second statement follows.

Let e_1, e_2 be the standard basis for $\Re^2$.

$$\begin{aligned}
& |\frac{\partial}{\partial x_1} g_s(x+\epsilon e_1) - \frac{\partial}{\partial x_1} g_s(x)| \\
= \; & \beta \int_{B_{\beta^{-\frac{1}{2}}}(x)} \left(\eta(\beta^{\frac{1}{2}}(x+\epsilon e_1 - y)) - \eta(\beta^{\frac{1}{2}}(x-y))\right) \frac{\partial}{\partial y_1} g_u(y) dy \\
\leq \; & \beta^{\frac{3}{2}} c_\eta \epsilon \int_{B_{\beta^{-\frac{1}{2}}}(x)} \frac{\partial}{\partial y_1} g_u(y) dy
\end{aligned}$$

Thus we get

$$\begin{aligned}
|\Delta g_s(x)| \; & \leq \; \beta^{\frac{3}{2}} c_\eta \int_{B_{\beta^{-\frac{1}{2}}}(x)} |\frac{\partial}{\partial y_1} g_u(y)| + |\frac{\partial}{\partial y_2} g_u(y)| dy \\
& \leq \; \sqrt{2}\pi \beta^{\frac{1}{2}} c_\eta L
\end{aligned}$$

This completes the proof of the lemma. □

4.3.3 Main Results

We are now ready to establish the most important estimate in the proof. One important consequence of the following lemma is that it shows that if there remains small pieces of the boundary $\mathcal{S}_f$ disjoint from $[\mathcal{S}_{g_u}]_\epsilon$ then they are sparsely placed.

Lemma 4.21 Let $g_u \in \mathrm{SBV}(\Omega)$ satisfy our assumptions and $K \subset \Omega$ be a compact set such that $K \cap \overline{\mathcal{S}}_{g_u} = \emptyset$. Define $\delta = \frac{1}{2}\mathrm{dist}(K, \mathcal{S}_{g_u} \cup \partial\Omega)$, denote by L the Lipschitz constant associated with $[K]_\delta$ and set $c = \pi(8(1 + L(1 + \sqrt{2}\pi c_\eta)))^2$. Now, given $0 < \gamma < \frac{1}{2}$ there exists a constant $\beta' < \infty$ such that if $\beta \geq \beta'$ and f minimizes $E(\beta)$ for some $g \in \Upsilon(\beta)$ then,

$$\sup_{x \in K} J(f - g_s, \beta, \beta^{-\gamma}, x) \leq c\beta^{-2\gamma}$$

Proof Assume the lemma is false. There exist a K and γ satisfying the conditions of the lemma and a sequence of quadruples $\{(g_n, f_n, \beta_n, x_n)\}$ such that $\beta_n \uparrow +\infty$, $E(f_n, \beta_n) = E^*(\beta_n)$ with $g = g_n \in \Upsilon(\beta_n)$, $x_n \in K$ and,

$$J(w_n, \beta_n, \beta_n^{-\gamma}, x_n) > c\beta_n^{-2\gamma} \tag{4.3.14}$$

for each n, where we have used the notation $w_n = f_n - g_s$. Note that since β depends on n so does g_s.

Without loss of generality we can assume that $\beta_n \geq 1$ and $2\beta_n^{-\gamma} + \beta_n^{-\frac{1}{2}} < \delta$, so that the following estimates,

$$\sup_{y \in [K]_{2\beta_n^{-\gamma}}} |g_s(y) - g_u(y)| \leq L\beta_n^{-\frac{1}{2}} \tag{4.3.15}$$

$$\sup_{y \in [K]_{2\beta_n^{-\gamma}}} |\nabla g_s(y)| \leq L \tag{4.3.16}$$

$$\sup_{y \in [K]_{2\beta_n^{-\gamma}}} |\Delta g_s(y)| \leq \sqrt{2}\pi c_\eta L\beta_n^{\frac{1}{2}} \tag{4.3.17}$$

hold by lemma 4.20. Defining $e = f - g_u$ as before and applying the estimates given above along with the triangle inequality we obtain,

$$\begin{aligned} J(w_n, \beta_n, 2\beta_n^{-\gamma}, x_n) &\leq 2\left\{\beta \int_{B_{2\beta_n^{-\gamma}}(x_n)} e_n^2 + \int_{B_{2\beta_n^{-\gamma}}(x_n)} |\nabla e_n|^2 \right. \\ &\quad \left. + \beta \int_{B_{2\beta_n^{-\gamma}}(x_n)} (g_u - g_s)^2 + \int_{B_{2\beta_n^{-\gamma}}(x_n)} |\nabla(g_u - g_s)|^2 \right\} \\ &\quad + \mathcal{H}^1(S_{f_n} \cap B_{2\beta_n^{-\gamma}}(x_n)) \\ &\leq 2(\beta \int_{B_{2\beta_n^{-\gamma}}(x_n)} e_n^2 + \int_{B_{2\beta_n^{-\gamma}}(x_n)} |\nabla e_n|^2) + 24L^2\pi\beta_n^{-2\gamma} \\ &\quad + \mathcal{H}^1(S_{f_n} \cap B_{2\beta_n^{-\gamma}}(x_n)) \end{aligned}$$

By lemma 4.15 and the corollary to lemma 4.14 we can assert that for n sufficiently large

$$J(w_n, \beta_n, 2\beta_n^{-\gamma}, x_n) \leq 1. \tag{4.3.18}$$

In this case we can conclude from proposition 4.18 that for each n we can choose $\rho_n \in (\beta_n^{-\gamma}, 2\beta_n^{-\gamma})$ such that,

$$J'(w_n,\beta_n,\rho_n,x_n) \leq 2\beta_n^{\gamma}\ln\frac{\beta_n^{2\gamma}}{c}J(w_n,\beta_n,\rho_n,x_n)$$

Let N_1 be such that if $n \geq N_1$ then 4.3.18 holds and also $\beta_n^{\gamma}\ln\frac{\beta_n^{2\gamma}}{c} < \frac{\beta_n^{\frac{1}{2}}}{32}$. We now have that for $n \geq N_1$,

$$J'(w_n,\beta_n,\rho_n,x_n) \leq \beta_n^{\frac{1}{2}}\frac{1}{16}J(w_n,\beta_n,\rho_n,x_n) \tag{4.3.19}$$

Let us define $\tilde{w}_n$ as in 4.3.13 with $\rho_n' = \rho_n - \beta_n^{-\frac{1}{2}}$ and the balls centered at x_n; i.e. we introduce polar coordinates r, θ centered at x_n and set,

$$\begin{aligned}
\Phi(\rho_n,\rho_n',r,\theta) &= 1 \wedge (\frac{r-\rho_n'}{\rho_n-\rho_n'} \vee 0) \\
\hat{w}_n(\rho_n,r,\theta) &= \begin{cases} w_n(r,\theta) & (r,\theta) \in \Omega\backslash\overline{B}_{\rho_n}(x_n) \\ w_n(\rho_n,\theta) & \text{otherwise} \end{cases} \\
\tilde{w}_n(\rho_n,\rho_n',r,\theta) &= \Phi(\rho_n,\rho_n',r,\theta)\hat{w}_n(\rho_n,r,\theta).
\end{aligned}$$

From lemma 4.19 we obtain,

$$J(\tilde{w}_n,\beta_n,\rho_n,x_n) \leq 2\beta_n^{-\frac{1}{2}}J'(w_n,\beta_n,\rho_n,x_n)$$

Now applying 4.3.19 we derive,

$$J(\tilde{w}_n,\beta_n,\rho_n,x_n) \leq \frac{1}{8}J(w_n,\beta_n,\rho_n,x_n) \tag{4.3.20}$$

Let $\tilde{f}_n = g_s + \tilde{w}_n$. Note that in $\Omega\backslash B_{\rho_n}(x_n)$ we have $\tilde{f}_n = f_n$. Since f_n is a minimizer of $E(f,\beta_n)$ we have, $E(f_n,\beta_n) \leq E(\tilde{f}_n,\beta_n)$. We can express this in terms of w_n as,

$$\begin{aligned}
J(w_n,\beta_n,\rho_n,x_n) \leq\ & J(\tilde{w}_n,\beta_n,\rho_n,x_n) + 2\beta_n\int_{B_{\rho_n}(x_n)}(\tilde{w}_n - w_n)(g_s - g) \\
&+2\int_{B_{\rho_n}(x_n)}(\nabla\tilde{w}_n - \nabla w_n)\cdot\nabla g_s.
\end{aligned}$$

Substituting from 4.3.20 we get,

$$\begin{aligned}
\frac{7}{16}J(w_n,\beta_n,\rho_n,x_n) \leq\ & \beta_n\int_{B_{\rho_n}(x_n)}(\tilde{w}_n - w_n)(g_s - g) + \\
&+\int_{B_{\rho_n}(x_n)}(\nabla\tilde{w}_n - \nabla w_n)\cdot\nabla g_s
\end{aligned} \tag{4.3.21}$$

Note that 4.3.19 and 4.3.18 imply that $\mathcal{S}_{w_n - \tilde{w}_n} \cap \partial B_{\rho_n}(x_n)$ has at most finitely many points. Thus we can apply the general result for SBV functions, equation 2.3.5, to get,

$$\begin{aligned}
\int_{B_{\rho_n}(x_n)} (\nabla w_n - \nabla \tilde{w}_n) \cdot \nabla g_s &= \int_{B_{\rho_n}(x_n)} D(w_n - \tilde{w}_n) \cdot \nabla g_s \\
&\quad - \int_{B_{\rho_n}(x_n)} J(w_n - \tilde{w}_n) \cdot \nabla g_s \\
&= \int_{B_{\rho_n}(x_n)} (w_n - \tilde{w}_n) \Delta g_s \qquad (4.3.22) \\
&\quad - \int_{B_{\rho_n}(x_n)} J(w_n - \tilde{w}_n) \cdot \nabla g_s.
\end{aligned}$$

where have used the notation,

$$\int_{B_\rho} Ju \cdot \phi = \int_{B_\rho \cap \mathcal{S}_u} (u^+ - u^-) \phi \nu_u \, d\mathcal{H}^1.$$

It is clear that $\sup_{x \in B_{\rho_n}(x_n)} (\tilde{w}_n^+(x) - \tilde{w}_n^-(x)) \leq \sup_{x \in B_{\rho_n}(x_n)} (f_n^+(x) - f_n^-(x))$. From lemma 4.17 we conclude that there exists N_2 sufficiently large so that if $n \geq N_2$ then $\sup_{x \in B_{\rho_n}(x_n)} (f_n^+(x) - f_n^-(x)) \leq \frac{1}{8L}$. We recall that $|\nabla g_s| \leq L$ for all $x \in [K]_{\rho_n}$ so we now have,

$$\begin{aligned}
\left| \int_{B_{\rho_n}(x_n)} Jw_n \cdot \nabla g_s \right| &\leq \frac{1}{8} \mathcal{H}^1(\mathcal{S}_{f_n} \cap B_{\rho_n}(x_n)) \\
&\leq \frac{1}{8} J(w_n, \beta_n, \rho_n, x_n) \qquad (4.3.23)
\end{aligned}$$

and

$$\begin{aligned}
\left| \int_{B_{\rho_n}(x_n)} J\tilde{w}_n \cdot \nabla g_s \right| &\leq \frac{1}{8} \mathcal{H}^1(\mathcal{S}_{\tilde{w}_n} \cap B_{\rho_n}(x_n)) \\
&\leq \frac{1}{8} J(\tilde{w}_n, \beta_n, \rho_n, x_n) \qquad (4.3.24)
\end{aligned}$$

Let $N = \max(N_1, N_2)$ and assume henceforth that $n \geq N$. Summing 4.3.24 with 4.3.23 and substituting from 4.3.20 we obtain,

$$\left| \int_{B_{\rho_n}(x_n)} J(w_n - \tilde{w}_n) \cdot \nabla g_s \right| \leq \frac{9}{64} J(w_n, \beta_n, B_{\rho_n}, x_n).$$

Substituting this into 4.3.22 and in turn substituting the result into 4.3.21 we obtain,

$$\frac{47}{128} J(w_n, \beta_n, \rho_n, x_n) \leq \int_{B_{\rho_n}(x_n)} (\tilde{w}_n - w_n)(\beta_n(g_s - g) + \Delta g_s) \qquad (4.3.25)$$

Since,

$$\lim_{n\to\infty} \beta_n^{\frac{1}{2}} \sup_{y\in[K]_{2\beta_n^{-\gamma}}} |g_u(y) - g(y)| = 0$$

we can assume that N is sufficiently large so that for $n \geq N$ $\sup_{y\in[K]_{2\beta_n^{-\gamma}}} |g_u(y) - g(y)| \leq \beta_n^{-\frac{1}{2}}$. From this and the estimates 4.3.15 and 4.3.17 we now have,

$$\sup_{x\in B_{\rho_n}(x_n)} |\beta_n(g_s - g) + \Delta g_s| \leq \beta_n^{\frac{1}{2}} \frac{1}{8}\sqrt{\frac{c}{\pi}} \tag{4.3.26}$$

Using the Schwartz inequality we can derive,

$$\begin{aligned} \int_{B_{\rho_n}(x_n)} |w_n| &\leq \beta_n^{-\frac{1}{2}} \sqrt{\pi\rho_n^2} (\beta_n \int_{B_{\rho_n}(x_n)} w_n^2)^{\frac{1}{2}} \\ &\leq \beta_n^{-\frac{1}{2}} \sqrt{\pi}\rho_n (J(w_n, \beta_n, \rho_n, x_n))^{\frac{1}{2}} \end{aligned} \tag{4.3.27}$$

Similarly,

$$\begin{aligned} \int_{B_{\rho_n}(x_n)} |\tilde{w}_n| &\leq \beta_n^{-\frac{1}{2}} \int_{\partial B_{\rho_n}(x_n)} |\tilde{w}_n| \, d\mathcal{H}^1 \\ &\leq \beta_n^{-1} \sqrt{2\pi\rho_n} (\beta_n \int_{\partial B_{\rho_n}(x_n)} w_n^2 \, d\mathcal{H}^1)^{\frac{1}{2}} \\ &\leq \beta_n^{-1} \sqrt{2\pi\rho_n} (J'(w_n, \beta_n, \rho_n, x_n))^{\frac{1}{2}} \\ &\leq \beta_n^{-\frac{3}{4}} \sqrt{2\pi\rho_n} (\frac{1}{16} J(w_n, \beta_n, \rho_n, x_n))^{\frac{1}{2}} \end{aligned} \tag{4.3.28}$$

where in the last step we have used 4.3.19. Combining 4.3.26, 4.3.27 and 4.3.28 and substituting into equation 4.3.25 we obtain,

$$\frac{47}{128} J(w_n, \beta_n, \rho_n, x_n) \leq \frac{1}{8}\sqrt{c}\rho_n (1 + \frac{\sqrt{2}}{4} \frac{\beta_n^{-\frac{1}{4}}}{\sqrt{\rho_n}}) (J(w_n, \beta_n, \rho_n, x_n))^{\frac{1}{2}}$$

Now, since $\rho_n \geq \beta_n^{-\gamma}$, $\gamma < \frac{1}{2}$ and $\beta_n > 1$ we have $(1 + \frac{\sqrt{2}}{4} \frac{\beta_n^{-\frac{1}{4}}}{\sqrt{\rho_n}}) < \frac{47}{32}$ and hence we obtain,

$$\frac{1}{4} J(w_n, \beta_n, \rho_n, x_n) < \frac{1}{8}\sqrt{c}\rho_n (J(w_n, \beta_n, \rho_n, x_n))^{\frac{1}{2}}$$

Noting $\rho_n \leq 2\beta_n^{-\gamma}$ we conclude,

$$J(w_n, \beta_n, \rho_n, x_n) < c\beta_n^{-2\gamma}$$

which contradicts 4.3.14. Q.E.D. □

We are now ready to demonstrate that the conditions required to prove $x \notin \mathcal{S}_f$ can be met simultaneously for each $x \in K \subset \Omega \backslash \overline{\mathcal{S}}_{g_u}$ when β is sufficiently large. We first recall some notation and some important results on SBV functions quoted in chapter 2. Let $u \in \mathrm{SBV}(\Omega)$. For every compact set $K \subset \Omega$ we set,

$$F(u,K) = \int_K |\nabla u|^2 + \mathcal{H}^1(\mathcal{S}_u \cap K)$$

and

$$\Phi(u,K) = \inf\{F(v,K) : v \in \mathrm{SBV}(\Omega), v = u \in \Omega \backslash K\}.$$

The deviation from minimality is defined as

$$\Psi(u,K) = F(u,K) - \Phi(u,K).$$

Lemma 4.22 There exist universal constants $\xi, \eta > 0$ such that if $u \in \mathrm{SBV}(\Omega)$, $B_\rho(x) \subset\subset \Omega$ for some $\rho > 0$, and each of the following three conditions hold;

$$F(u, \overline{B}_\rho(x)) \leq \xi\rho, \tag{4.3.29}$$

$$\lim_{t \to 0+} t^{-1}\Psi(u, \overline{B}_t(x)) = 0, \tag{4.3.30}$$

$$\Psi(u, \overline{B}_t(x)) \leq \eta t, \ \forall t \leq \rho, \tag{4.3.31}$$

then $x \notin \mathcal{S}_u$.

