Graduate Student Theses, Dissertations, & Professional Papers

Graduate School

2006

A three -dimensional variational approach to video segmentation

Aaron Benjamin Luttman
The University of Montana

Recommended Citation

Luttman, Aaron Benjamin, "A three -dimensional variational approach to video segmentation" (2006). *Graduate Student Theses, Dissertations, & Professional Papers*. 9592.
https://scholarworks.umt.edu/etd/9592

A THREE-DIMENSIONAL VARIATIONAL APPROACH TO VIDEO SEGMENTATION

by

Aaron Benjamin Luttman

B.S. Purdue University, U.S. 1998

M.S. University of Minnesota, U.S. 2000

presented in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

The University of Montana

May 2006

Approved by:

Committee Chair

Dean, Graduate School

5-5-06

Date

UMI Number: 3228605

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI ®

Luttman, Aaron B Ph.D., May 2006   Mathematics

A Three-dimensional Variational Approach to Video Segmentation

Committee Chair: John Bardsley, Ph.D.

In order to engage in photosynthesis, plant leaves absorb CO_2 via the opening of pores in their surfaces called *stomata*. Open stomata, however, result in the evaporation of H_2O, which is a detriment to plant function. Thus a particular leaf will seek stomatal apertures through which its need for CO_2 is balanced by its aversion to H_2O loss. In order to visualize a particular leaf's stomatal aperture, an experimentalist injects the leaf with dye so that it fluoresces when closing its stomata. The regions with a higher relative intensity correspond to areas in which the stomata are closed and the darker regions where the stomata are open. A camera is used to collect the emitted light, and a fluorescence pattern is measured. Images are recorded as these patterns change with time, resulting in a video sequence. The primary task of this work is to segment these video sequences into fluorescing and non-fluorescing regions. To do this, we propose a 3D segmentation method inspired by the *active contours without edges* approach of Chan and Vese [15]. The associated partial differential equations are solved within a level-set framework using a three-dimensional semi-implicit numerical scheme. Due to noise in the data, preprocessing is required prior to the segmentation step, and for this we use the PDE based denoising algorithm of [55] with L^1-fidelity as proposed by Chan and Esedoglu in [12].

ii

Acknowledgements

I have been blessed with more help in my research than any graduate student should be allowed, and there are many people to be thanked for that. For general help I have to thank Prof. Thomas Tonev who advised me in my research in Banach algebras which complemented my applied research nicely; along the same lines I thank also Scott Lambert for humoring me in many "math fights."

As far as the research outlined in this work is concerned, I was encouraged and helped along the way via correspondence with Professors Ron Kimmel of Technion in Israel and Luminita Vese of UCLA, two researchers who basically don't know me at all and yet were always quick with responses to any and all questions I threw at them - even when they were pretty dumb questions. I owe them thanks for that. Mr. Jan Erik Solem and Dr. Niels Christian Overgaard of the Applied Mathematics group at Malmö University were also helpful with answers to my questions regarding the differential geometry of surface manifolds.

I must also thank the people at the *Complexity and Stomatal Behavior* research lab at Utah State University for supplying me with the leaf data which is seen throughout this thesis. In particular my collaboration with Profs. David Peak and Keith Mott has been a great interdisciplinary experience from which I have learned a great deal about how to do and not to do science.

The Department of Mathematical Sciences at the University of Montana does not have the computational resources required by the segmentation algorithm outlined in this work, so I thank the *Molecular Computational Core Facility* at the University of Montana for allowing me to use their extensive computational resources for this work.

I have received financial support from the Department of Mathematical Sciences at the University of Montana, the Society of Industrial and Applied Mathematics, the Montana NSF EPSCoR program, and directly from the NSF under grant DMS-0504325. I am grateful for this financial support, as it allowed me to complete my work much faster than I would have been able to had I not received this support.

Of course the greatest contributions to my education, and this work in particular, have been made by my two advisors, Profs. Emily Stone and John Bardsley. They both always made themselves available to help me even when I didn't quite know what it was I needed help with. It is only through their intellectual, moral, and financial support that this work was possible. I am a much better mathematician and researcher for having worked with both of them.

The last person, and usually not the least, to be mentioned in many such theses is a spouse, and this work is no exception. I am not quite sure how Susannah put up with my talking math all the time, spending too much money on books, and working every weekend, but somehow she still loves me and for that she gets not only my thanks but also my love in return.

iii

Notations

$\mathbb{R}^n$	n-dimensional Euclidean space
$\mathbf{x}$	Vector in $\mathbb{R}^n$
$\|\mathbf{x}\|$	Euclidean norm of vector in $\mathbb{R}^n$
A	Matrix mapping $\mathbb{R}^n \to \mathbb{R}^m$
$\|A\|$	Spectral norm of matrix
$\langle \cdot, \cdot \rangle$	Euclidean inner product
M	Manifold or regular surface in $\mathbb{R}^n$
$\langle \cdot, \cdot \rangle_M$	Inner product on the manifold M
$\mathfrak{M}(S)$	Set of all regular surfaces contained in the set $S \subset \mathbb{R}^n$
$\mathcal{M}_0$	Surface manifold associated with a regular surface M_0
∂S	Topological boundary of the set $S \subset \mathbb{R}^n$
f	Function defined on a subset of $\mathbb{R}^n$
E	Surface functional
∇f	Gradient of function f
$\nabla_M f$	Gradient of function f with respect to the surface M
$\nabla_M E$	Gradient of surface functional E with respect to the surface M
$\boldsymbol{\alpha}(t)$	Parametric curve in $\mathbb{R}^n$
$\dot{\boldsymbol{\alpha}}(t)$	Time derivative of $\boldsymbol{\alpha}$
$D_{\mathbf{x}_0}\Phi(\mathbf{x})$	Directional derivative of Φ at $\mathbf{x}$ along the direction of $\mathbf{x}_0$
Ω	Image domain (in $\mathbb{R}^2$) or Video domain (in $\mathbb{R}^3$)
x, y	Pixel coordinates
z	Frame number in video sequence
$I(x, y)$	single image
$I(x, y, z)$	video sequence
t, τ	Artificial time parameter for surface evolution
Δt	PDE Time Step
$D_x^+ u$	Forward finite difference approximation to the first partial derivative of u with respect to x
$D_x^- u$	Backward finite difference approximation to the first partial derivative of u with respect to x
$D_x^0 u$	Centered finite difference approximation to the first partial derivative of u with respect to x
$D_{xx} u$	Centered finite difference approximation to the second partial derivative in the x direction
$\nabla^+ u, \nabla^- u$	Engquist-Osher upwind approximations to the spatial gradient of u