Proof This is just a combination of theorems 2.7 and 2.8 □

Lemma 4.23 Let $g_u \in \mathrm{SBV}(\Omega)$ satisfy our assumptions and $K \subset \Omega$ be a compact set such that $K \cap \overline{\mathcal{S}}_g = \emptyset$. There exists a constant $\beta' < \infty$ such that if $\beta \geq \beta'$ and f is a minimizer of $E(\cdot, \beta)$ with $g \in \Upsilon(\beta)$ then,

$$\mathcal{S}_f \cap K = \emptyset$$

Proof Assume the lemma is false. Then there exists a K satisfying the conditions of the lemma and a sequence of quadruples $\{(g_n, f_n, \beta_n, x_n)\}$ such that $\beta_n \uparrow +\infty$, $E(f_n, \beta_n) = E^*(\beta_n)$ with $g = g_n \in \Upsilon(\beta)$ and $x_n \in K \cap \mathcal{S}_{f_n}$. Define $\delta = \frac{1}{2}\mathrm{dist}(K, \mathcal{S}_{g_u} \cup$

$\partial\Omega$), denote by L the Lipschitz constant associated with $[K]_\delta$ and define c as in lemma 4.21. Fix any real γ satisfying $\frac{1}{4} < \gamma < \frac{1}{2}$. By lemma 4.21 we can assume,

$$J(w_n, \beta_n, \beta_n^{-\gamma}, x_n) < c\beta_n^{-2\gamma} \tag{4.3.32}$$

for each n, where we have again used the notation $w_n = f_n - g_s$. Furthermore, for convenience we make the assumption $c\beta_n^{-\gamma} < \frac{1}{4}$. From 4.3.32 we have $\mathcal{H}^1(\mathcal{S}_{f_n} \cap \beta_n^{-\gamma}) < c\beta_n^{-2\gamma}$. Thus $|\{\rho \in [\frac{1}{2}\beta_n^{-\gamma}, \beta_n^{-\gamma}] : \overline{\mathcal{S}}_f \cap \partial B_\rho(x_n) = \emptyset\}| > \frac{1}{2}\beta_n^{-\gamma} - c\beta_n^{-2\gamma} > \frac{1}{4}\beta_n^{-\gamma}$ by our assumption. We can thereby choose a $\rho_n \in [\frac{1}{2}\beta_n^{-\gamma}, \beta_n^{-\gamma}]$ such that $\overline{\mathcal{S}}_{f_n} \cap \partial B_{\rho_n}(x_n) = \emptyset$ and,

$$\beta \int_{\partial B_{\rho_n}(x_n)} w_n^2 \, d\mathcal{H}^1 + \int_{\partial B_{\rho_n}(x_n)} |\nabla w_n|^2 \, d\mathcal{H}^1 < 4c\beta_n^{-\gamma}. \tag{4.3.33}$$

Define,

$$\overline{w}_n = \sup_{x \in \partial B_{\rho_n}(x_n)} |w_n(x)|$$

and,

$$\overline{g - g_s} = \sup_{x \in \partial B_{\rho_n}(x_n)} |g(x) - g_s(x)|$$

From the existence and regularity results for minimizers of $E(\cdot,\cdot)$ we deduce w_n is C^1 on $\partial B_{\rho_n}(x_n)$ so from proposition 4.16 and 4.3.33 we conclude

$$\max_{x \in \partial B_{\rho_n}(x_n)} |w_n(x)| \le \sqrt{8c}\beta_n^{-\frac{1}{4}-\frac{\gamma}{2}}.$$

Our goal in the remainder of the proof is to show that the three conditions of lemma 4.22 are satisfied for n sufficiently large with $u = f$, $\rho = \rho_n$ and $x = x_n$, thus obtaining a contradiction with $x_n \in \mathcal{S}_{f_n}$. Now,

$$\begin{aligned} F(f_n, \overline{B}_{\rho_n}(x_n)) &= \int_{B_{\rho_n}(x_n)} |\nabla f_n|^2 + \mathcal{H}^1(\mathcal{S}_{f_n} \cap B_{\rho_n}(x_n)) \\ &\le 2(\int_{B_{\rho_n}(x_n)} |\nabla w_n|^2 + |\nabla g_s|^2) + \mathcal{H}^1(\mathcal{S}_{f_n} \cap B_{\rho_n}(x_n) \\ &\le 2(c + 4\pi L^2)\beta_n^{-2\gamma} \end{aligned}$$

where we have used the facts $|\nabla g_s| \le L$ and $\rho_n \le \beta^{-\gamma}$. Condition 4.3.29 is thus satisfied as long as $2(c + 4\pi L^2)\beta_n^{-\gamma} \le \xi$ which is clearly true for n sufficiently large.

Consider a fixed n, let $0 < t \leq \rho_n$ and let $v^t \in \mathrm{SBV}(\Omega)$ realize $\Phi(f_n, \overline{B}_t(x_n))$, i.e. $v^t(x) = f_n(x)$ for all $x \in \Omega \backslash \overline{B}_t(x_n)$ and $F(v^t, \overline{B}_t(x_n)) = \Phi(f_n, \overline{B}_t(x_n))$. Since f_n is a minimizer of $E(\cdot, \beta_n)$ we have,

$$\beta_n \int_{\overline{B}_t(x_n)} (f_n - g)^2 + F(f_n, \overline{B}_t(x_n)) \leq \beta_n \int_{\overline{B}_t(x_n)} (v^t - g)^2 + F(v^t, \overline{B}_t(x_n))$$

Let $\underline{g}$ be the infimum and $\overline{g}$ the supremum of g in $\overline{B}_{\rho_n}(x_n)$. By a simple truncation argument as was used in the proof of lemma 4.17 it is easy to establish,

$$\forall x \in \overline{B}_{\rho_n}(x_n),\ f(x) \in [\underline{g} - \overline{w}_n - \overline{g - g_s}, \overline{g} + \overline{w}_n + \overline{g - g_s}]$$

and essentially the same argument shows,

$$\forall x \in \overline{B}_{\rho_n}(x_n),\ v^t(x) \in [\inf_{x \in \overline{B}_{\rho_n}(x_n)} f_n(x), \sup_{x \in \overline{B}_{\rho_n}(x_n)} f_n(x)]$$

Thus we obtain,

$$\begin{aligned} \sup_{x \in \overline{B}_t(x_n)} |v^t - g| &\leq \overline{g} - \underline{g} + 2(\overline{w}_n + \overline{g - g_s}) \\ &\leq 2(\rho_n L + \sqrt{8c}\beta_n^{-\frac{1}{4} - \frac{\gamma}{2}} + L\beta_n^{-\frac{1}{2}}) + 4\|(g - g_u)\chi_{\overline{B}_{\rho_n}(x_n)}\|_\infty \end{aligned}$$

By assumption 4.3.2 and since $\gamma < \frac{1}{2}$ and $\rho_n \leq \beta_n^{-\gamma}$ there exists an N such that if $n \geq N$ then,

$$\sup_{x \in \overline{B}_{\rho n (x_n)}} |v^t - g| \leq (1 + 3L)\beta_n^{-\gamma}$$

and hence,

$$\Psi(f_n, \overline{B}_t(x)) \leq (1 + 3L)^2 \beta_n^{1-2\gamma} \pi t^2.$$

Thus condition 4.3.30 of lemma 4.22 is clearly satisfied. Also, $\Psi(f_n, \overline{B}_t(x)) \leq \gamma t$ as long as $(1 + 3L)^2 \beta_n^{1-2\gamma} \pi t^2 < \gamma t$ i.e. for all $t < \frac{\gamma}{(1+3L)^2 \pi \beta_n^{1-2\gamma}}$. Now since $\Psi \leq F$ and $F(f_n, B_t(x_n)) \leq F(f_n, B_{\rho_n}(x_n))$ we have $\Psi(f_n, B_t(x_n)) \leq 2(c + \pi L^2)\beta_n^{-2\gamma}$ and for $t > \frac{2(c+\pi L^2)\beta_n^{-2\gamma}}{\gamma}$ we have $\Psi(f_n, \overline{B}_t(x)) \leq \gamma t$. Thus condition 4.3.31 of lemma 4.22 is satisfied if $\frac{2(c+\pi L^2)\beta_n^{-2\gamma}}{\gamma} < \frac{\gamma}{(1+3L)^2 \pi \beta_n^{1-2\gamma}}$. Since $\gamma > \frac{1}{4}$ this inequality is satisfied for n sufficiently large and the proof is now complete. Note that had we set $\gamma > \frac{1}{3}$ then the first bound would have been sufficient since for n large enough we would have $\rho_n \leq 2\beta_n^{-\gamma} \leq \frac{\gamma}{(1+3L)^2 \pi \beta_n^{1-2\gamma}}$

□

Finally we are ready to state the theorem to which the previous effort has been directed. Lemma 4.23 almost gives the theorem directly, the only problem is that as stated the lemma requires K be disjoint from the boundary of Ω. Fortunately the result can be extended to the boundary by a reflection argument.

Theorem 4.24 Let $g_u \in \mathrm{SBV}(\Omega)$ satisfy our assumptions and assume Ω is a rectangle. Given $\epsilon > 0$ there exists a constant $\beta' < \infty$ such that if $\beta \geq \beta'$ and f is a minimizer of $E(\cdot, \beta)$ with $g \in \Upsilon(\beta)$ then,

$$\mathcal{S}_f \subset [\mathcal{S}_{g_u}]_\epsilon.$$

Proof Let Ω_3 be a rectangle with the same center and proportions as Ω but 3 times the length. Similarly define Ω_2 with twice the length of Ω. We can define g_3 and f_3 on Ω_3 by reflection of g and f respectively. We note that g_3 satisfies our assumptions on Ω_3. Clearly f_3 minimizes,

$$E(u, \beta) = \beta \int_{\Omega_3} (g_3 - u)^2 + \int_{\Omega_3} |\nabla u|^2 + \mathcal{H}^1(\mathcal{S}_u)$$

over all $u \in \mathrm{SBV}(\Omega_3)$. Define $A = \overline{\Omega} \cap [\mathcal{S}_g]_\epsilon$ and let A_3 be the reflection of A onto Ω_3. $\overline{\Omega}_2 \backslash A_3$ is a compact set in Ω_3 disjoint from $\overline{\mathcal{S}}_{g_3}$. Thus, by lemma 4.23 there exists $\beta' < \infty$ such that if $\beta > \beta'$ then $\mathcal{S}_{f_3} \cap \overline{\Omega}_2 \backslash A_3 = \emptyset$ and hence $\mathcal{S}_f \subset [\mathcal{S}_{g_u}]_\epsilon$. □

The next lemma establishes that the opposite containment also holds.

Lemma 4.25 Let $g_u \in \mathrm{SBV}(\Omega)$ satisfy our assumptions and assume Ω is a rectangle. Given $\epsilon > 0$ there exists a constant $\beta' < \infty$ such that if $\beta \geq \beta'$ and f is a minimizer of $E(\cdot, \beta)$ with $g \in \Upsilon(\beta)$ then,

$$\mathcal{S}_{g_u} \subset [\mathcal{S}_f]_\epsilon.$$

Proof Assume that the lemma is false. Then there exists a $\epsilon > 0$, a sequence of minimizers $\{f_n\}$ with $\beta_n \to \infty$ and $g(n) \in \Upsilon(\beta_n)$ such that there is a corresponding sequence of points $x_n \in \mathcal{S}_{g_u}$ such that $\mathrm{dist}(x_n, \mathcal{S}_{f_n}) \geq \epsilon$. Let y be a cluster point of the sequence $\{x_n\}$. We can find $x \in \mathcal{S}_{g_u}$ satisfying $|x - y| < \frac{\epsilon}{2}$. Thus there exists a subsequence $\{f_{n_k}\}$ such that $\mathrm{dist}(x, \mathcal{S}_{f_{n_k}}) \geq \frac{\epsilon}{2}$. This then contradicts weak convergence of Jf_{n_k} to Jg_u which was proved in lemma 4.13. □

We are now ready to conclude the proof the theorem 4.12.

Proof **Of theorem 4.12** That $\lim_{\beta\to\infty} \sup_{g\in\Upsilon(\beta)} d_H(\mathcal{S}_f(\beta), \mathcal{S}_{g_u}) = 0$ follows from theorem 4.24 and lemma 4.25. The result of $\lim_{\beta\to\infty} \sup_{g\in\Upsilon(\beta)} |\mathcal{H}^1(\mathcal{S}_f) - \mathcal{H}^1(\mathcal{S}_{g_u}| = 0$ was established in lemma 4.14. This lemma together with the basic bound on E given in 4.3.4 implies that $\lim_{\beta\to\infty} \sup_{g\in\Upsilon(\beta)} \beta \int_\Omega (f-g)^2 = 0$. This concludes the proof. □

4.4 A Hybrid Case

As a final addition we wish to consider a mixed case in which the underlying image is piecewise constant but the functional we minimize is that associated with the piecewise smooth case. The reason for doing this will be made clear in chapter 5 where we propose various segmentation algorithms. Essentially all of the results carry over from the piecewise smooth case. Since $\nabla g_c = 0$ in this case things are greatly simplified.

4.4.1 Problem Formulation

As in the piecewise smooth case we treat the variational problem in the SBV setting. The difference here is that we weigh the smoothing term.

$$E(f,\beta,\lambda) = \beta \int_\Omega (g-f)^2 + \lambda \int_\Omega |\nabla f|^2 + \mathcal{H}^1(\mathcal{S}_f).$$

We will see that the only assumption we need make on λ is that it be bounded below by some constant $\lambda^* > 0$. Without loss of generality we set $\lambda^* = 1$. This greatly simplifies the statements of the results to follow. Our assumptions on the domain are as before i.e. Ω is a rectangle. The underlying image which we denote g_c is piecewise constant. The assumptions on it can be stated as follows,

Assumption 1 $g_c \in L^\infty(\Omega) \cap \mathrm{SBV}(\Omega), \nabla g_c = 0, \mathcal{H}^1(\mathcal{S}_{g_c}) < \infty.$

The Noise Model

As before the class of observed images is denoted by $\Upsilon(\beta)$. For this mixed case we assume the same noise model as in the piecewise smooth case i.e.

$$\lim_{\beta\to\infty} \sup_{g\in\Upsilon(\beta)} \beta \int_\Omega (g-g_c)^2 = 0, \tag{4.4.1}$$

and,

$$\forall\epsilon > 0,\ \lim_{\beta\to\infty} \sup_{g\in\Upsilon(\beta)} \|(g-g_c)(1-\chi_{[\mathcal{S}_{g_c}]_\epsilon})\|_\infty = 0 \tag{4.4.2}$$

We can now state the limit theorem for this case. It is essentially the same as in the piecewise smooth case.

Theorem 4.26 Under our stated assumptions, as $\beta \to \infty$ $\{\mathcal{S}_f(\beta)\}$ converges to $\mathcal{S}_{g_c}$ with respect to the Hausdorff metric, and $\mathcal{H}^1(\mathcal{S}_f(\beta)) \to \mathcal{H}^1(\mathcal{S}_{g_c})$. We mean by this that for any $\epsilon > 0$ there exists $\beta' < \infty$ such that if $\beta > \beta'$ and f is a minimizer of E with $\lambda \geq 1$ for some $g \in \Upsilon(\beta)$, then $d_H(\mathcal{S}_f, \mathcal{S}_{g_c}) < \epsilon$ and $|\mathcal{H}^1(\mathcal{S}_f) - \mathcal{H}^1(\mathcal{S}_{g_c})| < \epsilon$. Furthermore $\sqrt{\beta}(f-g_c)$ converges to 0 in $L^2(\Omega)$.

4.4.2 Preliminary Results

The results of this section are essentially the same as in the piecewise smooth case. The only thing we need check is the dependence on λ, and the effect of setting $\nabla g_c = 0$.

Let $E^*(\beta,\lambda)$ denote the minimal value of $E(f,\beta,\lambda)$. By substituting g_c for f we get the following bound,

$$E^*(\beta,\lambda) \leq \beta \int_\Omega (g-g_c)^2 + \mathcal{H}^1(\mathcal{S}_{g_c}). \tag{4.4.3}$$

Lemma 4.27 If $g_n, f_n, \beta_n, \lambda_n$ are sequences such that $\beta_n \uparrow +\infty$, $\lambda_n \geq 1$ for all n, and $E(f_n,\beta_n,\lambda_n) = E^*(\beta_n,\lambda_n)$ with $g = g_n \in \Upsilon(\beta_n)$ then,

$$\begin{aligned} f_n &\to g_c &&\text{in } L^1(\Omega) \\ Jf_n &\to Jg_c &&\text{weakly as radon measures} \\ \nabla f_n &\to 0 &&\text{weakly in } L^1(\Omega;\Re^2). \end{aligned}$$

Proof The proof is the same as in lemma 4.27. We need the uniform lower bound on λ so that L. Ambrosio's compactness theorem for SBV(Ω) can be applied. $\square$

Lemma 4.28 If $g_n, f_n, \beta_n, \lambda_n$ are sequences such that $\beta_n \uparrow +\infty$, $\lambda_n \geq 1$ for all n, and $E(f_n, \beta_n, \lambda_n) = E^*(\beta_n, \lambda_n)$ with $g = g_n \in \Upsilon(\beta_n)$ then,

$$\begin{aligned}\lim_{n\to+\infty} \lambda_n \int_\Omega |\nabla f_n|^2 &= 0\\ \lim_{n\to+\infty} \mathcal{H}^1(\mathcal{S}_{f_n}) &= \mathcal{H}^1(\mathcal{S}_{g_c})\end{aligned}$$

Proof We can argue exactly as in lemma 4.14 with g_u replace by g_c (which satisfies $\nabla g_c = 0$) to get the second statement, after noting $\lambda_n \geq 1$. We also have,

$$\limsup_{n\to\infty} \lambda_n \int_\Omega |\nabla f_n|^2 + \mathcal{H}^1(\mathcal{S}_{f_n}) \leq \mathcal{H}^1(\mathcal{S}_{g_c})$$

from the bound 4.4.3. This establishes the first statement. $\square$

Corollary If $A \subset \Omega$ is any borel set such that $\text{dist}(A, \overline{\mathcal{S}}_{g_c} \cup \partial\Omega) > 0$ then,

$$\lim_{n\to+\infty} \mathcal{H}^1(\mathcal{S}_{f_n} \cap A) = 0.$$

Proof The proof is the same as for the Corollary to lemma 4.14. $\square$

Lemma 4.29 Let $K \in \Omega$ be any compact set disjoint from $\overline{\mathcal{S}}_{g_c} \cup \partial\Omega$ and $g_n, f_n, \beta_n, \lambda_n$ be sequences such that $\beta_n \uparrow +\infty$, $\lambda_n \geq 1$ and $E(f_n, \beta_n, \lambda_n) = E^*(\beta_n, \lambda_n)$ with $g = g_n \in \Upsilon(\beta_n)$. It then follows that

$$\lim_{n\to+\infty} \sup_{x\in K} |f_n^+(x) - f_n^-(x)| = 0.$$

Proof The proof is the same as in lemma 4.17 with g_u replaced by g_c. The constant L can be set to 0 and e_n is here defined as $f_n - g_c$ which satisfies $\lim_{n\to\infty}(\int_\Omega e_n^2 + \int_\Omega |\nabla e_n|^2) = 0$. $\square$

We need to make a slight change of notation from what was used earlier. For $\overline{B}_\rho(x) \in \Omega$ and $u \in$ SBV(Ω) we define,

$$\begin{aligned}J(u,\beta,\lambda,\rho,x) &= \beta \int_{B_\rho(x)} u^2 + \lambda \int_{B_\rho(x)} |\nabla u|^2 + \mathcal{H}^1(\mathcal{S}_u \cap B_\rho(x))\\ &\text{and}\\ J'(u,\beta,\rho,x) &= \beta \int_{\partial B_\rho(x)} u^2 \, d\mathcal{H}^1 + \lambda \int_{\partial B_\rho(x)} |\nabla u|^2 \, d\mathcal{H}^1 + \mathcal{H}^0(\mathcal{S}_u \cap \partial B_\rho(x))\end{aligned}$$

We note that proposition 4.18 holds for J and J' with the new definition as well as with the previous one.