$\nabla_0 u$	Gradient of u approximated using centered finite differences
$u_{i,j},\ u_{i,j,k}$	Function u evaluated at the discrete point (x_i, y_j), (x_i, y_j, z_k)
$L^p(\Omega)$	Measurable functions f such that $|f|^p$ is Lebesgue integrable on Ω
$C^m(\Omega)$	Continuous functions on Ω whose partial derivatives exist and are continuous up to the mth order
$C_0^m(\Omega)$	Functions in $C^m(\Omega)$ with compact support
$M(t)$	Regular surface evolution
$\Phi(\mathbf{x})$	Level set representation of regular surface
$\Phi(\mathbf{x}, t)$	Level set representation of regular surface evolution
$\Psi(\mathbf{x})$	Variation of a level set representation $\Phi(\mathbf{x})$
v_n	Normal velocity of regular surface evolution
$\mathbf{v}_n$	Normal velocity of variational regular surface evolution
$\delta(x)$	Dirac Delta distribution
$H(x)$	Heaviside function

v

Contents

List of Figures

xi

Chapter 1

Introduction

If you're not worrying that something you're making will come out badly, or that you won't be able to understand something you're studying, then it isn't hard enough.

Paul Graham

Image processing is a field that today spans the disciplines of computer science, electrical engineering, physics, and applied mathematics. It started as a subfield of electrical engineering, as developing the hardware for visualizing digital signals was necessary before algorithms for image analysis would be relevant. After the hardware was developed - personal computers, monitors, and digital capture devices - computer scientists began working with digital signals, because the computational aspects of image analysis were then the primary difficulties. Once computers became fast enough that standard numerical algorithms such as the Fast Fourier Transform and image compression techniques could be performed quickly, the field of image processing morphed once again.

Today the main difficulties in digital signal analysis are caused by a lack of mathematical

1

models and techniques for solving the applied problems that arise on an almost daily basis. This drought of mathematics has drawn many mathematicians and physicists into image processing in the last decade, and things have changed to the point where even the *Society of Industrial and Applied Mathematics* has its own working group on *Imaging Science*, which is the current buzz-phrase used to describe the point at which the methods of engineering, mathematics, and computer science meet to solve problem involving digital images and videos.

Mathematical biology has also been a booming field in the last ten years. It has been said that the relationship between mathematics and biology in the 21st century will be analogous to the relationship between mathematics and physics in the 20th century, and we are already witnessing an expansion of biological theory based on the introduction of new mathematical methods. Some biological subfields, such as cellular and molecular biology, have seen an especially dramatic impact. According to *Mathematics and Biology: The Interface*, published by the Dutch Society of Theoretical Biology, "Molecular biology itself can trace its origins to the infusion of physical scientists into biology with the inevitable infusion of mathematical tools" [1].

Mathematical biology and imaging science have converged in the field of medical imaging, a field that at one time included only biologists but is now equally influenced by mathematicians and engineers. New mathematical models for processing images from MRI, CAT, and other medical scans have brought scientists together from all three fields to develop methods for interpreting these diagnostic techniques.

Botany is also a field that has benefitted from an increased involvement by mathematicians in recent years. New and improved mathematical models in biochemistry have allowed botanists to analyze the chemistry and physics of plants in ways that were not possible even a decade ago. New models in population dynamics have also allowed plant ecologists to better predict plant growth and loss in a diverse range of ecosystems.

The primary work of this thesis sits at one of the first crossing points of mathematics and imaging science in the field of botany. Though, as was mentioned, mathematics and imaging science have been used a great deal in biology, and mathematics and botany have paired together to address many new problems in recent years, there are few examples of all three of these fields coming together to analyze applied scientific problems. The reason for this is that it is not obvious that imaging plants is of scientific interest.

Not all biochemical processes that go on inside of a plant can be measured directly. If an indirect measurement method can be designed so that the data being measured is energy in the form of emitted photons from a physical process, then the methodologies from imaging science become immediately relevant, as imaging science focuses on the analysis of data captured by CCD devices (digital cameras), which measure photon intensities.

One particular example of a biochemical or physical process that occurs in leaves that cannot be measured directly is the phenomenon of *stomatal patchiness*. Leaves have special pores in their surfaces - called *stomata* - for absorbing carbon dioxide directly from the air, and it often occurs that groups of stomata work synchronously. It is not known how the stomata synchronize their reactions to the environment, since it can be shown that this happens even when the environment is not synchronized. An experiment can be designed in which a leaf will fluoresce when it is adjusting its stomata. An image of this fluorescence is captured using a camera and can be analyzed using the mathematical methods designed for image processing. It is these methods and their application to this particular botanical problem that form the focus of this work.