4.4.3 Main Results

Lemma 4.30 Let $g_c \in \mathrm{SBV}(\Omega)$ satisfy our assumptions and $K \subset \Omega$ be a compact set such that $K \cap \overline{S}_{g_c} = \emptyset$. Define $\delta = \frac{1}{2}\mathrm{dist}(K, (S_{g_c} \cup \partial\Omega))$. Let $0 < \gamma < \frac{1}{2}$. For any sequences $\beta_n \uparrow \infty$, $\lambda_n \geq 1$ and f_n such that f_n is a minimizer for $E(\beta_n, \lambda_n)$ with $g \in \Upsilon(\beta_n)$ the following holds,

$$\lim_{n\to\infty} \sup_{x\in K} \beta_n^{2\gamma} J(f_n - g_c, \beta_n, \lambda_n, \beta_n^{-\gamma}, x) = 0$$

Proof Assume the lemma is false. There exists a K and γ satisfying the conditions of the lemma, a constant $\eta > 0$ and a sequence of quintuples, $\{(g_n, f_n, \beta_n, \lambda_n, x_n)\}$ such that $\beta_n \uparrow +\infty$, $\lambda_n \geq 1$, $E(f_n, \beta_n, \lambda_n) = E^*(\beta_n, \lambda_n)$ with $g = g_n \in \Upsilon(\beta_n)$, $x_n \in K$ and,

$$J(w_n, \beta_n, \lambda_n, \beta_n^{-\gamma}, x_n) > \eta\beta_n^{-2\gamma} \tag{4.4.4}$$

for each n, where we have used the notation $w_n = f_n - g_c$.

Since $\lim_{n\to+\infty} \beta_n \int_\Omega w_n^2 + \lambda_n \int_\Omega |\nabla w_n|^2 = 0$ the corollary to lemma 4.28 yields for n sufficiently large,

$$J(w_n, \beta_n, \lambda_n, 2\beta_n^{-\gamma}, x_n) \leq 1. \tag{4.4.5}$$

Thus there exists an N_1 such that if $n \geq N_1$ then 4.4.5 holds and also $\beta_n^{\gamma} \ln \frac{\beta_n^{2\gamma}}{\eta} < \frac{\beta_n^{\frac{1}{2}}}{32}$. We now conclude from 4.4.4 and proposition 4.18 that for each $n \geq N_1$ we can choose $\rho_n \in (\beta_n^{-\gamma}, 2\beta_n^{-\gamma})$ such that,

$$J'(w_n, \beta_n, \lambda_n, \rho_n, x_n) \leq \beta_n^{\frac{1}{2}} \frac{1}{16} J(w_n, \beta_n, \lambda_n, \rho_n, x_n) \tag{4.4.6}$$

Let us define $\tilde{w}_n$ as in 4.3.13 with $\rho_n' = \rho_n - \beta_n^{-\frac{1}{2}}$ and the balls centered at x_n. From lemma 4.19 and 4.4.6 we derive,

$$J(\tilde{w}_n, \beta_n, \lambda_n, \rho_n, x_n) \leq \frac{1}{8} J(w_n, \beta_n, \lambda_n, \rho_n, x_n) \tag{4.4.7}$$

Let $\tilde{f}_n = g_c + \tilde{w}_n$ in $B_{\rho_n}(x_n)$ and f_n elsewhere. Since f_n is a minimizer of $E(f, \beta_n, \lambda_n)$ we have, $E(f_n, \beta_n, \lambda_n) \leq E(\tilde{f}_n, \beta_n, \lambda_n)$. We can express this in terms of w_n as,

$$J(w_n, \beta_n, \lambda_n, \rho_n, x_n) = J(\tilde{w}_n, \beta_n, \lambda_n, \rho_n, x_n) + 2\beta_n \int_{B_{\rho_n}(x_n)} (\tilde{w}_n - w_n)(g_c - g)$$

Substituting from 4.4.7 we get,

$$\frac{7}{16} J(w_n, \beta_n, \lambda_n, \rho_n, x_n) \leq \beta_n \int_{B_{\rho_n}(x_n)} (\tilde{w}_n - w_n)(g_c - g) \tag{4.4.8}$$

As in lemma 4.21 we obtain,

$$\int_{B_{\rho_n}(x_n)} |w_n| \leq \beta_n^{-\frac{1}{2}} \sqrt{\pi} \rho_n (J(w_n, \beta_n, \lambda_n, \rho_n, x_n))^{\frac{1}{2}} \tag{4.4.9}$$

and

$$\int_{B_{\rho_n}(x_n)} |\tilde{w}_n| \leq \beta_n^{-\frac{3}{4}} \sqrt{2\pi\rho_n} (\frac{1}{16} J(w_n, \beta_n, \lambda_n, \rho_n, x_n))^{\frac{1}{2}} \tag{4.4.10}$$

Let $\vartheta_n = \beta_n^{-\frac{1}{2}} \sup_{y \in [K]_{2\beta_n^{-\gamma}}} |g_c(y) - g(y)(n)|$. Combining 4.4.9 and 4.4.10 and substituting into equation 4.4.8 we obtain,

$$\frac{7}{16} J(w_n, \beta_n, \lambda_n, \rho_n, x_n) \leq \vartheta_n \sqrt{\pi} \rho_n (1 + \frac{\sqrt{2}}{4} \frac{\beta_n^{-\frac{1}{4}}}{\sqrt{\rho_n}}) (J(w_n, \beta_n, \lambda_n, \rho_n, x_n))^{\frac{1}{2}}$$

Now, since $\rho_n \geq \beta_n^{-\gamma}$, $\gamma < \frac{1}{2}$ and $\beta_n \geq 1$ we have $(1 + \frac{\sqrt{2}}{4} \frac{\beta_n^{-\frac{1}{4}}}{\sqrt{\rho_n}}) < \frac{7}{4}$ and hence we obtain,

$$\frac{1}{4} J(w_n, \beta_n, \lambda_n, \rho_n, x_n) < \vartheta_n \sqrt{\pi} \rho_n (J(w_n, \beta_n, \lambda_n, \rho_n, x_n))^{\frac{1}{2}}$$

Since,

$$\lim_{n \to \infty} \beta_n^{\frac{1}{2}} \sup_{y \in [K]_{2\beta_n^{-\gamma}}} |g_c(y) - g(y)| = 0$$

and $\rho_n \leq 2\beta_n^{-\gamma}$ we conclude,

$$\lim_{n \to \infty} \beta_n^{2\gamma} J(w_n, \beta_n, \lambda_n, \rho_n, x_n) = 0$$

which contradicts 4.4.4. Q.E.D.

□

Lemma 4.31 Let $g_c \in \text{SBV}(\Omega)$ satisfy our assumptions and $K \subset \Omega$ be a compact set such that $K \cap \overline{\mathcal{S}}_g = \emptyset$. There exists a constant $\beta' < \infty$ such that if $\beta \geq \beta'$ and f is a minimizer of $E(\cdot, \beta, \lambda)$ with $\lambda > 1$ then,

$$\mathcal{S}_f \cap K = \emptyset$$

Proof Assume the lemma is false. There exists a K satisfying the conditions of the lemma and a sequence of quintuples $\{(g_n, f_n, \beta_n, \lambda_n, x_n)\}$ such that $\beta_n \uparrow +\infty$, $\lambda_n > 1$, $E(f_n, \beta_n, \lambda_n) = E^*(\beta_n, \lambda_n)$ with $g = g_n \in \Upsilon(\beta_n)$ and $x_n \in K \cap \mathcal{S}_f$. Define $\delta = \frac{1}{2}\text{dist}(K, \mathcal{S}_{g_c} \cup \partial\Omega)$. Fix any real γ satisfying $\frac{1}{4} < \gamma < \frac{1}{2}$. By lemma 4.30 we can assume,

$$J(w_n, \beta_n, \beta_n^{-\gamma}, x_n) < \beta_n^{-2\gamma} \tag{4.4.11}$$

for each n, where we have again used the notation $w_n = f_n - g_c$. Furthermore, for convenience we make the assumption $\beta_n^{-\gamma} < \frac{1}{4}$. From 4.4.11 we have $\mathcal{H}^1(\mathcal{S}_{f_n} \cap \beta_n^{-\gamma}) < \beta_n^{-2\gamma}$. Thus $|\{\rho \in [\frac{1}{2}\beta_n^{-\gamma}, \beta_n^{-\gamma}] : \overline{\mathcal{S}}_f \cap \partial B_\rho(x_n) = \emptyset\}| > \frac{1}{2}\beta_n^{-\gamma} - \beta_n^{-2\gamma} > \frac{1}{4}\beta_n^{-\gamma}$ by the assumption just made . We can thereby choose a $\rho_n \in [\frac{1}{2}\beta_n^{-\gamma}, \beta_n^{-\gamma}]$ such that $\overline{\mathcal{S}}_{f_n} \cap \partial B_{\rho_n}(x_n) = \emptyset$ and,

$$\beta_n \int_{\partial B_{\rho_n}(x_n)} w_n^2 \, d\mathcal{H}^1 + \lambda_n \int_{\partial B_{\rho_n}(x_n)} |\nabla w_n|^2 \, d\mathcal{H}^1 < 4c\beta_n^{-\gamma}. \tag{4.4.12}$$

Define,

$$\overline{w}_n = \max_{x \in \partial B_{\rho_n}(x_n)} |w_n(x)|$$

and,

$$\overline{g - g_c} = \max_{x \in \partial B_{\rho_n}(x_n)} |g(x) - g_c(x)|$$

From the existence and regularity results for minimizers of $E(\cdot,\cdot)$ we know w_n is C^1 on $\partial B_{\rho_n}(x_n)$ so from proposition 4.16 and 4.4.12 we conclude

$$\max_{x \in \partial B_{\rho_n}(x_n)} |w_n(x)| \leq \sqrt{8\eta}\beta_n^{-\frac{1}{4}-\frac{\gamma}{2}}\lambda_n^{-\frac{1}{4}}.$$

Our goal in the remainder of the proof is to show that the three conditions of lemma 4.22 are satisfied for n sufficiently large with $u = \sqrt{\lambda} f$, $\rho = \rho_n$ and $x = x_n$, thus obtaining a contradiction with $x_n \in \mathcal{S}_{f_n}$. Now,

$$\begin{aligned} F(\sqrt{\lambda_n} f_n, \overline{B}_{\rho_n}(x_n)) &= \lambda_n \int_{B_{\rho_n}(x_n)} |\nabla f_n|^2 + \mathcal{H}^1(\mathcal{S}_{f_n} \cap B_{\rho_n}(x_n)) \\ &= \lambda_n \int_{B_{\rho_n}(x_n)} |\nabla w_n|^2 + \mathcal{H}^1(\mathcal{S}_{f_n} \cap B_{\rho_n}(x_n)) \\ &\leq \beta_n^{-2\gamma} \end{aligned}$$

Condition 4.3.29 is thus satisfied as long as $\beta_n^{-\gamma} \leq \xi$ which is clearly true for n sufficiently large.

Consider a fixed n, let $0 < t \leq \rho_n$ and let $\sqrt{\lambda_n}v^t \in \mathrm{SBV}(\Omega)$ be any function that realizes $\Phi(\sqrt{\lambda_n}f_n, \overline{B}_t(x_n))$, i.e. $v^t(x) = f_n(x)$ for all $x \in \Omega\backslash\overline{B}_t(x_n)$ and $F(\sqrt{\lambda_n}v^t, \overline{B}_t(x_n) = \Phi(\sqrt{\lambda_n}f_n, \overline{B}_t(x_n))$. Since f_n is a minimizer of $E(\cdot, \beta_n, \lambda_n)$ we have,

$$\beta_n \int_{\overline{B}_t(x_n)} (f_n - g)^2 + F(\sqrt{\lambda_n}f_n, \overline{B}_t(x_n)) \leq \beta_n \int_{\overline{B}_t(x_n)} (v^t - g)^2 + F(\sqrt{\lambda_n}v^t, \overline{B}_t(x_n))$$

By a simple truncation argument it is easy to establish,

$$\forall x \in \overline{B}_{\rho_n}(x_n),\ f(x) \in [-\overline{w}_n - \overline{g - g_c}, \overline{w}_n + \overline{g - g_c}]$$

and essentially the same argument shows,

$$\forall x \in \overline{B}_{\rho_n}(x_n),\ v^t(x) \in [\inf_{x \in \overline{B}_{\rho_n}(x_n)} f_n(x), \sup_{x \in \overline{B}_{\rho_n}(x_n)} f_n(x)]$$

Thus we obtain,

$$\begin{aligned} \sup_{x \in \overline{B}_t(x_n)} |v^t - g| &\leq 2(\overline{w}_n + \overline{g - g_c}) \\ &\leq 2(\sqrt{8c}\beta_n^{-\frac{1}{4}-\frac{\gamma}{2}}\lambda_n^{-\frac{1}{4}} + 2\|(g - g_c)\chi_{\overline{B}_{\rho_n}(x_n)}\|_\infty \end{aligned}$$

By assumption 4.4.2 and since $\gamma < \frac{1}{2}$ and $\rho_n \leq \beta_n^{-\gamma}$ there exists an N such that if $n \geq N$ then,

$$\sup_{x \in \overline{B}_{\rho_n(x_n)}} |v^t - g| \leq \pi^{-1}\beta_n^{-\gamma}$$

and hence,

$$\Psi(f_n, \overline{B}_t(x)) \leq \beta_n^{1-2\gamma}t^2.$$

Thus condition 4.3.30 of lemma 4.22 is clearly satisfied. Also, $\Psi(f_n, \overline{B}_t(x)) \leq \gamma t$ as long as $\beta_n^{1-2\gamma}t^2 < \gamma t$ i.e. for all $t < \gamma\beta_n^{2\gamma-1}$. Now since $\Psi \leq F$ and $F(f_n, B_t(x_n)) \leq F(f_n, B_{\rho_n}(x_n))$ we have $\Psi(f_n, B_t(x_n)) \leq \beta_n^{-2\gamma}$ and as long as $t > \gamma^{-1}\beta_n^{-2\gamma}$ we have $\Psi(f_n, \overline{B}_t(x)) \leq \gamma t$. Thus condition 4.3.31 of lemma 4.22 is satisfied if $\gamma^{-1}\beta_n^{-2\gamma} < \gamma\beta_n^{2\gamma-1}$. Since $\gamma > \frac{1}{4}$ this inequality is satisfied for n sufficiently large and the proof is now complete. □

We are now ready to conclude the proof the theorem 4.26.

Proof **Of theorem 4.26** To show $\lim_{\beta\to\infty}\sup_{g\in\Upsilon(\beta)} d_H(\mathcal{S}_f(\beta), \mathcal{S}_{g_c}) = 0$ we need to show $\mathcal{S}_{g_c} \subset [\mathcal{S}_f]_\epsilon$ and also $\mathcal{S}_f \subset [\mathcal{S}_{g_c}]_\epsilon$. Both of these statements follow from arguments exactly paralleling those use in theorem 4.24 and lemma 4.25. The result of $\lim_{\beta\to\infty}\sup_{g\in\Upsilon(\beta)} |\mathcal{H}^1(\mathcal{S}_f) - \mathcal{H}^1(\mathcal{S}_{g_c}| = 0$ was established in lemma 4.28. This lemma together with the basic bound on E, 4.4.3 implies that $\lim_{\beta\to\infty}\sup_{g\in\Upsilon(\beta)} \beta\int_\Omega (f-g)^2 = 0$. This concludes the proof.

□

Chapter 5

Scale Independent Segmentation

In this chapter we propose an algorithm for segmenting images which attempts to locate boundaries accurately independently of the "scale" of the features being sought. We will develop the algorithm beginning with a conceptual sketch and concluding with simulation results. In the first section we present a paradigm, of which our method is but one possible implementation, which contains within it the key ideas and most of the intuition behind our approach. The following section contains a more formal description of the algorithm. Each step of the paradigm is expanded into a mathematical formalism, which indicates, without explicitly representing, the main computational steps of the algorithm. It is at that point that we relate the algorithm to the limit theorems. An analogy between the form of the limit theorems and the structure of the algorithm will be drawn.

We also address the issue of computation in this chapter. The questions we ask are: how can one actually determine a segmentation and what are the differences between the mathematical formalism that has been the center of attention thus far in the thesis and a truly computational model? The second section of this chapter contains a model on which we build our computations. This model is based on the Γ–convergent approximation to the variational problem. The next step we take towards a computational algorithm is the consideration of the issue of discretization. There are some some fundamental difficulties associated with faithfully discretizing the variational formulation and we point out some of them in Section 5.3. With these difficulties in view we propose a particular discrete model on which our simulations are based. Finally in the last section we consider some of the remaining issues concerned with the computation. There are several options one has in performing the computation which are not resolved by the formal mathematical model. We discuss how selecting among these options changes the performance of the algorithm. This last section also contains our simulation results which illustrate and support the discussion.

5.1 A Paradigm

It was pointed out in Chapter 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 other segmentation techniques as well.) Consider for example the image of square such as in Figure 1.5. The Γ which minimizes E_0 will be the empty set whenever $\frac{\alpha}{\beta} < \frac{a}{2}$ is satisfied. Now suppose the goal of the segmentation was to recover objects only above a certain scale. Consider Figure 5.1 for example. If one were trying to find objects on the scale of the larger square and not those on the scale of the smaller square then it is necessary to incur an error at the corners of at least $(\sqrt{2}-1)b$ as illustrated in Figure 5.1.

Now, the limit theorems proved in the preceding chapter state that as the parameters tend to appropriate limits ($\beta \to \infty$ for example) 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 theorems do not provide us with any means of circumventing the scale/accuracy trade–off because they state that the limit of the minimizing Γ 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. Our response to this question is affirmative and takes the form of the algorithm outlined below.

Recall that in the limit theorems we have quantified the amount of noise which can be allowed while retaining convergence of the boundaries. For example, in the

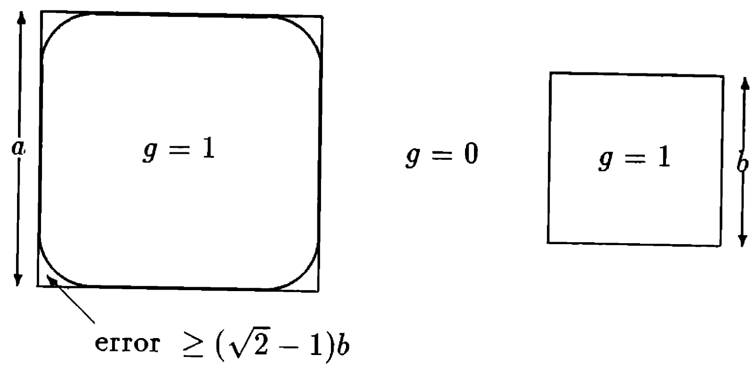


Figure 5.1: Segmentation of Two Squares with $a > \frac{\alpha}{\beta} > b$

piecewise constant and the piecewise smooth cases we considered sequences β_n, g_n such that $\{g_n\}$ converges to the ideal image g_∞ according to $\lim_{n\to\infty} \beta_n \int_\Omega (g_n - g_\infty)^2 = 0$, where Ω is the domain of the image, and $\lim_{n\to\infty} \sqrt{\beta_n}\|(1 - \chi_{[S_{g_\infty}]_\epsilon})(g_n - g_\infty)\|_\infty = 0$ for any $\epsilon > 0$ (where S_{g_∞} is the support set of the discontinuities of g_∞). Now, let Φ_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 Chapter 4, under some mild regularity assumptions on g_∞, 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

$$\lim_{n\to\infty} \beta_n r_n = 0 \tag{5.1.1}$$

and ϑ_n being a sequence of constants satisfying

$$\lim_{n\to\infty} \sqrt{\beta_n}\vartheta_n = 0. \tag{5.1.2}$$

Suppose our goal now is to obtain a coarse scale segmentation of the image. We can think of the data as the ideal coarse scale approximation corrupted by "noise" which will consist of small scale features as well as what is normally considered as noise, measurement noise for example. Now we envision an algorithm which attempts to recover the coarse scale segmentation by taking the same limit as in the limit theorems. The algorithm must not only minimize the variational problem while the parameters are varied, it must also remove the noise in accordance with the parameter variations as dictated by the limit theorems.

The fallacy with this project as we have described it is that we do not have a definition of the ideal coarse scale approximation. We will propose our algorithm and its result will constitute the coarse scale approximation. Nevertheless the algorithm we will propose will proceed in a manner consistent with the scenario just described and it is within this scenario that it can be best understood. The central questions concerning the implementation are how the "noise" is to be removed and how we will assure that the coarse scale approximation found is a "good" one. The idea we suggest is to use the solutions of the variational problem as a rough approximation to our "ideal" coarse scale image. Given a solution to the variational problem we will smooth the data outside of some neighborhood of the boundaries belonging to the solution which are tacitly assumed to be close to the ideal ones. Within the same neighborhood we will leave the image essentially unchanged. In this way the information required for good localization of large scale boundaries found by the variational approach is retained in its original form while smaller scale features and noise disjoint from these regions, where detailed information is not required, are smoothed out. The rates which govern this procedure are naturally derived from the limit theorems. It is the smoothing which attempts to provide for the circumventing of the scale/accuracy trade–off.

We outline the paradigm below in a schematic fashion. We emphasize however that the paradigm is conceptual and does not directly indicate our intended computational procedure. A proposal for a method of computation will be detailed later. The paradigm requires within it two key operations or procedures:

P1 *The Minimization* E *(or* E_0*) to produce* f *and* Γ *with the parameters* α *and* β *as input variables*

P2 *The updating or altering of the image by smoothing outside some neighborhood of* Γ *and updating the parameters which provide the data for P1 for resolution*

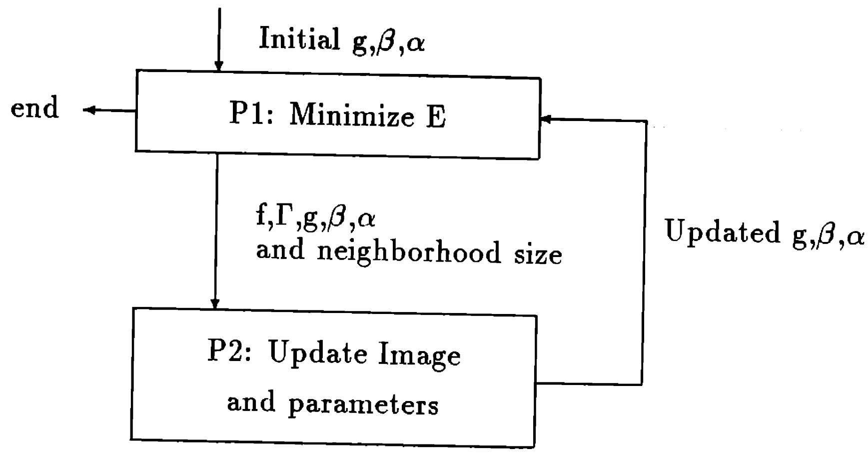


Figure 5.2: A Schematic for Scale Independent Segmentation

at smaller scale.

The interaction between these two procedures is illustrated in Figure 5.2. The algorithm begins by minimizing E with the function g set to the original data and the parameters α and β chosen to provide a segmentation of the desired scale. Once the minimizing f and Γ 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 Γ while leaving it unchanged inside the neighborhood. Since f is a smooth approximation to g which respects the boundaries Γ a simple smoothing technique might be to take a convex combination of f and g (this is what we have done in our simulations). Simultaneously, we replace our parameters β and α as if to find segmentations on a smaller scale. We then re–solve the problem of minimizing E, the difference being that we use the updated image and parameter values. The hope is that we should detect essentially the same boundaries as before only now with finer resolution. This procedure can then be iterated on until sufficient accuracy is attained.

Within this paradigm there still exists considerable flexibility. How one does the smoothing or chooses neighborhoods is somewhat free. Also, we have not stated how we will find a minimizer of the variational problem. Of course, in order for the procedure to work the various parameters and operations must be coordinated properly. It is at this point that the limit theorems become useful. If we remove the update of the image than the paradigm would simply be a taking of the same limit as taken in the limit theorems. The deficiency with the limit theorems with regard to their direct application in doing the segmentation, as we have mentioned, is the fact that the limits predict the convergence of the solution Γ to the entire discontinuity set of the image. If the image was noisy this could effectively result in boundaries being put almost everywhere. The update of the image plays a role similar to the rescaling of the noise as required by the limit theorems i.e. it is the smoothing step which we use in *P2* which provides the cure for the deficiency in the limit theorem as applied to segmentation. 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.

With small scale features thus interpreted as noise the limit theorems then tell us how quickly this noise must decay as the parameters tend to their limits i.e. the conditions 5.1.1 and 5.1.2 should be satisfied for the sequence of images produced. The minimal amount of smoothing which must be performed and the rate of decay of the size of the neighborhood of the conjectured boundaries in which smoothing is not done is thus determined.

5.1.1 The Piecewise Constant Case

In this section we present a more detailed implementation of the paradigm discussed in the previous section for the case in which we are seeking a piecewise constant approximation and we do so by minimizing E_0. According to the paradigm the algorithm produces sequences $f_n, \Gamma_n, g_n, w_n, \beta_n$ such that f_n, Γ_n are found by minimizing $E_0(\beta_n, g_n)$. The procedure is initialized by setting $g_0 = g$ where g is the original data and by choosing w_0 which denotes the initial neighborhood size within which we suppress the smoothing. The remaining quantities, g_n, β_n and the neighborhood sizes w_n are defined according to the following schedule,

$$\begin{aligned} g_{n+1} &= g_n + h_n(x)a\epsilon(f_n - g_n) \\ \beta_{n+1} &= (1-\epsilon)^{-2}\beta_n \end{aligned}$$

$$w_{n+1} \;=\; (1-b\epsilon)^2 w_n$$

where a and b are constants and h_n is a function. The constants $a \simeq 1$ and $b \simeq 1$ will satisfy $1 \leq a, b$. 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 $1-\chi_{[\Gamma_n]_{w_n}}$. The parameter w_0, as we mentioned, is a predefined constant and represents the initial estimate of the error in the boundary locations. The formalism we have just presented is a discrete one i.e. we produce a discrete sequence of images and parameters. We could just as well have made the parameter n continuous and represented the process with differential equations. In this sense we can say that ϵ represents the step size of the algorithm.

We mention in passing that an alternative formula for the update of g_n is,

$$g_{n+1} = (1-h_n(x))g_0 + h_n(x)(1-a\epsilon)g_n + h_n(x)a\epsilon f_n.$$

This formulation keeps some of the original data in a neighborhood of the set Γ_n and may be preferable depending on how one performs step *P1* of the paradigm.

We claim that for $a, b > 1$ these rates are consistent with the limit theorems proved in Chapter 4. Assume that the sequence $\{\Gamma_n\}$ produced by the algorithm converges in the Hausdorff metric to some Γ_∞ and hence that g_n converges to some piecewise constant g_∞. If we identify r_n with w_n and ϑ_n with the value of $\|(1-\chi_{[\Gamma_\infty]_\epsilon})(g_n-g_\infty)\|_\infty$ (where $\epsilon > 0$ is arbitrary), then r_n satisfies 5.1.1 and ϑ_n satisfies 5.1.2 as long as $b, a > 1$. In fact we obtain $\beta_n r_n = O((\frac{1-b\epsilon}{1-\epsilon})^{2n})$ and $\sqrt{\beta_n}\vartheta_n = O((\frac{1-a\epsilon}{1-\epsilon})^n)$.

By completely different means we can show that if $b > 1$ and $a \geq 1$ then the value of E_0^n remains bounded above so the algorithm which has been described is apparently well behaved in the sense that the quantity of boundary admitted into the solution is bounded. To justify this claim we mention the fact that the boundary Γ which represents a minimizer of E_0 is a finite number of rectifiable curves all joined either at their end points or to $\partial\Omega$ and that this implies the following inequality,

$$\forall\epsilon > 0, \;\; |[\Gamma]_\epsilon \cap \Omega| \leq 2\epsilon\mathcal{H}^1(\Gamma) \tag{5.1.3}$$

This result is a simple generalization of proposition 2.6. Now consider a single update of the algorithm. We have minimized E_0 and found f and Γ. We define the new image, which we denote with the superscript $'$, by,

$$g' = \begin{cases} (1-a\epsilon)g + a\epsilon f & \text{for } x \in \Omega \backslash [\Gamma]_w \\ g & \text{for } x \in [\Gamma]_w \end{cases}$$

where w is the width of the neighborhood in which we keep g unchanged. The parameter update is $\beta' = (1-\epsilon)^{-2}\beta$, and $\alpha' = \alpha$. We denote the new cost function by E_0'. To bound E_0' we consider setting $\Gamma' = \Gamma$ and $f' = f$.

$$\begin{aligned} E_0'(f,\Gamma) &= \beta' \int_\Omega (f-g')^2 + \alpha \mathcal{H}^1(\Gamma) \\ &= (\frac{1-a\epsilon}{1-\epsilon})^2 \beta \int_\Omega (f-g)^2 + \alpha \mathcal{H}^1(\Gamma) + (\beta' - \beta) \int_{\Omega \cap [\Gamma]_w} (f-g)^2 \\ &\leq E_0(f,\Gamma) + (\beta' - \beta) \int_{\Omega \cap [\Gamma]_w} (f-g)^2 \\ &\leq E_0(f,\Gamma) + c\epsilon\beta |\Omega \cap [\Gamma]_w| ||g||_\infty^2 \end{aligned}$$

where $c = (\frac{1}{1-\epsilon} + \epsilon) \simeq 1$. Now from equation 5.1.3 we obtain,

$$\begin{aligned} E_0'(f,\Gamma) &\leq E_0(f,\Gamma) + c\epsilon\beta\, 2w\mathcal{H}^1(\Gamma)||g||_\infty^2 \\ &\leq (1 + c\epsilon\beta\, 2w||g||_\infty^2) E_0(f,\Gamma) \end{aligned}$$

At each successive stage we reduce w by a factor of $(1-b\epsilon)^2$. Thus if E_0^{n*} denotes the cost after the nth stage we can iterate on the argument given above to establish,

$$E_0^{n*} \leq E_0^{0*} \Pi_{i=0}^n (1 + 2c\epsilon(\frac{1-b\epsilon}{1-\epsilon})^{2i} \beta_0 w_0 ||g||_\infty^2).$$

Now for $x > 0$, $\ln(1+x) \leq x$ so,

$$\begin{aligned} \Pi_{i=0}^n (1 + 2c\epsilon(\frac{1-b\epsilon}{1-\epsilon})^{2i} \beta_0 w_0 ||g||_\infty^2) &\leq \exp \sum_{i=0}^n 2c\epsilon(\frac{1-b\epsilon}{1-\epsilon})^{2i} \beta_0 w_0 ||g||_\infty^2 \\ &\leq \exp \frac{2c\epsilon\beta_0 w_0 ||g||_\infty^2}{1 - (\frac{1-b\epsilon}{1-\epsilon})^2} \\ &\leq \infty. \end{aligned}$$

This shows that the total boundary remains bounded. As ϵ tends to zero the bound converges to $\exp \frac{\beta_0 w_0 ||g||_\infty^2}{b-1}$. Thus as b tends to 1 this bound blows up. The parameter b thus has some control over the stability of the algorithm. A larger value b promotes the stabilization of the boundaries which is obiously a dersirable property. On the other hand it is not clear whether an a priori bound in the length of the boundary is desirable. In the continuous domain setting, which is what we are considering here, a boundary can have arbitrarily large length since it can be highly irregular. Why then should we bound the length of our solution? Thus it appears there may be a trade off between the predictable stability of the algorithm and the level of detail it can be expected to resolve. For $b = 1$ we do not have a finite bound on the length of the limit boundary. This then is a borderline case perhaps admitting convergence to the 'true' boundaries. However, the limit theorems do not provide a clear interpretation of this case. Fortunately in all practical situations there is a lower bound on what constitutes meaningful resolution and a bound on the length of the boundaries becomes inconsequential once large enough. It is worth noting however that by making b larger we can guarantee greater stability of the algorithm but the cost is a reduction in attainable resolution.

In more general terms we can argue why we expect the set of boundaries found to remain essentially unchanged throughout the iterations of the algorithm. For simplicity, consider what would happen if we set $h_n(x) = 1$ for all n i.e. $w_0 = 0$ and also $a = 1$. It is not very difficult to check that the solution f_0, Γ_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 when the image is a square as in Figure 1.5. 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. Also it is reasonable to expect that the boundaries found by solving the updated problem will lie inside the set $\chi_{[\Gamma_n]_{w_n}}$ and that for n large it should be approximately true that $\chi_{[\Gamma_n]_{w_n}} \subset \chi_{[\Gamma_n]_{w_{n-1}}}$. When this expectation is fulfilled there is effectively no difference between the two possible update formulas for g_n. For actual computation the first version of the update makes the algorithm more predictable and hence more easily stabilized. The second formula offers the possibility of better localization and the growth of boundaries. In our simulations we have used the second form of the update equation.