1.1 Organization

In Chapter 2, the biological problem associated with stomatal patchiness is addressed. The corresponding scientific experiment that is designed to capture the fluorescence data is also

discussed. The goal of this work is to segment each video of leaf fluorescence that is the result of a run of the leaf experiment. The data provided by the experiment must be preprocessed before it can be used in the segmentation algorithm, and Chapter 3 details the three primary preprocessing steps. Chapter 4 discusses the starting point for image segmentation in the form of some of the traditional methods for image and video segmentation. Chapters 5 and 6 are dedicated to the background theory from differential geometry and the calculus of variations that is required for understanding the video segmentation algorithm. Chapters 7 and 8 discuss the mathematical and computational aspects of the segmentation algorithm, and the results and future work are presented in Chapter 9.

Chapter 2

The Leaf

Carbon dioxide (CO_2) and water (H_2O) are the two essential resources required for a plant to engage in photosynthesis. Water is absorbed from the ground through a plant's roots, but CO_2 must be absorbed directly from the air. For this purpose leaves have pores in their surfaces called *stomata* that open and close to allow more, or less, CO_2 to be absorbed. This process is monotonic in the sense that the more open the stomata are, the more CO_2 is absorbed, and the more CO_2 absorbed the better. A leaf does not keep its stomata continually open, because H_2O is lost through the stomata via evaporation, which is a detriment to leaf function.

What this means for a leaf is that each stoma must continually regulate its aperture in order to ensure that a sufficient amount of CO_2 is absorbed and that only a tolerable amount of H_2O is lost. According to Peak, et. al. [50], "A central paradigm of plant biology is that, in the face of spatially heterogeneous and temporally varying environmental conditions, a plant continually adjusts its stomatal aperture so that, over time, it maximizes CO_2 uptake for a fixed amount of water loss." Moreover, it is thought that this optimization occurs over the entire leaf rather than for each stoma individually. It is not yet understood how plants solve this global optimization problem, but it is hypothesized (cf. [50]) that the stomatal aperture depends

on external environmental conditions and on interactions between neighboring stomata. The primary motivation for the latter claim is the observation - made in over 200 plant species - that, even under spatially varying environmental conditions, stomatal apertures are often synchronized into spatially extended patches, even though patchiness often results in less than optimal local CO_2 uptake for fixed water loss.

There is no well-understood physical or bio-chemical process by which the stomata in a leaf can be synchronized, as a leaf has no large-scale information transport network through which the stomata can communicate.

2.1 The Experiment

In order to better understand the processes by which a leaf's stomatal apertures are regulated, it is natural to attempt to visualize the process. This can be done using a technique for visualizing photosynthesis known as chlorophyll fluorescence. This process is explained in detail in [27, 60], but we outline the main ideas here.

A *Xanthium strumarium L.* (cocklebur) leaf is placed in a light-penetrating chamber such that temperature is completely controlled and gasses can be independently regulated and measured both on the top and bottom surfaces of the leaf. A dye is injected into the leaf, so that it fluoresces when its stomata close. The fluorescence happens at wavelengths above 700 nm, and hence it is measured using a camera with a long-pass filter that measures light in the appropriate spectrum. Stomatal dynamics are initiated by decreasing the concentration of H_2O on the upper surface of the leaf. This causes the leaf to close its stomata, since a decrease in humidity will correspond to an increase in H_2O loss through evaporation. The closing of stomata also causes a decrease in the concentration of CO_2 in the leaf, since the leaf cannot absorb as much as when the stomata are open. This decrease of CO_2 reduces the amount of photosynthesis occurring locally, which results in an increase of the measured

fluorescence. Thus there is an inverse relationship between photosynthesis and fluorescence. Since the density of stomata ($\approx$20,000 per cm^2) is of the same order of magnitude as the density of image pixels ($\approx$40,000 per cm^2), it is reasonable to assume that the measured change in intensity at each pixel roughly corresponds to local changes in stomatal aperture. In this way, fluorescence can be interpreted as an inverse measure of average stomatal aperture size.

As the stomatal apertures of the leaf are adjusted, images of the fluorescence can be measured (in the experiments analyzed here the images are taken approximately every 20 seconds). The measured data is then a grayscale video composed from these still images. Each experiment lasts approximately 6 hours, resulting in videos on the order of 800-1000 images. All experiments were performed by the *Complexity and Stomatal Behavior* research lab at Utah State University; further details can be found in [50]. Several example images from different experimental runs can be seen in Figure 2.1.

The most common result of the experiment is for stomatal apertures across the entire leaf to adjust to a uniform size in direct reaction to environmental conditions, and this steady-state is usually found quite quickly (approximately 30-60 minutes). It is sometimes observed, however, that spatial groups of stomata will synchronize their apertures, forming "patches." The stomata in these patches then act as a unit, opening and closing in synchrony. Moreover, each patch can change size and position in time. In most cases, after 1-2 hours of oscillations in patch size and shape, the uniform steady state mentioned above is found by the leaf, i.e. all stomata settle to the same uniform configuration. On rare occasions, however, no steady state is found for as long as the data is measured (approximately 6 hours). This suggests that the solution state for the global optimization is unstable and that the leaf is sensitive to microscopic changes in gaseous configurations.
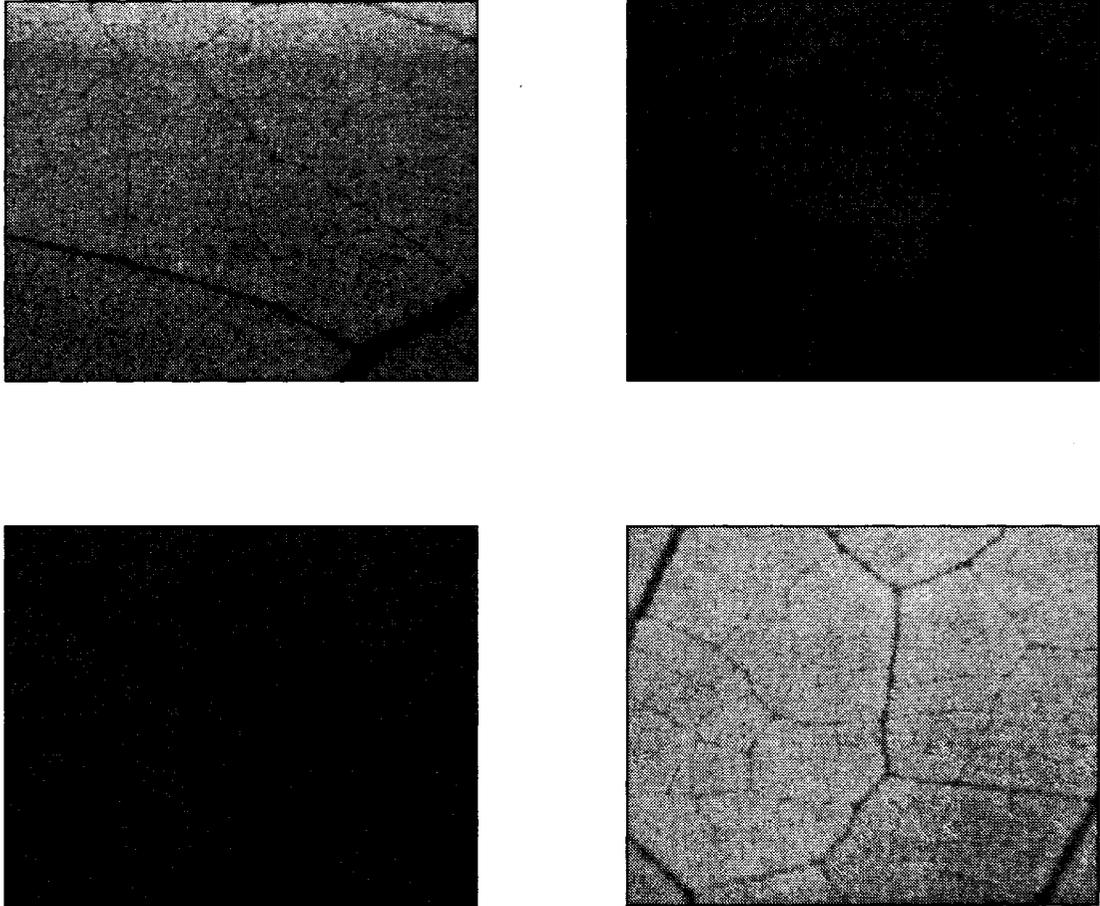
Figure 2.1: Sample Images from Four Runs of the Visualization Experiment

2.2 The Goals of This Work

The primary goal of this work is to develop an algorithm that extracts the actively fluorescing regions in each frame of a leaf video. The algorithm should pick out the regions in each image of a leaf sequence that are bright due to stomatal dynamics but not segment regions that are bright or dark independent of the dynamics, such as the vein regions which always appear dark (see Figure 2.1). It is hoped that a map of the fluorescence over the time of an experimental run can give insight into the possible dynamical processes - physiological, bio-chemical, or physical - that induce the oscillations in the sizes and shapes of stomatal patches.

Chapter 3

Video Preprocessing

The raw video data returned by a run of the leaf experiment requires several preprocessing steps before it can be analyzed and segmented. These preprocessing steps are the focus of this chapter. In particular there are two types of noise that must be offset before the data adequately represents the fluorescence being captured in the experiment. The first type of noise, which is common to all digital signal capture devices, is the high frequency random noise associated with digital signal instrumentation. The second type of noise is low-frequency and is caused by uncontrolled lighting conditions in the lab where the experiment is being run. In what follows, $I(x, y)$ will denote a single, gray-scale (8 bit) image that takes integer values from 0 to 255, and $\Omega \subset \mathbb{R}^2$ will denote the image domain.

3.1 Image Denoising

A CCD camera captures an image by counting photons of visible wavelengths on each of its pixels. Any counting procedure is an inherently noisy process, and noise must be removed from the images before they can be analyzed. Details of noise statistics for CCD cameras can

9

be found in [62].

3.1.1 Classical Methods for Image Denoising

The standard technique for image denoising is convolution with a Gaussian kernel. The Gaussian kernel with standard deviation σ is given by $G_\sigma(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left((-x^2 - y^2)/2\sigma^2\right)$, and the smoothed image, I_σ, corresponding to the convolution with G_σ is

$$I_\sigma(x_0, y_0) = (I * G_\sigma)(x_0, y_0) = \int_\Omega I(x,y)G_\sigma(x_0 - x, y_0 - y) \, dxdy.$$

The function G_σ is infinitely differentiable, and if $I \in L^2(\mathbb{R}^2)$ is convolved with $G_\sigma \in C_0^\infty(\mathbb{R}^2)$, which denotes the space of infinitely differentiable functions on $\mathbb{R}^2$ with compact support, then $I_\sigma \in C_0^\infty(\mathbb{R}^2)$. This is a standard result of real analysis (cf. [72] Theorem (9.3)).