5.1.2 The Piecewise Smooth Case

The piecewise smooth case admits a discussion very much like that provided for the piecewise constant case. There is a slight complication here in that we cannot assert the inequality 5.1.3 due to the possible existence of cracked tips i.e. boundaries which terminate at an interior point of Ω. If D is the number of connected components in Γ then we have the following,

$$\forall \epsilon > 0, \quad |[\Gamma]_\epsilon \cap \Omega| \leq 2\epsilon \mathcal{H}^1(\Gamma) + D\pi\epsilon^2 \tag{5.1.4}$$

We can see that without some regularity assumptions on the minimizing Γ we have no control over the area of an ϵ neighborhood. Our proposed computational scheme differs from that given above in that we do not actually determine Γ but rather we minimize the Γ–convergent approximation to the variational problem discussed in Section 3.3. This approximation bears a close tie to Minkowski content so we can gain control over the size of the neighborhood this way. We will discuss this further in Section 5.4. If we proceed under the assumption that D remains bounded then by following essentially the same line as in the piecewise constant case we can reach basically the same conclusion although we obtain a slightly weaker bound.

$$E^{n*} < \left(E^0(f,\Gamma) + \frac{c\epsilon\beta D\pi w_0^2}{1-(\frac{1-b\epsilon}{1-\epsilon})^4} \right) \exp \frac{2c\epsilon\beta_0 w_0 \|g\|_\infty^2}{1-(\frac{1-b\epsilon}{1-\epsilon})^2}$$

5.1.3 The Hybrid Case

The hybrid case in which we consider producing piecewise constant approximations to an image by minimizing E presents difficulties that at present we do not know how to surmount except in the one-dimensional version of the problem. We can make some progress if we make further regularity assumptions on minimizing Γ. For example if we require that each subdomain satisfy the restricted cone property [1] then we can assert the existence of a complete orthonormal set of eigenvectors satisfying

Neumann boundary conditions for the operator Δ, on L_2 of that domain. If we then consider a single step of the algorithm we can see that the difficulty arises from the spectral dependence of the feedback. For example let Γ be a minimizer of E. We will concentrate on a fixed subdomain of $\Omega\backslash\Gamma$ which we denote A. Assume there exists a complete set of orthonormal eigenvectors satisfying Neumann boundary conditions associated with the elliptic operator Δ on $L_2(A)$. We will denote a generic real (nonpositive) eigenvalue of this operator by λ_i and it's associated eigenvector by h_i. We then have the following unique representations of f and g in $L_2(A)$,

$$\begin{aligned} f &= \sum_{i=1}^{\infty} c_i^f h_i \\ g &= \sum_{i=1}^{\infty} c_i^g h_i \end{aligned}$$

where as a consequence of the Euler–Lagrange equations we have,

$$c_i^f = \frac{\beta}{\beta - \lambda_i} c_i^g.$$

It is unclear at this point what the schedule for the parameters β and α should be. The limit theorem for this case does not dictate the relation between α and β except that the ratio $\frac{\beta}{\alpha}$ should tend to infinity to obtain accurate localization of the boundaries. Any algorithm based on this idea should be designed to encourage a piecewise constant limit and therefore it might be suggested that one should let β tend to zero. Consider updating the image as in the other cases;

$$g' = \begin{cases} (1 - a\epsilon)g + a\epsilon f & \text{for } x \in \Omega\backslash[\Gamma]_w \\ g & \text{for } x \in [\Gamma]_w \end{cases}$$

For the sake of explicitness, consider updating the parameters according to the schedule,

$$\begin{aligned} \alpha' &= (1 - \epsilon)^3 \alpha \\ \beta' &= (1 - \epsilon)\beta. \end{aligned}$$

As before in order to get a bound on the change in cost we consider leaving Γ unchanged. The difficulty in this case arises because we are effectively increasing the contribution of the smoothing term to the functional so we need to explicitly determine the reduction in the smoothing term after the update and not just consider the fidelity term, which was sufficient to handle the piecewise constant and piecewise smooth cases.

Define,

$$\hat{g} = (1 - a\epsilon)g + a\epsilon f$$

and consider as a candidate f' the solution to

$$\Delta f' = \beta'(f' - g)$$

with the Neumann boundary conditions. Now we have $\hat{g} = \sum_{i=1}^{\infty} c_i^{\hat{g}} h_i$ and,

$$c_i^{\hat{g}} = \left(1 + \epsilon a \frac{-\lambda}{\beta - \lambda}\right) c_i^g$$

and we thus get,

$$\beta' \int_A (f' - g)^2 + \int_A |\nabla f'|^2 = \sum_{i=0}^{\infty} \frac{-\lambda_i \beta'}{\beta' - \lambda_i} (1 + a\epsilon \frac{\lambda_i}{\beta - \lambda_i})^2 (c_i^g)^2.$$

From this expression we see how the reduction in cost can depend on the spectral gap associated with the operator Δ in relation to the size of β'. In one dimension stability of the algorithm (in the sense that the number of break points remains bounded) follows from the fact that this spectral analysis is rigorous and that the spectral gap for an interval of length l is bounded below by $\frac{\pi}{l}$. In two dimensions we do not know a-priori that a spectral gap of sufficient size exists for this method of proof to go through. We fail to get sufficient reduction in the cost for a stepwise argument like this one to work when most of the cost in the integral terms is associated with small values of $|\lambda_i|$ or equivalently, with the smoothing term. If we make some assumptions about the partition of the energy between the fidelity and the smoothing term then we can use the stepwise arguments we used earlier. Consider for example the situation when the two terms are equal. Leaving f unchanged we obtain a new cost of approximately $(1 - \frac{2a+1}{2})\epsilon$ times the earlier one. Thus for $a > \frac{5}{2}$ we would predict stability of the algorithm in the sense that the cost remains bounded. It appears that for the hybrid case more study is required to determine what the form of the update equations should be. A more sophisticated smoothing mechanism might be warranted.

5.2 A Computational Model

In this section we propose a more explicit realization of the paradigm. This realization is well suited only for the piecewise smooth case and perhaps also for the hybrid case. Our proposal is to use the Γ–convergent approximation to the variational formulation mentioned in the introduction as a means to generate both the boundary locations and their neighborhoods. Within this format we re–examine the stability issue addressed earlier.

The function v appearing in the Γ–convergent approximation is such that $1 - (1 - v^2)^n$ has the appearance of a smoothed neighborhood of the boundaries. The boundaries themselves can be identified with those locations where $(1 - v^2)^n \simeq 0$. A neighborhood of the boundaries can be defined simply as $\{x : 1 - (1 - v^2)^n > t\}$ with $t \simeq \frac{1}{2}$ for example. In Chapter 2 we reviewed some calculations presented in [5] to capture the essential form of minimal v. These calculations were necessary in order to prove the second condition required for the Γ–convergence which in our case is essentially,

$$\exists (f_n, v_n) \rightarrow (f, 0) \;\; \liminf_{n \rightarrow \infty} E^n(f_n, v_n) \leq E(f)$$

The nearly optimal v_n was constructed by setting it equal to 1 in a tiny neighborhood of $\mathcal{S}_f$ and then letting it decay as a function of distance from that neighborhood. In Chapter 2 we showed that by thresholding the nearly optimal $(1 - v^2)^n$ at a level t we obtain a set which we can interpret as a neighborhood of an optimal Γ of size w_t where,

$$\frac{n w_t}{2} \simeq \int_{-\ln \sqrt{t}}^{\infty} \frac{\exp -r}{r} \, dr$$

This means that we can define a neighborhood of Γ to be $\{x : (1 - v^2)^n > t\}$ and that this should approximate $[\Gamma]_{w_t}$ where $w_t \simeq \frac{2}{n} \int_{-\ln \sqrt{t}}^{\infty} \frac{\exp -r}{r} \, dr$. The parameter w discussed in the previous section which represents the width of the regions in which we leave our data unchanged can now be replaced by n and t. For our simulations we set $t = \frac{1}{2}$ and we then have the relation $w_t \simeq \frac{1.6}{n}$. Keeping t fixed also allows us to relate the neighborhood size directly to the cost of the boundary.

Earlier, within the formal description of the algorithm in the piecewise smooth case we used the expression 5.1.4 to relate the area of the neighborhood in which the image is not smoothed to the length of the boundary. The term $D\pi\epsilon^2$ was included because of the possibility of cracked tips. In the model we are now considering we can relate $|\{x : (1-v^2)^n > \frac{1}{2}\}|$ directly to the cost associated with the boundary. Since $(1-v^2)^n > \frac{1}{2} \Rightarrow n^2v^2 > n^2(1-(\frac{1}{2})^{\frac{1}{n}}) \simeq n \ln 2$ we get (approximately),

$$|\{x : (1-v^2)^n > \frac{1}{2}\}| < \frac{1}{n \ln 2} \int_\Omega \frac{n^2v^2}{16}$$

This result now allows us to treat the piecewise smooth case much as in the piecewise constant case. Assuming $w \propto \frac{1}{n}$ and $b > 1$ we have $\beta/n \to 0$ and because of this the stability analysis for the piecewise smooth case can proceed much as it did for the piecewise constant case. However, there is a different problem with this model; it is difficult to bound the change in the cost associated with the smoothing term and the boundary term when n is increased.

The procedure *P1* of our paradigm has now been reduced to the minimization of the following functional,

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

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,

$$\begin{aligned} \frac{\partial f}{\partial t} &= c_f(\nabla \cdot ((1-v^2)^n \nabla f) - \beta(f-g)) \\ \frac{\partial v}{\partial t} &= c_v(\alpha \nabla \cdot ((1-v^2)^n \nabla v) + n(|\nabla f|^2 + \alpha |\nabla v|^2)(1-v^2)^{n-1} v - \frac{\alpha n^2}{16} v) \end{aligned}$$

with Neumann boundary conditions. The parameters c_f and c_v are arbitrary positive constants which control the rate of descent. These equations resemble the non-linear filtering scheme of Perona and Malik [30] which was reviewed in Chapter 1. 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 f 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 as we shall see in Section 5.4 the dependence of the solution on the initial conditions, and the relative rate constant 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 behaves very much like the non-linear filtering scheme of Perona and Malik [30]. 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 5.1. Suppose we initialize the descent equations with f set to the solution of $\Delta f = \beta(f - g)$ and β 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 α 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.

5.3 Discretization

In this section we consider the problem of discretizing the functional associated with the variational formulation of the segmentation problem. We also present a particular discretization based on the model presented in the previous section and which was used to produce our simulation results.

Under what conditions can one say that a discrete formulation of a problem faithfully approximates the original continuous one? Clearly one should require that as the discretization becomes finer and finer the discrete problem should converge, in some sense to the continuous one. A reasonable formulation of the convergence issue in our case could be made in the framework of Γ–convergence for example. Many discrete versions of energy minimizing formulations of the segmentation problem have been proposed (see [7, 17, 26, 29]). However, it seems that these discrete versions do not properly approximate the continuous problem as the lattice spacing tends to zero. Thus these discrete formulations in the limit fail to capture properties of the continuous formulation (such as rotational invariance.)

Besides this convergence property there are other desirable properties discrete versions of the variational problem should have. In most situations we would like to represent both the boundaries and the approximating function on grids of the same scale. Also one would hope that all calculations can made locally, i.e. that the energy functional can be written as the sum of terms each of which depends only a finite set of values of the discretized problem which are localized in the two dimensional array. For the original variational problem it appears that both of these goals i.e. convergence of the discrete functional to the continuous one and a local computation model cannot be met for the length term with the usual representation of the boundary.

Consider for example a tiling of the plane with squares whose sides have length δ. Each square is identified by its center which we assume lies on δZ^2. We consider as an approximation of a curve in the plane the set squares it intersects. Thus as $\delta \to 0$ the approximation converges to the curve in Hausdorff metric. For every δ we have a map,

$$T_\delta : 2^{\Re^2} \longrightarrow 2^{Z^2}$$

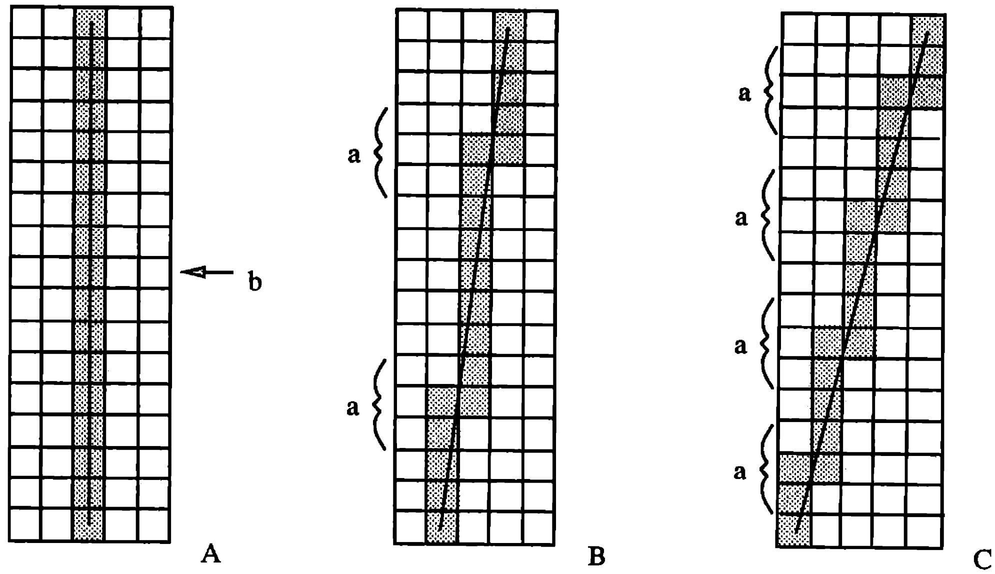
where $(i,j) \in T_\delta(A) \iff A \cap \{(x,y) : |x - \delta i| \leq 2\delta, |y - \delta j| \leq 2\delta\} \neq \emptyset$. To calculate an approximation to 'length' we consider translation invariant local functions on $Z^2 \times 2^{Z^2}$ of radius K. By this we mean functions, $F : Z^2 \times 2^{Z^2} \longrightarrow \Re$ such that $F(x,B) = F(x-y, B-y)$ for all $x,y \in Z^2$ and $B \in 2^{Z^2}$ (where we have interpreted B as a subset of Z^2). We are interested in computational models which are local of radius K. We express this be requiring $F(x,B) = F(x,A)$ whenever $B \cap \{y : \|y-x\|_\infty \leq K\}$ $= A \cap \{y : \|y-x\|_\infty \leq K\}$. The estimate of the length of A is defined by the functional,

$$L^\delta(A) = \delta \sum_{x \in Z^2} F(x, T_\delta(A))$$

We claim that such functions can never asymptotically recover length. We will demonstrate this for $K = 1$ i.e. a nearest neighbor model. However, the proof technique will work for any K. Figure 5.3 illustrates the idea and for simplicity we will refer to it while sketching the proof. Define,

$$\begin{aligned} A &= \{x \in \Re^2 : -1 \leq x_2 \leq 1, x_1 = 0\} \\ B &= \{x \in \Re^2 : -1 \leq x_2 \leq 1, x_2 = 8x_1\} \\ C &= \{x \in \Re^2 : -1 \leq x_2 \leq 1, x_2 = 4x_1\} \end{aligned}$$

Now it is not too difficult to see that if $\lim_{\delta \to 0} L^\delta(A) = 2$ then the sum of the $F(x, T_\delta(A))$ over the x in a row like 'b' in Figure 5.3 must equal 1. If it is also true that $\lim_{\delta \to 0} L^\delta(B) = 2\sqrt{1 + \frac{1}{64}}$ then the sum of the $F(x, T_\delta(A))$ over the x in three rows such as 'a' indicated in Figure 5.3 must equal $\sqrt{8^2+1} - 5$ since almost all of the other rows must contribute the same as row 'b'. (The only remaining significant rows are those near the end of the line but it can be easily seen that asymptotically



If $\lim_{\delta\to 0} L^\delta(A) = 2$ then each row 'b' contributes δ

If $\lim_{\delta\to 0} L^\delta(B) = 2\sqrt{1+\frac{1}{64}}$ also converges the rows 'a' must contribute $\delta(\sqrt{8^2+1}-5)$

If $L^\delta(A)$ and $L^\delta(B)$ converge then $L^\delta(C)$ converges to $\frac{1}{2}(\sqrt{8^2+1}-4)$

Figure 5.3: Estimates of Length

the contribution from these rows must be negligible.) These two results imply that $\lim_{\delta\to\infty} L^\delta(C) = \frac{1}{2}(\sqrt{8^2+1}-4)$ which is not the length of C. It is obvious that this argument can be applied for any K. Thus there does not exist a local functional of the form we have considered which can asymptotically recover length, even for straight lines.

In [23] it was shown that a Γ–convergent discrete approximation of the length term is possible by using a discrete approximation of Minkowski content. For a fixed lattice spacing of δ one can approximate $\mathcal{M}^1_\epsilon$ of a curve γ by counting the number of lattice sites within distance ϵ of the discrete approximation to γ. Now for this approximation to be accurate it is necessary that $\epsilon \gg \delta$. If the quantity $\frac{\epsilon}{\delta}$ is kept fixed while δ decreases to zero then the result discussed above implies this approximation cannot converge to length even for straight lines. It was shown in [23] that if $\frac{\epsilon}{\delta} \to \infty$ while $\epsilon \to 0$ then the discrete approximation to Minkowski content mentioned above will actually Γ–converge to Minkowski content on the space of closed sets having finitely many connected components.

We propose to discretize by finite elements the function v in the Γ-convergent approximation mentioned in Section 3.3. Because, roughly speaking, the boundaries are spread out over a region whose width is proportional to $\frac{1}{n}$ we conjecture that if δ is the lattice spacing it is necessary and sufficient that as $n \to \infty$ that $\delta \to 0$ such that $n\delta \to 0$ for faithful recovery of the length term.

We return now to the consideration of a more realistic situation, a fixed lattice. Consider the behavior of v in the Γ–convergent approximation for large n. For each point i in the array there is a term in the cost proportional to $n^2 v_i^2$. 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, 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 $(1 - v^2)^n$ results in an increase in the smoothness of f i.e. a reduction in $|\nabla f|$. This causes yet a further increase in $(1 - v^2)^n$. Thus we see that the discretized version of this approximation becomes unreliable for large n.

For the simulations presented in this thesis we discretized by finite elements f, g and v in a manner described below. We define the discrete version of f and g on a single lattice 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 5.4 indicates the assignment of variables to lattice values. For convenience we label the variables as in a matrix. 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 we partition the variables associated with v into two sets, vh and vv, corresponding in some sense to horizontal and vertical edge elements. The assignments are as in Figure 5.4. A suitable discrete version of E in terms of these variables is the following,

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

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

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 $vv_{i,j}$.