Thus I_σ is a smoothed version of I. This smoothing process is called *isotropic diffusion*, because it is computed by locally averaging in all radial directions equally without regard to image content. If such an approach were used on images such as those in Figure 3.1, the result would be that the veins would be smoothed into the inter-vein regions. One of the primary goals of the video segmentation is to determine the effect of the veins on the patchiness of the fluorescence, so the vein boundaries must be preserved with sharp edges. Hence Gaussian convolution is not appropriate for this application.

3.1.2 ROF Denoising

Several image denoising techniques have been developed that allow for removal of noise while retaining sharp edges in an image. These techniques smooth only locally within regions of relatively homogeneous intensity. A few examples are the variations of anisotropic diffusion methods such as those found in [4,11,51].

Image 2

Image 150

Image 275

Image 400
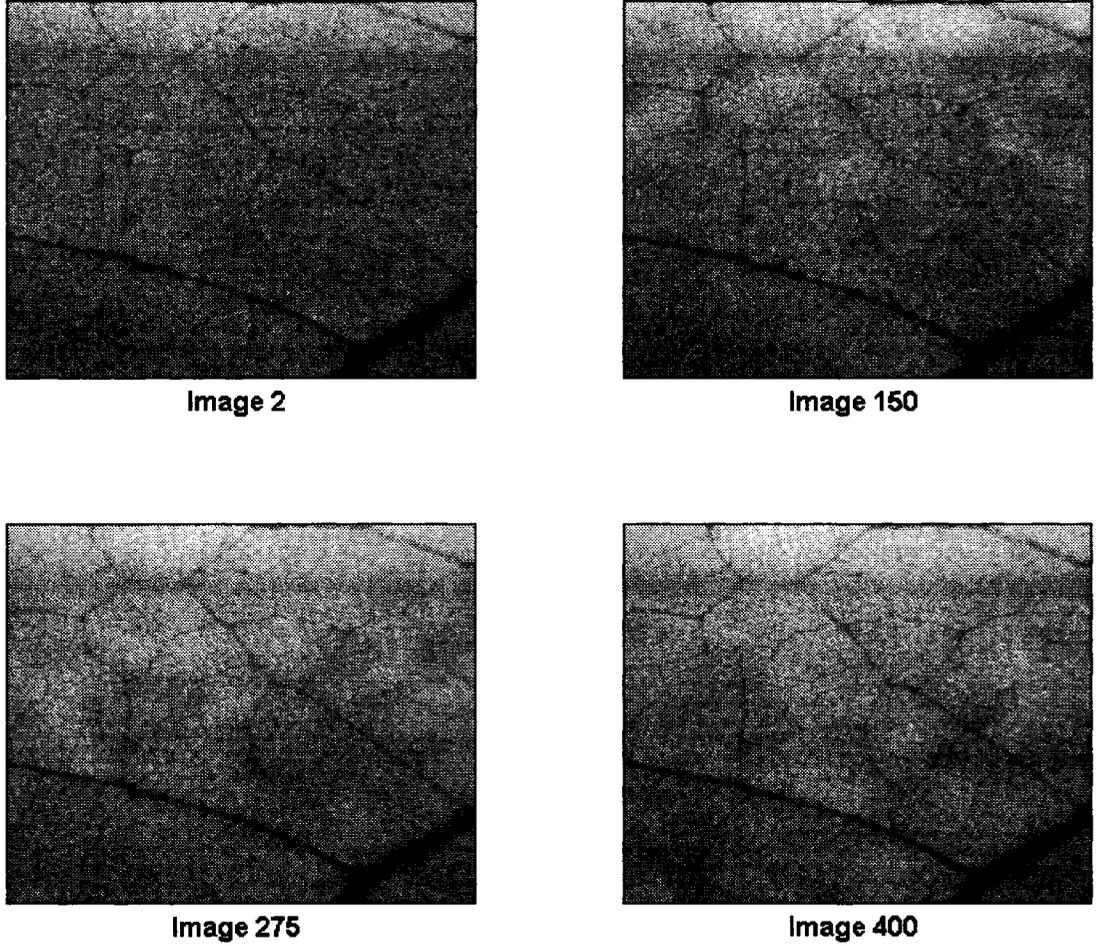
Figure 3.1: Four Images from Experiment with Patchiness

The most well-known edge-preserving method for denoising images is the method of Rudin-Osher-Fatemi (ROF) [55]. The goal is to compute an image $I(x,y)$ that is a local minimizer of the functional

$$E(I) = \int_\Omega |\nabla I| \, d\mathbf{x} + \lambda \int_\Omega (I_0 - I)^2 \, d\mathbf{x}. \tag{3.1}$$

Here $\mathbf{x} = (x,y)$,

$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right),$$

and $I_0(\mathbf{x})$ is the measured image. The first term in (3.1) is a smoothing term, which ensures that the resulting $I(x,y)$ will be differentiable, and the second term is the *fidelity* term, which

ensures that $I(x, y)$ will approximate $I_0(\mathbf{x})$ in $L^2(\Omega)$.

Before detailing the minimization process for this particular functional, we first outline some of basic results that ensure that the steady-state solution of an evolution equation is a local minimizer of a functional. First we require some notation.

Definition 3.1.1. *Let* $F \colon \mathbb{R}^n \to \mathbb{R}^n$ *be a differentiable vector field,*

$$F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_n(\mathbf{x})) = (f_1(x_1, \ldots, x_n),\ f_2(x_1, \ldots, x_n),\ \ldots,\ f_n(x_1, \ldots, x_n)).$$

Then

$$\nabla \cdot F = \left(\frac{\partial f_1}{\partial x_1}, \ \ldots, \ \frac{\partial f_n}{\partial x_n} \right).$$

Now we define the variational derivative of a functional with respect to its argument.

Definition 3.1.2 (Gelfand, [26], p. 28). *Let* $\Omega \subset \mathbb{R}^n$, *and*

$$E(I) = \int_\Omega F(\mathbf{x}, I, \nabla I)\ d\mathbf{x}.$$

Then the variational derivative of E with respect to I *is given by*

$$\frac{\delta E}{\delta I} = \frac{\partial}{\partial I} F(\mathbf{x}, I, \nabla I) - \nabla \cdot \left(\frac{\partial}{\partial (\nabla I)} F(\mathbf{x}, I, \nabla I) \right). \tag{3.2}$$

A necessary, but not generally sufficient, condition for I to be a local minimizer of E is that the variational derivative of E is zero at I. This condition,

$$\frac{\partial}{\partial I} F(\mathbf{x}, I, \nabla I) - \nabla \cdot \left(\frac{\partial}{\partial (\nabla I)} F(\mathbf{x}, I, \nabla I) \right) = 0,$$

is called the *Euler-Lagrange equation for E*. It should be noted here that the derivatives of F in (3.2) are not required to exist in the classical sense but only in the weak sense. If the derivatives exist *almost everywhere,* in the sense of Lebesgue measure, then they exist in the

weak sense. Note also that I and ∇I are treated as independent variables.

The variational derivative (cf. [26]) obeys the standard rules for classical derivatives such as the chain rule and product rule. Thus, in order to compute the variational derivative of a functional $E(I(\mathbf{x}, t))$ with respect to t, it is only necessary to apply the chain rule,

$$\frac{\delta E}{\delta t} = \frac{\delta E}{\delta I} \frac{\partial I}{\partial t}.$$

Here t is an artificial time parameter, we assume that I is differentiable with respect to t in the classical sense. This allows us to develop a minimization procedure for E by evolving I according an evolution equation that will ensure that the variational derivative of E with respect to t is non-positive. This is guaranteed when $\partial I / \partial t = -\delta E / \delta I$. Then

$$\frac{\delta E}{\delta t} = \frac{\delta E}{\delta I} \frac{\partial I}{\partial t} = -\frac{\delta E^2}{\delta I},$$

which is clearly non-positive, and a minimization procedure for $E(I(\mathbf{x}, t))$ is given by solving the initial-value problem

$$\frac{\partial I}{\partial t} = -\frac{\delta E}{\delta I}, \quad I(\mathbf{x}, 0) = I_0(\mathbf{x}).$$

In order to determine the minimization for (3.1), it is only necessary to compute $\delta E / \delta I$.

Let $F(\mathbf{x}, I, \nabla I) = |\nabla I| + \lambda (I_0 - I)^2$, then, by (3.2),

$$\frac{\delta E}{\delta I} = \frac{\partial}{\partial I} F(\mathbf{x}, I, \nabla I) - \nabla \cdot \left(\frac{\partial}{\partial (\nabla I)} F(\mathbf{x}, I, \nabla I) \right) = -2\lambda (I_0 - I) - \nabla \cdot \frac{\nabla I}{|\nabla I|}.$$

Thus the minimization procedure for $E(I)$ is given by

$$\frac{\partial I}{\partial t} = \nabla \cdot \left(\frac{\nabla I}{|\nabla I|} \right) + \lambda (I_0 - I), \tag{3.3}$$

where the factor of 2 has been absorbed into the factor λ.

If, rather than computing the fidelity as a measure in $L^2(\Omega)$, we measure the fidelity in $L^1(\Omega)$, (3.1) becomes

$$E(I) = \int_\Omega |\nabla I| \, d\mathbf{x} + \lambda \int_\Omega |I_0 - I| \, d\mathbf{x}. \tag{3.4}$$

This is a variant of the ROF scheme proposed by Chan and Esedoglu in [12]. A direct calculation shows that the variational derivative of E is given by

$$\frac{\delta E}{\delta I} = -\nabla \cdot \left(\frac{\nabla I}{|\nabla I|} \right) - \lambda \frac{I_0 - I}{|I_0 - I|},$$

so the minimization procedure is given by the initial value problem

$$\frac{\partial I}{\partial t} = \nabla \cdot \left(\frac{\nabla I}{|\nabla I|} \right) + \lambda \frac{I_0 - I}{|I_0 - I|}, \quad I(\mathbf{x}, 0) = I_0(\mathbf{x}). \tag{3.5}$$

Both with the traditional ROF and the Chan-Esedoglu variant, λ is a weighting factor for the fidelity term.

Minimizing the functional in (3.4) according to (3.5) is numerically unstable, so a regularized version of the functional must be minimized instead. The regularized functional,

$$E(I) = \int_\Omega \sqrt{|\nabla I|^2 + \epsilon} \, d\mathbf{x} + \int_\Omega \sqrt{|I_0 - I|^2 + \delta} \, d\mathbf{x}, \tag{3.6}$$

is minimized by evolving I according to

$$\frac{\partial I}{\partial t} = \nabla \cdot \left(\frac{\nabla I}{\sqrt{|\nabla I|^2 + \epsilon}} \right) + \lambda \frac{I_0 - I}{\sqrt{|I_0 - I|^2 + \delta}}, \tag{3.7}$$

subject to $I(\mathbf{x}, 0) = I_0$. The factors ϵ and δ are regularization parameters.