We can derive the discrete form of the Euler–Lagrange equations for this system by differentiating the expression above with respect to the various elements. In particular we get,

$$
\begin{aligned}
\frac{\partial}{\partial f_{i,j}}E &= 2\delta^2\beta(f_{i,j}-g_{i,j}) \\
&\quad +2(f_{i,j}-f_{i+1,j})(1-vv_{i,j}^2)^n \\
&\quad +2(f_{i,j}-f_{i-1,j})(1-vv_{i-1,j}^2)^n \\
&\quad +2(f_{i,j}-f_{i,j+1})(1-vh_{i,j}^2)^n \\
&\quad +2(f_{i,j}-f_{i,j-1})(1-vh_{i,j-1}^2)^n
\end{aligned}
$$

and

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

and similarly,

$$
\frac{\partial}{\partial vh_{i,j}}E = -2\,n\,vh_{i,j}(1-vh_{i,j}^2)^{n-1}\,(f_{i,j}-f_{i+1,j})^2
$$

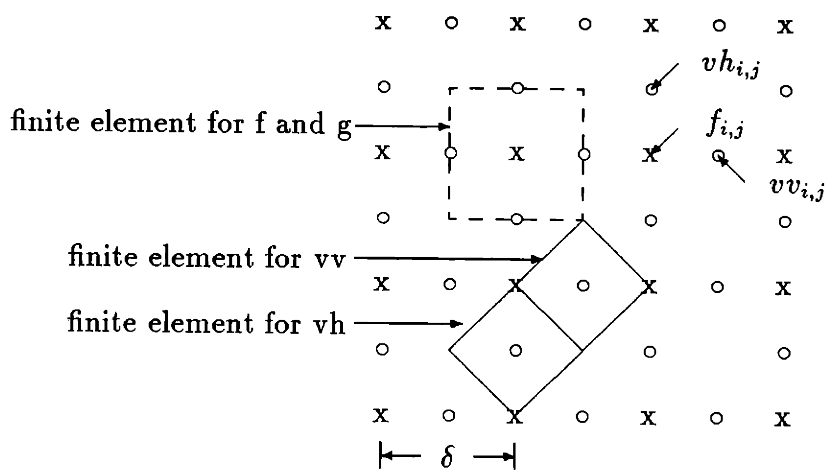


Figure 5.4: Lattice Variables and Finite Elements

$$
\begin{aligned}
&-\alpha n\, vh_{i,j}\,(1 - vh_{i,j}^2)^{n-1} \sum_{(i',j')\in\mathcal{N}_h(i,j)} (vh_{i,j} - vv_{i',j'})^2 \\
&+\alpha \sum_{(i',j')\in\mathcal{N}_h(i,j)} (1 - vh_{i,j}^2)^n (vh_{i,j} - vv_{i',j'}) \\
&+\alpha \sum_{(i',j')\in\mathcal{N}_v(i,j)} (1 - vv_{i',j'}^2)^n (vh_{i,j} - vv_{i',j'}) \\
&+\alpha \frac{\delta^2 n^2}{16} vh_{i,j}.
\end{aligned}
$$

Normally the domain Ω will be rectangular in shape. For those elements on the boundary of the domain we can evaluate the expressions above by assuming that Neumann boundary conditions are satisfied.

5.4 Computation and Simulations

In the preceding section we have set up a discrete model for the Γ–convergent approximation to the variational formulation of the segmentation problem. A discrete computation based on the parabolic equations associated with this formulation takes the following form,

$$
f_{i,j}^{t+1} - f_{i,j}^{t} \quad = \quad -c_f \frac{\partial}{\partial f_{i,j}} E \tag{5.4.1}
$$

$$vv_{i,j}^{t+1} - vv_{i,j}^{t} = -c_v \frac{\partial}{\partial vv_{i,j}} E \tag{5.4.2}$$

$$vh_{i,j}^{t+1} - vh_{i,j}^{t} = -c_v \frac{\partial}{\partial vh_{i,j}} E \tag{5.4.3}$$

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 assignment,

$$\begin{aligned} c_f &= 2\left(\delta^2\beta + (1 - vv_{i,j}^2)^n + (1 - vv_{i-1,j}^2)^n + \right. \\ &\quad \left. (1 - vh_{i,j}^2)^n + (1 - vh_{i,j-1}^2)^n\right) \end{aligned}$$

in keeping with standard relaxation algorithms.

An important consideration concerning the choice of c_v is the dependence it should have on n. As was pointed out in the preceding section the relative rates of the evolution of f and v can effect the solution. Also in our algorithm we vary n both to sharpen the boundaries and to reduce the size of the neighborhood which controls the smoothing of the image. Thus as much as possible we would like to be able to produce an n invariant version of the equations in the sense that the only variations in n we observe should be those predicted by the mathematical model and not merely artifacts of the way we perform the computation. For notational convenience in the examination of this issue we will dispense with the distinguishing between vh and vv and will denote a generic edge element simply by $v_{i,j}$.

The function $(1 - v^2)^n$ indicates the presence or absence of edges. This function makes a transition from approximately 1 to approximately 0 in the neighborhood of an edge. Thus it is reasonable to expect that most of the computation relevant to the decision of where to place boundaries will occur with $(1 - v^2)$ taking values in a range of (0,1) which will be largely independent of n. Consider the first two terms of $\frac{\partial}{\partial v_{i,j}} E$. Each term has as a factor the function $n v_{i,j}(1 - v_{i,j}^2)^{n-1}$. As $(1 - v_{i,j}^2)^n$ tends to either 0 or 1 this factor tends to 0. The maximum of this factor occurs at approximately $(1 - v_{i,j}^2)^n = e^{-\frac{1}{2}}$. The the other factor in the first term is proportional to $\|\nabla f\|^2$. Clearly this term, being always negative, is promoting the appearance of edges. The second term, also negative, has a factor proportional to $\|\nabla v\|^2$. Now $\|\nabla v\|^2$ will be large in the neighborhood of the edges. This term therefore has a positive feedback effect favoring the creation of edges. Consider initializing the descent with $(1-v^2)^n = 0.5$. A locally large value of $\|\nabla f\|^2$ will, as a consequence of the first term tend to increase $\|\nabla v\|^2$ locally. The second term will then promote even further the local increase in v. The third and fourth terms of $\frac{\partial}{\partial v_{i,j}}E$ are smoothing terms, i.e. they tend to smooth v. The fifth term tends to reduce v in proportion to its value and hence also has a smoothing effect. Thus the last three terms favor the elimination of edges while the first two promote their creation. Thus we see that our expectation was correct; the value of $(1-v^2)^n$ at which the descent equations is most sensitive to the potential creation of boundaries occurs in the environ of $(1-v_{i,j}^2)^n = e^{-\frac{1}{2}}$. Thus to determine how c_v should be scaled with n we consider $(1-v_{i,j}^2)^n \propto 1$.

If we consider $(1-v_{i,j}^2)^n \propto 1$ it follows that $v_{i,j} \propto \frac{1}{\sqrt{n}}$. Thus the fourth term in $\frac{\partial}{\partial v_{i,j}}E$ i.e.,

$$\alpha \frac{\delta^2 n^2}{16} v_{i,j}$$

varies in proportion to $n^{1.5}$. Now consider the second term which has the form,

$$-\alpha n\, v_{i,j}\,(1-v_{i,j}^2)^{n-1} \sum_{(i',j')\in\mathcal{N}(i,j)} (v_{i,j}-v_{i',j'})^2.$$

We know that the width of the boundaries decays like $\frac{1}{n}$. This means that if x and y are neighboring sites on one side of and near a boundary (i.e. at these sites $(1-v^2)^n$ is bounded away from 0 and 1), that $(1-v_x^2)^n - (1-v_y^2)^n \propto n$. To a first order approximation this implies $v_x^2 - v_y^2 \propto 1$ as n varies. Since $v_x + v_y$ will vary in proportion to $\frac{1}{\sqrt{n}}$ it follows that $v_x - v_y \propto \sqrt{n}$. Thus the term quoted above also varies in proportion to $n^{1.5}$. We expect that the third and fourth terms i.e.,

$$\begin{aligned} &\alpha \sum_{(i',j')\in\mathcal{N}(i,j)} (1-v_{i,j}^2)^n (v_{i,j}-v_{i',j'}) \\ + \;&\alpha \sum_{(i',j')\in\mathcal{N}(i,j)} (1-v_{i',j'}^2)^n (v_{i,j}-v_{i',j'}) \end{aligned}$$

which is the discrete representation of $(1-v^2)^n\Delta v$, will also vary in proportion to $n^{1.5}$ in the neighborhood of the edges. In any case we have clearly seen that in the

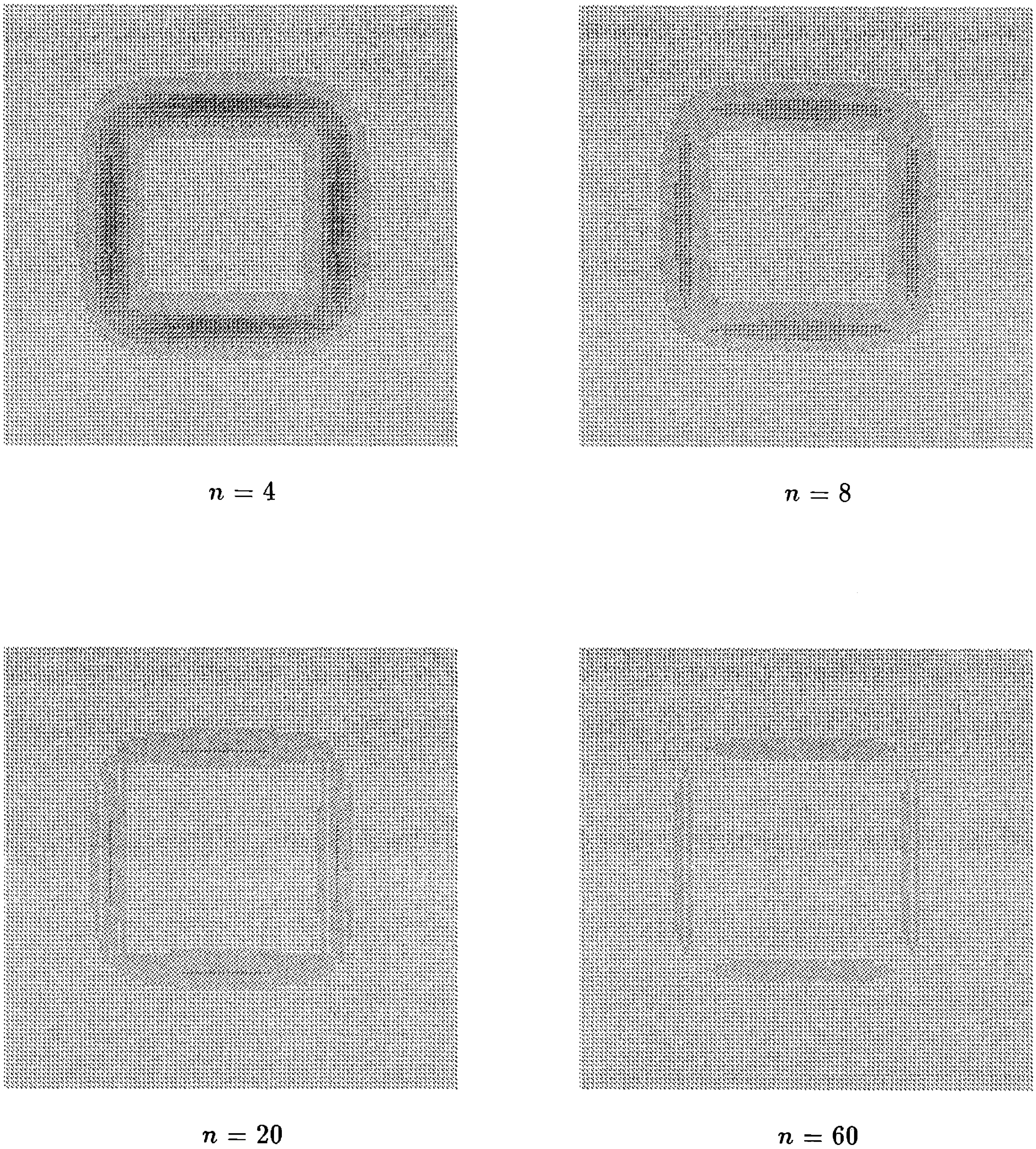
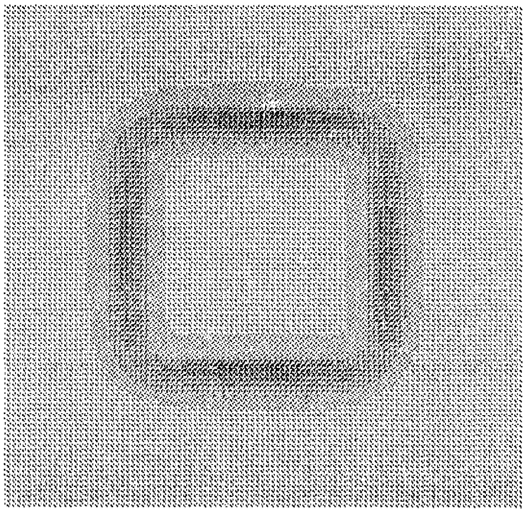
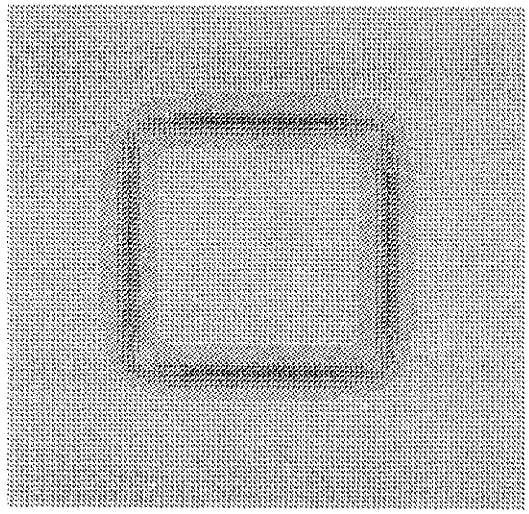


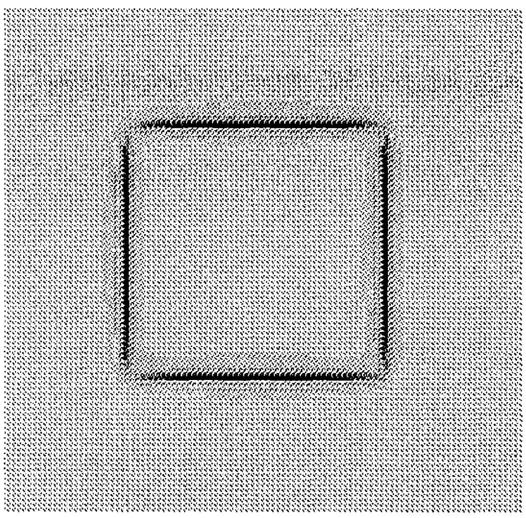
Figure 5.5: Simulation Dependence on n



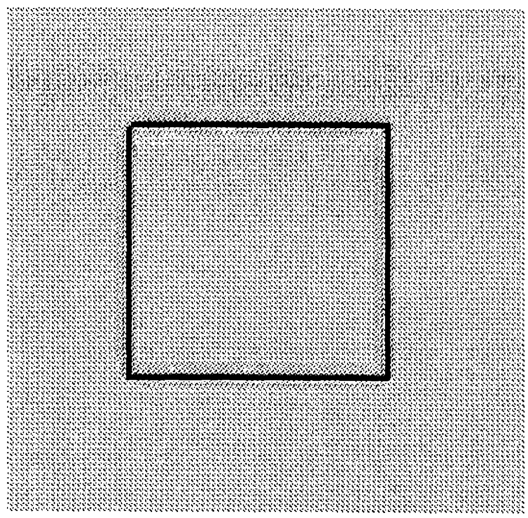
$n = 4, \beta = 2$



$n = 8, \beta = 4$



$n = 20, \beta = 10$



$n = 60, \beta = 30$

Figure 5.6: Simulation Dependence on n and β

most volatile part of the descent, when boundaries are being placed, the magnitude of $\frac{\partial}{\partial v_{i,j}}E$ will be proportional to $n^{1.5}$. For our simulations we have therefore let the stepsize c_v vary like $n^{-1.5}$.

Of course the first term in $\frac{\partial}{\partial v_{i,j}}E$ namely,

$$n\, v_{i,j}\,(1 - v_{i,j}^2)^{n-1}(f_{i,j} - f_{i',j'})^2$$

will vary like $\sqrt{n}$. In general then the early stages of the descent which should be dominated by this term become gradually slower if we scale c_v like $n^{-1.5}$. In our simulations we have found it useful to do presmoothing by iterating on the update equations for f before initiating those for v. This provides for the smoothing out of tiny objects in g. The alternative, i.e. no presmoothing tends to make the descent more sensitive to the height of discontinuities and less sensitive to the geometry of the object to which they are associated. In other words presmoothing, i.e. using a different initial condition on f, can effect which local minima of E^n will be found. Presmoothing will suppress the placing of boundaries at edges of objects having relatively small support. Now, in our algorithm β increases with n. Because of this the presmoothing becomes less and less severe. True discontinuities in g get smoothed out over a range approximately in proportion to $\frac{1}{\sqrt{\beta}}$. Thus the initial relevant values of $||\nabla f||^2$ will be proportional to β and hence proportional to n. Thus with our rescaling of β with n we can obtain a descent in which all the important terms of $\frac{\partial}{\partial v_{i,j}}E$ important to the early, decisive stages of the computation vary like $n^{1.5}$. In Figure 5.5 we show the results of a simulation on the image of a square carried in the same manner as the 'Lenna' simulations (see Figure 5.7) for various values of n. We see that increasing n tends slightly to reduce the sensitivity to an edge. In Figure 5.6 we have repeated the same simulation except that in this case we have increased β in proportion to n. We see that the increase in β dominates the effect of increasing n and we recover more and more of the boundary with increasing n and β.

We have argued that near edges $v_{i,j} - v_{i',j'}$ varies as $\sqrt{n}$. Obviously, however, in the discrete problem there is a strict upper bound of 1 on this difference. Again we see that the discrete formulation becomes unreliable for large n.