Image 2

Image 150

Image 275

Image 400
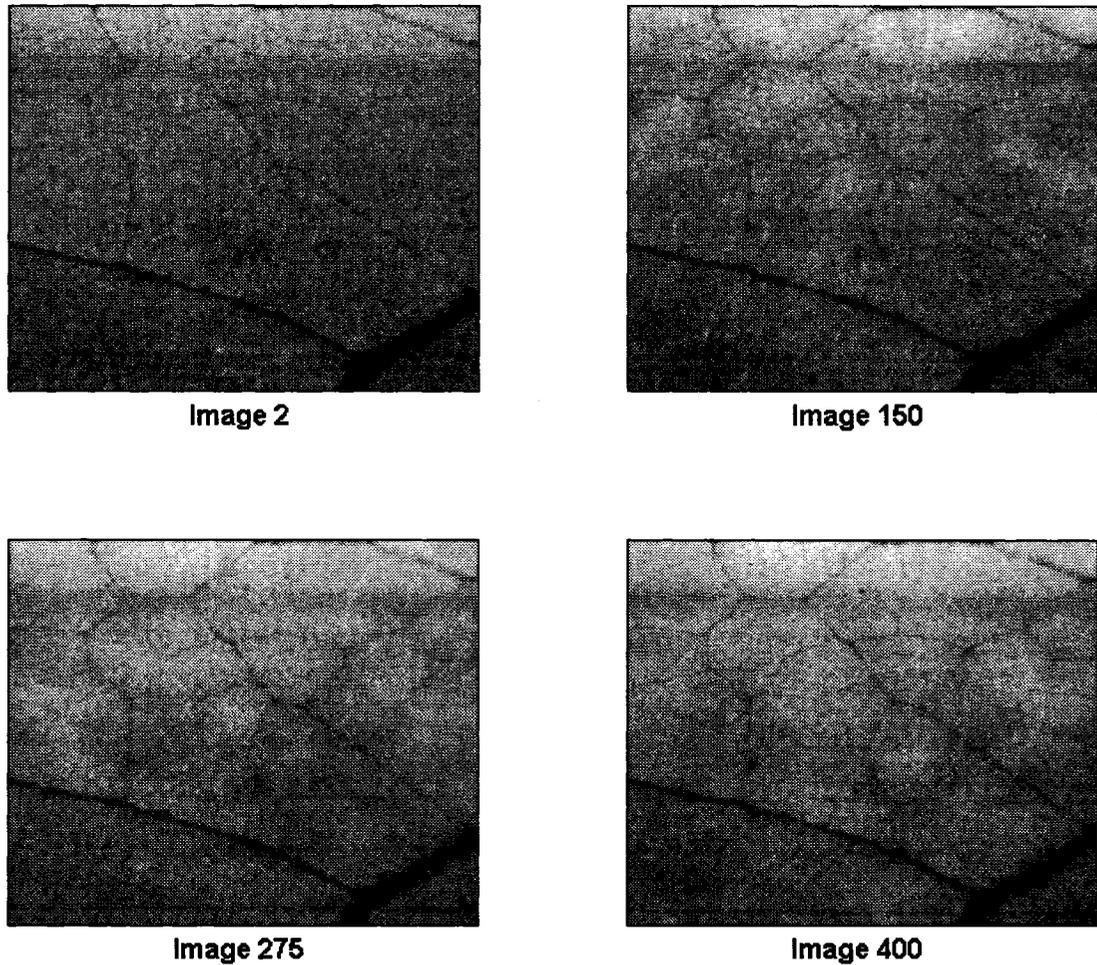
Figure 3.2: Denoised Images from Experiment with Patchiness

3.1.3 Numerical Scheme

Each image in the input video sequence is individually denoised by evolving it according to equation (3.7). This must be done numerically. Let $I_{i,j}^n$ represent the pixel in row i and column j at the nth numerical iteration and $I_0(i,j)$ be the value of the measured image at the $(i,j)^{\text{th}}$ pixel. Then equation (3.7) can be discretized explicitly in time to yield the following

explicit numerical scheme as given in [12]:

$$\frac{I_{i,j}^{n+1} - I_{i,j}^{n}}{\Delta t} = D_x^- \left(\frac{D_x^+ I_{i,j}^n}{\left((D_x^+ I_{i,j}^n)^2 + (D_y^+ I_{i,j}^n)^2 + \epsilon \right)^{1/2}} \right) + \tag{3.8}$$

$$D_y^- \left(\frac{D_x^+ I_{i,j}^n}{\left((D_x^+ I_{i,j}^n)^2 + (D_y^+ I_{i,j}^n)^2 + \epsilon \right)^{1/2}} \right) + \lambda \left(\frac{I_0(i,j) - I_{i,j}^n}{\left((I_0(i,j) - I_{i,j}^n)^2 + \delta \right)^{1/2}} \right),$$

where the spatial forward and backward finite difference approximations to the partial derivatives are given by

$$D_x^+ I_{i,j}^n = I_{i+1,j}^n - I_{i,j}^n, \qquad D_y^+ I_{i,j}^n = I_{i,j+1}^n - I_{i,j}^n, \tag{3.9}$$
$$D_x^- I_{i,j}^n = I_{i,j}^n - I_{i-1,j}^n, \qquad D_y^- I_{i,j}^n = I_{i,j}^n - I_{i,j-1}^n.$$

These notations will be used throughout the sequel.

Note that Δx and Δy do not appear in this numerical scheme, because the natural, pixel-based discretization, $\Delta x = \Delta y = 1$, is used.

Denoised versions of the images in Figure 3.1 can be seen in Figure 3.2.