5.4.1 Simulation Results

We have simulated the algorithm developed in this chapter on the image shown in Figure 5.8. The size of the image is 230×216 pixels. The image was taken from a bitmap of 920×864 bits and each 4×4 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 4×4 block is scaled so that the range of g lies within $[0, 2]$. 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 difficult one because many of the edges are blurred and there are regions of texture. The flow chart in Figure 5.7 shows the details of the computations. We have performed the simulation for several scales. For one scale which we denoted 'Scale 3' we have sampled the functions, f, g and $(1 - v^2)^n$ 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. Even though β is increased by a factor of almost 200 the particular boundaries which are found do not change except in the fine detail of the localization. We also present the final results of the algorithm for different scales. In figure 5.11 we display solutions obtained for several scales. Notice that the set of boundaries found is essentially monotonic in scale.

In Figure 5.12 we have plotted the evolution of the first and second terms of E as a function of β for Scale 3. We note that the first term (i.e. the term $\beta \int_\Omega (f - g)^2$) decreases while the second ($\beta \int_\Omega (1 - v^2)^n |\nabla f|^2$) increases with β. This effect is controlled somewhat by the constants a and b in the feedback equations. Smaller values of these constants should result in relatively less variation in these terms. We can report that the simulation of 'Scale 3' was repeated with a and b both set to 1 and that we observed the same trends in the two terms as before although the rates of change as a function of β were smaller, as one would expect.

The second set of simulation results were made on the synthetic image of one square occluded by another. The algorithm used was of the hybrid type. The algorithm segmented out the larger square removing the smaller one. This occurred even though the edge height of the smaller square i.e. the intensity of the smaller square, was larger then the difference between the intensity of the larger square and that of the smaller. The displays in this case are three dimensional mesh plots. The computation was performed somewhat differently for this simulation. There was no pre-smoothing, we simply alternated between the update equations for f and those for v. Also, we initialized each stage by setting $(1 - v^2) = 0.1$.

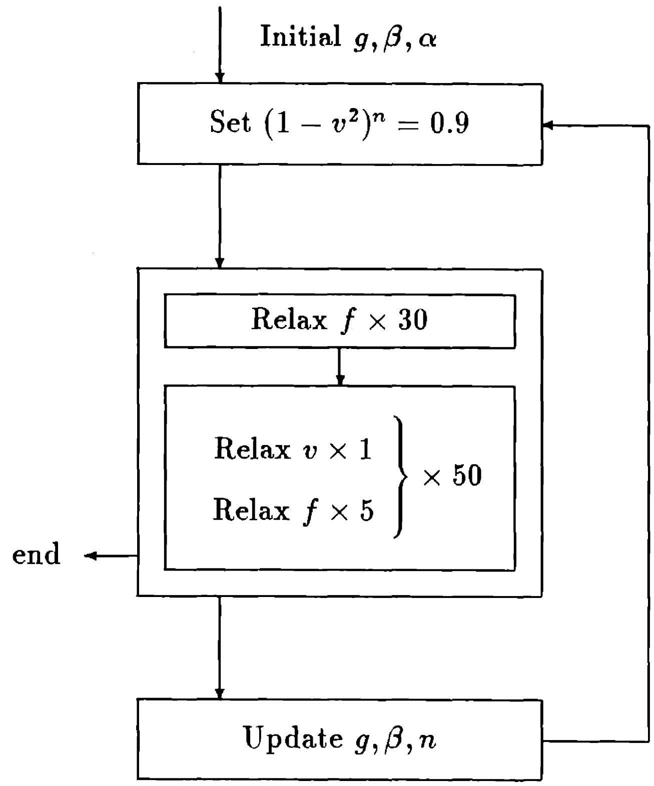


Figure 5.7: 'Lenna' Simulation Flowchart

δ	0.05
Range of g	$[0.0, 2.0]$
Stepsize c_v	$\frac{0.3}{\alpha n^{1.5}}$
ϵ	0.3
Gain a	1.2
Gain b	1.3

Table 5.1: General Parameters for 'Lenna' Simulations

'Scale'	α	Initial β	Final β	Initial n	Final n
1	0.01	4.0	84.1	3.0	117.3
2	0.0075	7.0	147.2	4.0	156.6
3	0.005	10.0	210.3	5.0	195.6
4	0.005	40.0	841.2	10.0	391.1

Table 5.2: Parameters for Simulation 'Lenna: All Scales'

Sample Number	α	β	n
1	0.005	10.0	5.0
2	0.005	18.4	10.4
3	0.005	62.2	45.1
4	0.005	210.3	195.6

Table 5.3: Parameters for Simulation 'Lenna: Scale 3'

Figure 5.8: “Lenna” Data Used for Simulations

Sample 1

Sample 2

Sample 3

Sample 4

Figure 5.9: Lenna Scale 3: Samples of f

Sample 1

Sample 2

Sample 3

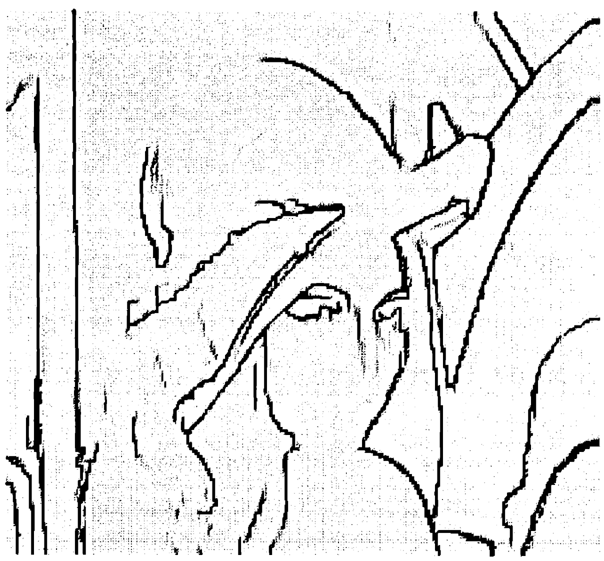
Sample 4

Figure 5.10: Lenna Scale 3: Samples of Updated g

Sample 1

Sample 2

Sample 3



Sample 4

Figure 5.11: Lenna Scale 3: Samples of $1 - (1 - v^2)^n$

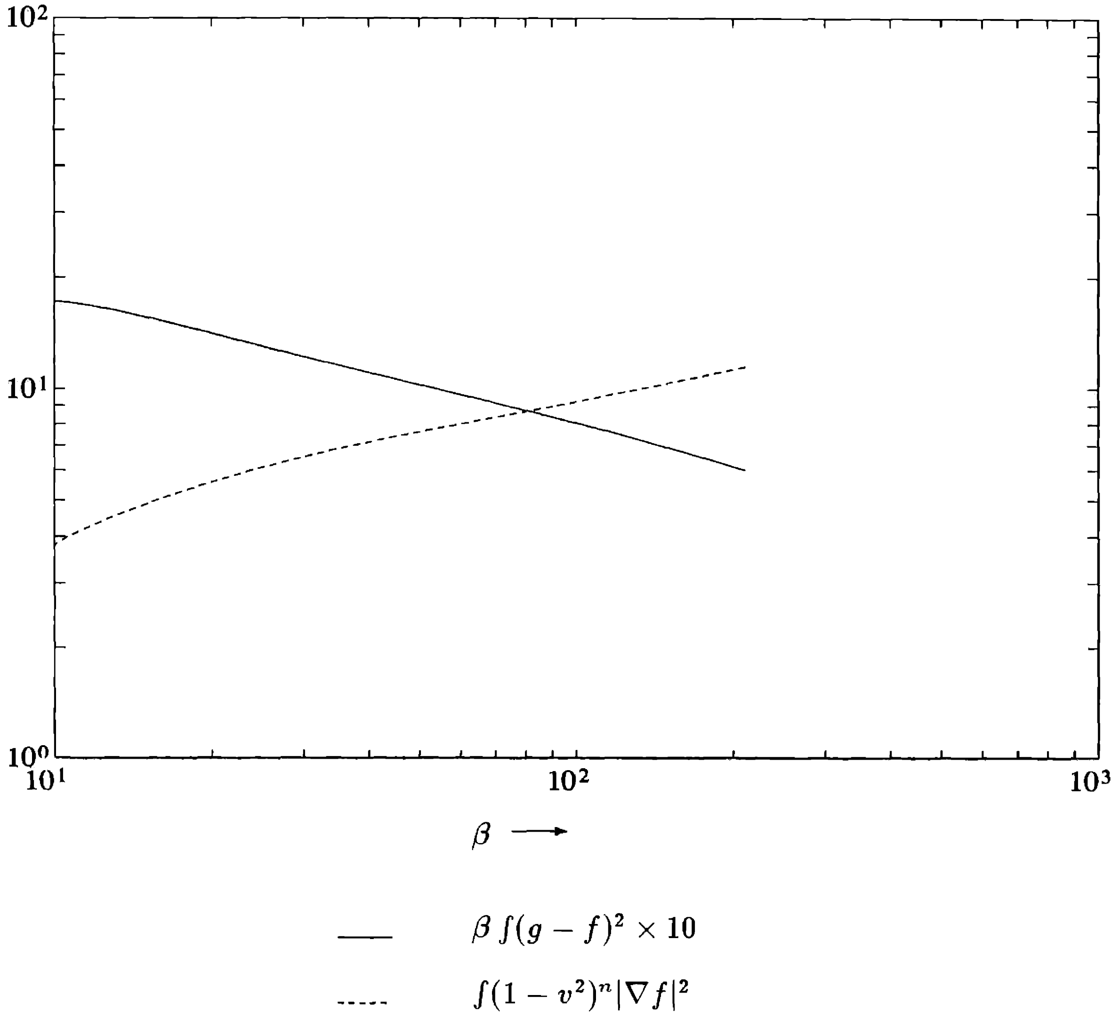
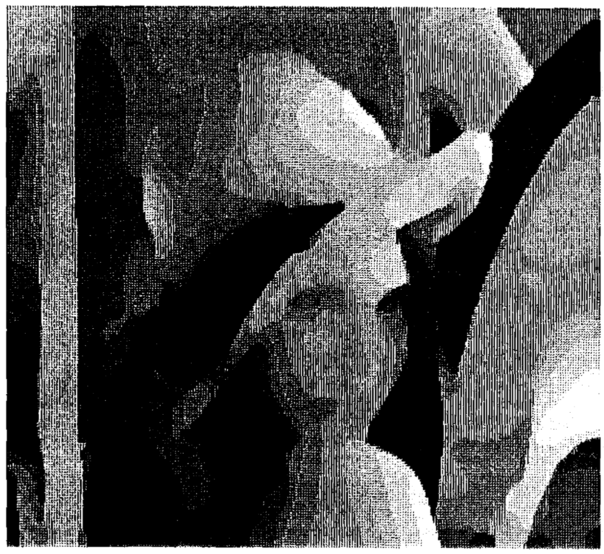


Figure 5.12: Evolution of Energy Terms for 'Lenna', Scale 3

δ	0.12
Range of g	[0.0,2.0]
Stepsize c_v	0.1
ϵ	0.02
Gain a	1.0
Gain b	1.0

Table 5.4: General Data for Occluded Square Simulations

Scale 1



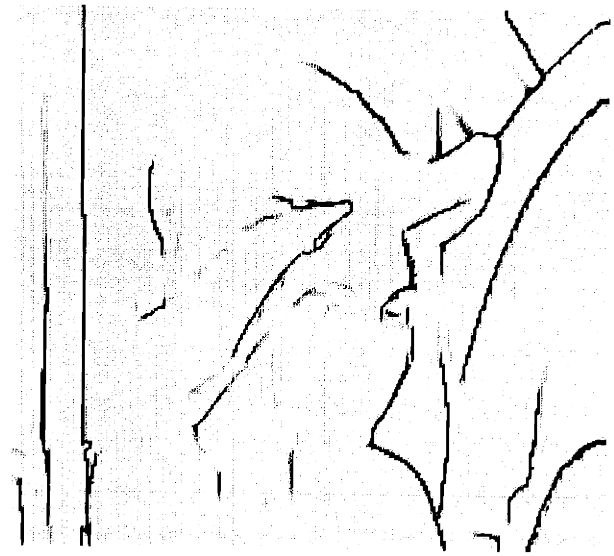
Scale 2

Scale 3

Scale 4

Figure 5.13: The Approximation f on Various Scales

Scale 1



Scale 2

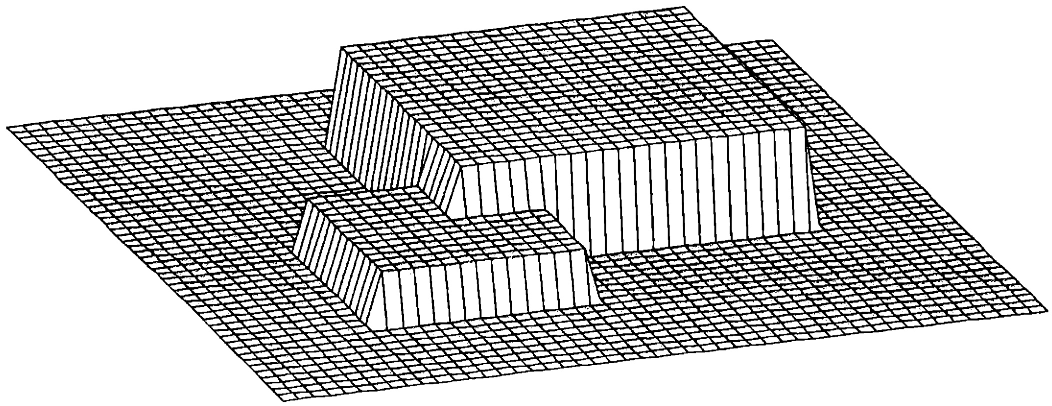
Scale 3

Scale 4

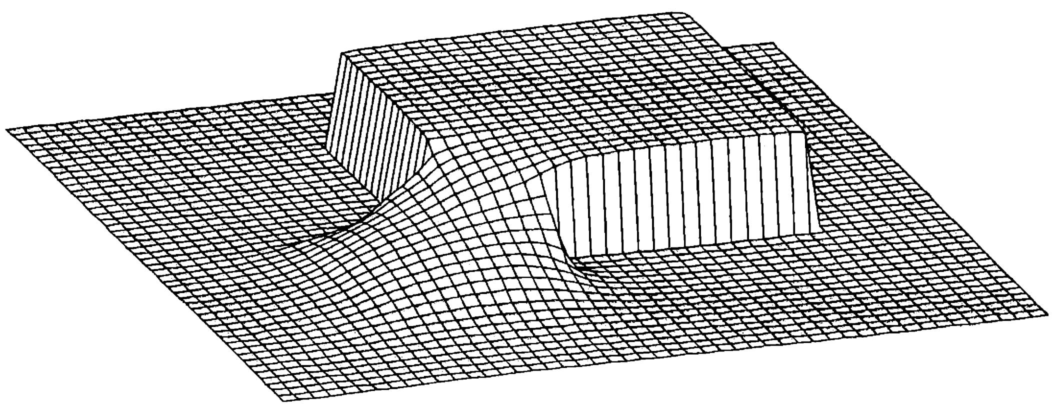
Figure 5.14: 'Lenna' Boundaries Various Scales

Sample Number	α	β	n
1	3	1.0	2.5
2	1.45	0.78	4.0
3	0.20	0.40	15.4
4	0.03	0.22	51.8

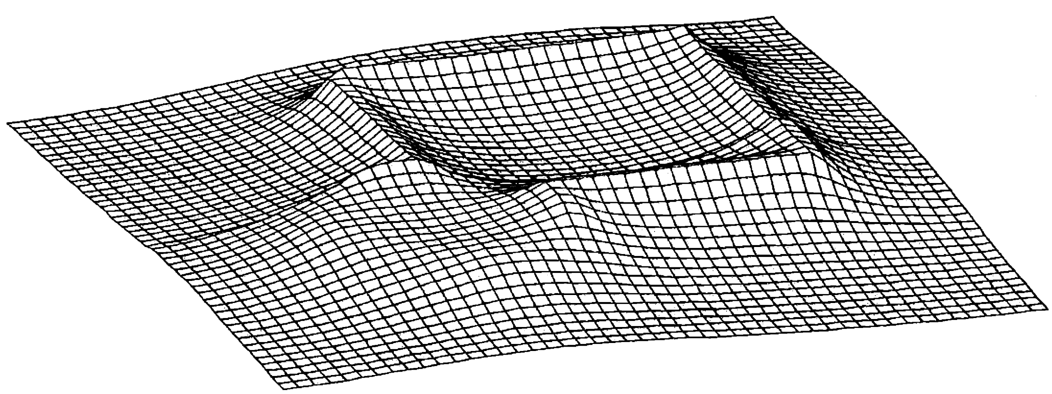
Table 5.5: Parameters for Simulation Occluded Square



Data g

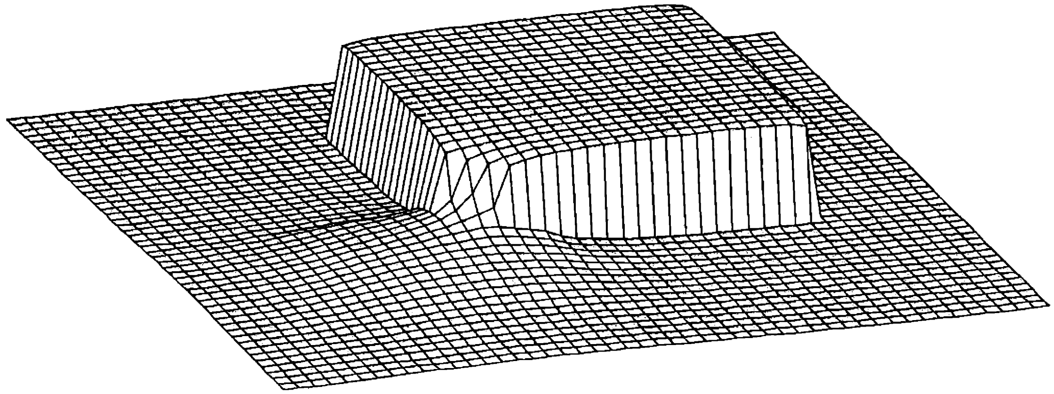


Sample 1: f

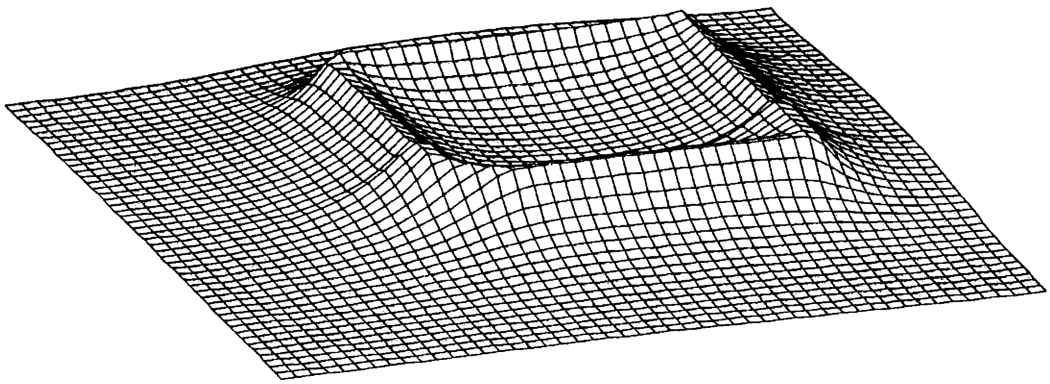


Sample 1: $1-(1-v^2)^n$

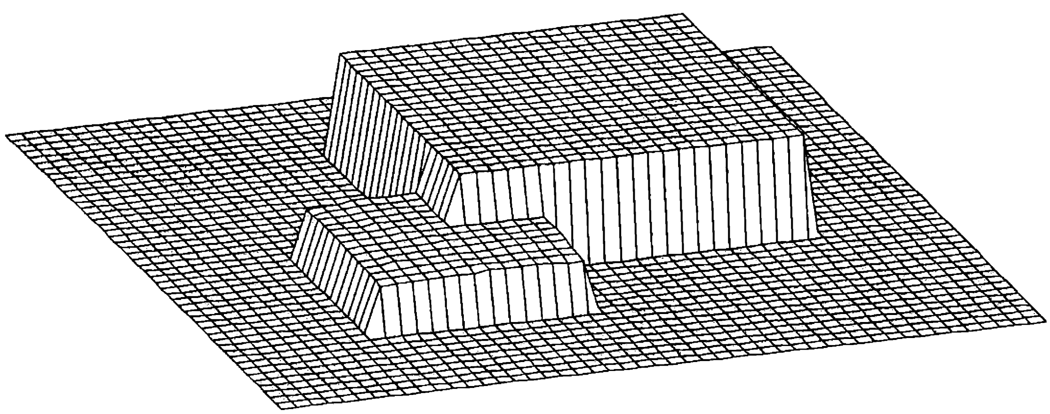
Figure 5.15: Occlude Simulations 1



Sample 2: f

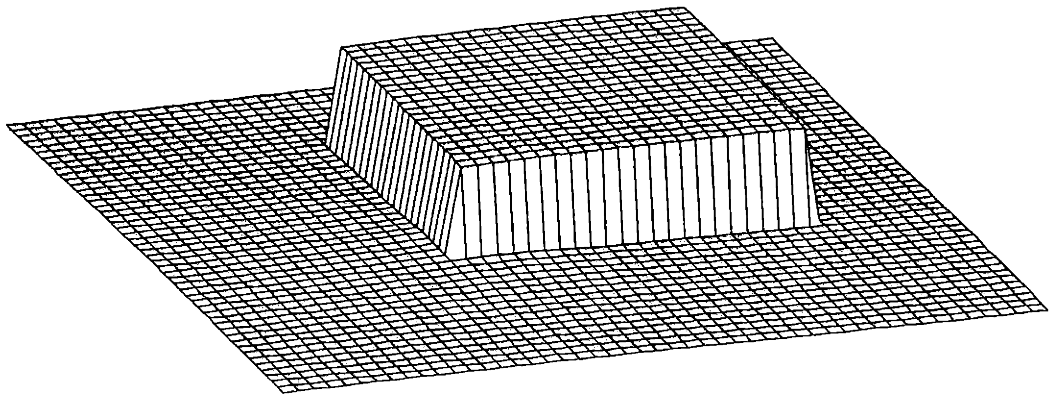


Sample 2: $1-(1-v^2)^n$

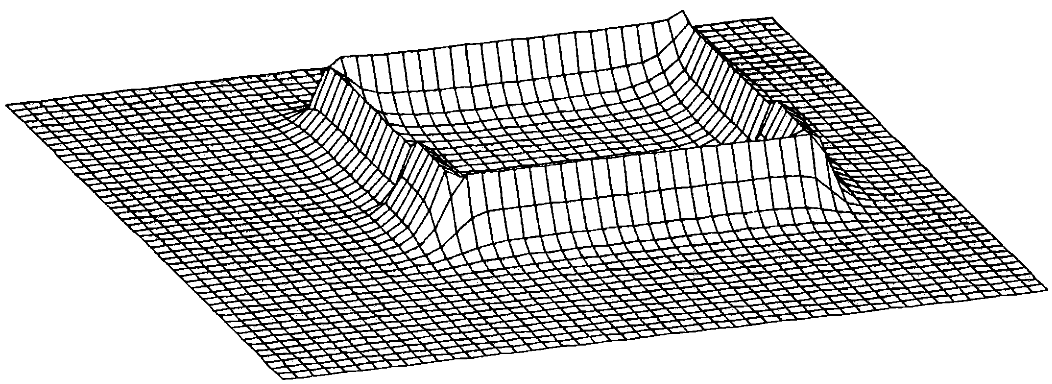


Sample 2: Updated g

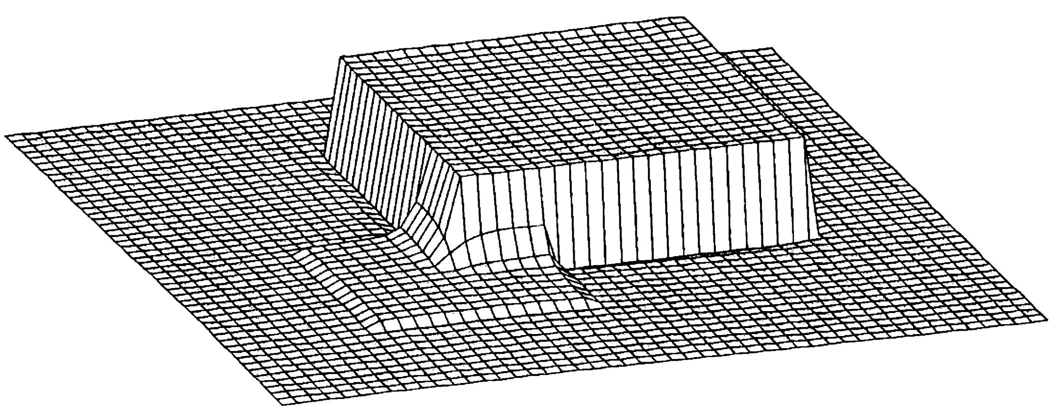
Figure 5.16: Occlude Simulations 2



Sample 3: f

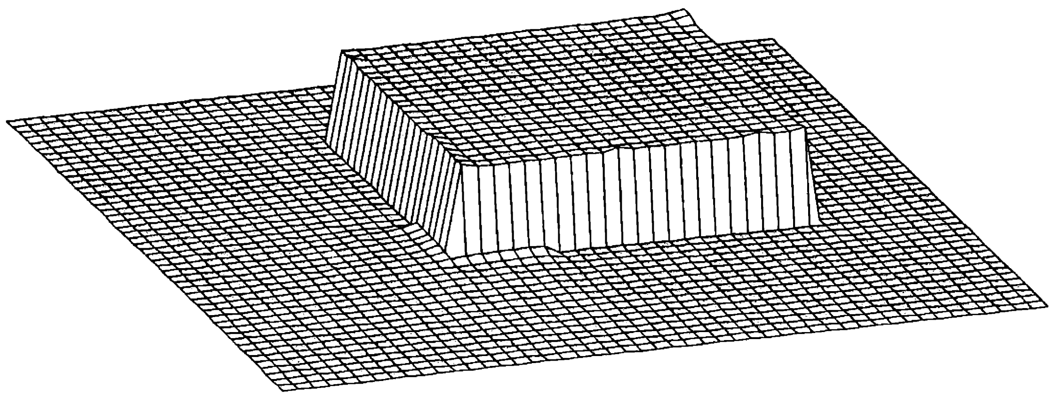


Sample 3: $1-(1-v^2)^n$



Sample 3: Updated g

Figure 5.17: Occlude Simulations 3



Sample 4: Updated g

Figure 5.18: Occlude Simulations 4

Chapter 6

Conclusions

6.1 Contributions

The topic of this thesis is the approximation of images, which we have considered to be real valued functions defined on a two dimensional domain, by piecewise smooth and piecewise constant functions. A basic goal pervading almost all of the results is to improve upon other techniques such as the Variational Approach, the Markov Random Field Approach and Non-linear Filtering Approach in the sense that the location of the boundaries or edges is provided on all scales with the same accuracy which is usually reserved only for the finest scales. It is in this sense that the segmentations we provide are "scale independent". Each of the methods mentioned above have parameters which under suitable conditions can be related to the 'scale' of the segmentation obtained. In each case there are certain ad hoc choices in the model which introduce 'errors' into the solutions i.e. the edges found will not track discontinuities in the data. Generally speaking these errors vary with the 'scale'. That is, as the scale for which the parameters are set tends to the microscopic the errors one observes in the location of the edges decreases. It is considered desirable to be able to obtain coarse scale approximations so that prominent features can be identified. Thus there is a trade off; larger scale implies larger errors. Our basic idea in this regard is to improve on these techniques by letting the 'scale' for which the parameters are set tend to the microscopic, forcing locally accurate boundary locations, while simultaneously altering the data presented to the technique to force the solution to remain on a coarse scale when viewed on a global level.

In the work presented in the thesis we have focussed on the variational method for basically two reasons. First, it is fundamental in the sense that the other techniques can be interpreted as effectively solving a problem similar to it. Second, because within the framework afforded us by the many rich mathematical results which are available for the variational approach we were able to develop a coherent basis for the algorithm which implements the scale independent segmentation.

The variational method consists in minimizing functionals of the form,

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

where g represents the image, f a piecewise smooth approximation and Γ the set of edges found in the image. For a fixed α 'scale' can be equated to $\frac{1}{\sqrt{\beta}}$. The central idea of the algorithm is to gradually increase β while smoothing the data g outside of some neighborhood of the boundaries which are found by minimizing E. This smoothing is effected by replacing g with some pointwise convex combination of f and g. In effect we introduce feedback from f into g in the variational formulation along with a parameter schedule.

To motivate and govern the algorithm mentioned above we have proved a theorem which characterizes what relation the solutions f, Γ will bear to the data g asymptotically as $\beta \to \infty$. Essentially we have shown that is g is a piecewise smooth function (piecewise $C^{0,1}$) then as $\beta \to \infty$ the optimal boundaries Γ_β will converge with respect to the Hausdorff metric to the support set of the discontinuities in g, and $\sqrt{\beta}(f-g)$ will converge to zero in $L^2(\Omega)$. Furthermore we have characterized the quantity of noise and smearing which can be permitted to corrupt the image yet still have the limit theorem hold. That is, we rescale the admissible noise and smearing as a function of β in such a way that the limit theorems still hold true, and the scaling we use appears to be optimal (i.e. the least conservative possible). The limit theorems have also been established for the piecewise constant version of the variational formulation of the segmentation problem. In deciding the amount of smoothing we must do and how fast the neighborhood size must decrease we have drawn an analogy with the limit theorem. The smoothing has the same role in the algorithm as the rescaling of the noise did in the limit theorems. We treat the lack of smoothing in the neighborhood of the boundaries as a smearing effect. The limit theorems then dictate how much smoothing is required and how fast the neighborhood size should decay in order to effect convergence of the boundaries to the 'correct' location without introducing spurious (i.e. finer scale) boundaries. The proofs of the limit theorems can be found in Chapter 4. The algorithm is developed in Chapter 5.

An actual implementation of the algorithm requires the minimization of E. Most practical computational methods will not be able to find global minimizers of E, but they will try to find "good" local minima. The algorithm can still be developed in such a framework as long as the computational procedure is kept consistent so that as the parameters are changed essentially the same local minima is found. Since we have implemented the smoothing by taking a convex combination of g and f we tend to make the data 'favor' even more the minima f. This helps to make the procedure more robust.

There is a basic principle which can be gleaned from this work. The original model which was conceived for segmenting images was the class of piecewise smooth functions. The variational formulation provides an ad hoc model for selecting such a function to approximate a given function. In Chapter 3 we reviewed some results in [29] attained using the calculus of variations which describe constraints on the structure of the solutions to the variational problem. In particular the geometry of the boundaries of minimal solutions is restricted to have only certain structures. For example, curves never exhibit sharp corners; at most three curves meet at a point and they will do so only such that their tangents make angles of 120° with each other. Thus certain desirable structures such as t–junctions and corners tend to be distorted. Now, the calculus of variations results are obtained by considering local changes in a potential solution. It is believed that when viewed globally the solutions will be "good". The limit theorems provide a means of making this assertion formally. The fact the boundaries converge to the discontinuity set of the image means that asymptotically essentially any boundary geometry can be recovered (excluding fractal sets, i.e. sets of infinite length). The particular topology i.e. that induced by the

Hausdorff metric, in which this convergence occurs is important. The algorithm tries to exploit the fact that the distortions in the boundaries are local effects with respect to this topology and that the Hausdorff distance between the "true" boundaries and those provided by solving the variational problem decays to zero in the limit $\beta \to \infty$. Thus the algorithm tries to circumvent the artifacts of the ad hoc model used to produce piecewise smooth approximations by taking the path pointed out by the limit theorems.

In this thesis we have also addressed the issue of computation. To (locally) minimize E we propose using a slightly different model. In [5] it was shown that there exists a sequence of functionals which converge to E in the sense of Γ–convergence. We have implemented the algorithm by employing a sequence of functionals In this approximation the usual representation of the boundaries as a one dimensional subset of the domain of definition of the image is replace by a function defined on the entire domain of definition. In our simulations we have locally minimized the approximating functional using local coordinate relaxation on both the function which approximates the image and that which approximates the boundaries. The approximations are parametrized by a real number n. The function $1-(1-v^2)^n$ has the appearance of a smoothed characteristic function of a neighborhood of the boundaries, which can be identified with the set where $(1-v^2) \simeq 0$. We have shown that the neighborhood size varies in proportion to $\frac{1}{n}$. By using the approximating functional we thus automatically determine a neighborhood of the boundaries simply by looking at a level set of $(1-v^2)^n$. Furthermore this computational model has the advantage that the boundaries are represented as a function and can be computed by a finite element approximation. The details concerning our implementation of this scheme and some simulation results are presented in Chapter 5. Imbedding the algorithm in the sequence of approximations provides an elegant and natural computational framework for doing scale independent segmentation.

Some of the fundamental questions associated with the variational formulation of the segmentation problem are still open. The most important of these concerns the existence of minimizers to E with boundaries consisting of a finite set of regular (e.g.

$C^{1,1}$) curves. The question of existence of minimizers to "weak" formulations of the problem have been essentially settled. Our contribution is an existence result which can also be applied to other variational problems such as the "weak plate" version of the segmentation problem. This result depends however, on the number of connected components of the boundary being finite. The proof of the existence theorem is given in Chapter 3. In the same chapter we have reviewed some existence results which appeared subsequently due to De Giorgi-Carriero-Leaci [10] which employed the SBV framework, which had been developed largely by Ambrosio [4]. It is within this framework that the piecewise smooth verion of the limit theorem was proved.

6.2 Further Work

Further work suggested by this thesis can be categorized according to how distant from the contents of the thesis the suggestions are. There are ideas at both extremes of this spectrum. We will organize them by starting closer to home and lifting our sights gradually further.

It is clear that the computational scheme developed here has considerable potential. It is quite versatile in the sense that varying initial conditions and the relative rates of evolution of the function f and the function v can effect quite different solutions and hence extract quite different information. There has been considerable interest in the non–linear filtering approach to segmentation. We have pointed out the connections between the non–linear filtering approach and the solving of the Γ–convergent approximation by local relaxation. This connection could be exploited to the further understanding of both. Perhaps variations in the form of the equations for v, for example, could produce decision variables of another type i.e. a detector for things other than the existence or non–existence of an edge. Another possibility is the detection of regions of texture. When a textured region is smoothed out the detail is lost. However a textured regions possesses many edges on a small scale. If we implement our algorithm with pre–smoothing as was done in most of the simulations then we would not detect the small objects which comprise a texture until we operated on a very small scale. However with no presmoothing we might expect that a textured region would produce a large amount of boundary at a somewhat larger scale. If the boundaries were 'spread out' sufficiently i.e. if $\frac{1}{n}$ were large enough compared to the size of the objects in the textured region, then over the textured region we would have $(1 - v^2)^n \simeq 0$. This region could then be detected by segmenting the function $(1 - v^2)^n$ as if it were an image itself. In any case a thorough investigation of the type of system of coupled differential (or difference) equations such as studied and used here might prove very fruitful. It may be that the splitting of the filtering problem into the evolution of v and the evolution of f might make this formulation a more revealing object of study than the 'usual' non-linear filtering equations.

Other possible developments include the formulating of a stochastic version of the evolution equations for the functions v and f leading to a Random Field interpretation of the approximating functionals and eventually perhaps to a Random Field interpretation of the the original (continuous domain) variational problem.

A more practical domain in which some of the ideas developed in this thesis might be applied is in the development of hardware for early vision computations. The form of the equations for solving for v are relatively simple and may admit efficient analog VLSI implementation such as discussed in [21].

It would be of great interest to develop the ideas which have grown up in the context of the segmentation problem in other early vision problems. The depth from stereo problem and motion tracking problem seem particularly suited to piecewise smooth or piecewise constant models. An extension of the ideas in this thesis would require the consideration of vector valued functions.

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 Γ–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] C.A. Rogers, *Hausdorff Measure.*
Cambridge University Press. 1970

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

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

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

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

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

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

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