3.2 Median-change Normalization

Each run of the leaf experiment lasts approximately six hours, and the lighting conditions in the lab are not tightly regulated. The result of this is that there are global effects seen in each leaf video that must be accounted for and offset before a leaf video can be processed.

Pixel intensities in a particular leaf video change for two reasons. The first reason - and that which is being analyzed - is the stomatal dynamics, which can cause intensities to both decrease and increase. The second reason intensities change is global lighting effects, which

can also cause both increases and decreases in intensity in time but only a global increase or global decrease for a given frame. If we assume that less than half of the pixels in a given frame in the video sequence are increasing in intensity due to dynamics and less than half are decreasing due to dynamics, then the median intensity change is caused by the global lighting effects for that frame. It is believed that this is a reasonable assumption for the leaf data based on direct observation. In order to offset these effects, the median intensity change must be removed. This means that each image $I(x, y, z)$ in the video sequence, where x and y are pixel coordinates and z corresponds to the frame number in the sequence, is replaced with $I^*(x, y, z)$ where

$$I^*(x, y, z) = I(x, y, z) - \underset{(x,y)}{\text{median}} \left(I(x, y, z) - I^*(x, y, z - 1) \right).$$

This is the *median-change normalization*. Note that this is not equivalent to normalizing each image to have the same median value. There is no change at the first image in a video sequence, so $I^*(x, y, 1) = I(x, y, 1)$.

After each image in the video sequence is denoised and the entire video sequence is median-change normalized, the entire video sequence is linearly scaled so that the intensity values take on the full range $[0, 255]$. Since each image in the preprocessed video sequence was computed via (3.7) prior to the median-change normalization, the data is continuous and is not limited to integer values on the interval $[0, 255]$. Fully preprocessed images can be seen in Figure 3.3 and should be compared with those in Figures 3.1 and 3.2.
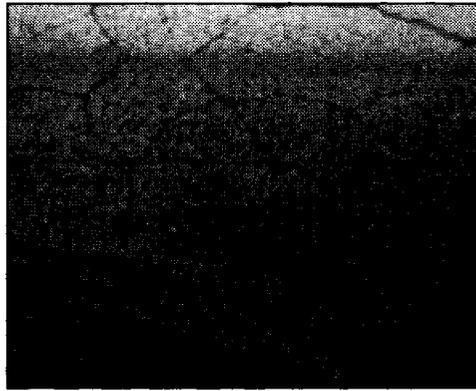
Image 2
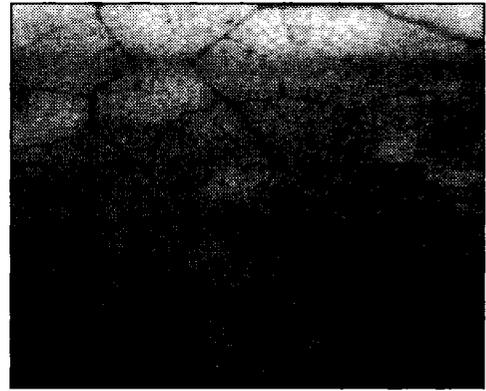
Image 150
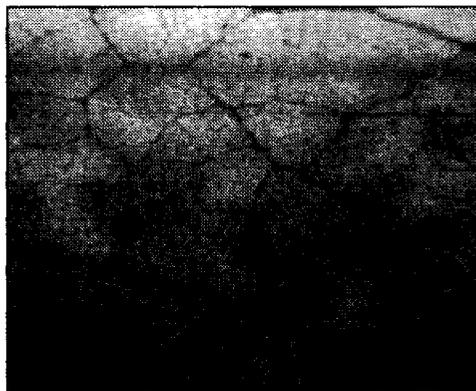
Image 275
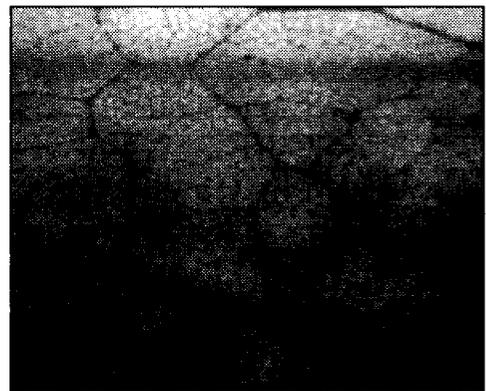
Image 400

Figure 3.3: Fully Proprocessed Images from Experiment with Patchiness

Chapter 4

Classical Segmentation Techniques

Image segmentation is the term used for the extraction of salient features from digital still images or digital video. A salient feature is any information in the image or video that is to be extracted, such as edges, corners, or regions of homogeneous intensity. Image segmentation is one of the primary problems in computer vision and image processing, and there are many techniques and approaches for segmenting different types of images based on the specific features being segmented. Before developing the method of Chapter 7, we first recall some of the classical techniques used for image and video segmentation.

4.1 Image Segmentation Techniques

4.1.1 Threshold-based Techniques

The first techniques for segmenting gray-scale images were based on thresholding, a process that yields a binary segmentation of the input image. A user would *a priori* select an intensity value so that all intensities above that value were "bright" and all below that value were

19

"dark." The primary reason a technique of this type does not work well is that it requires the intensity scales to be the same across an entire image. Regions are segmented globally, and locally bright or dark regions would not necessarily be segmented. If regions that are only locally bright or dark are to be segmented and the intensities scales are not the same across an entire image, then a threshold-based segmentation must be done locally on subregions, a process that requires a great deal of user interaction.

4.1.2 Edge-Based Segmentation Techniques

Since segmenting based strictly on intensity values is not optimal for many types of data, the next natural step is to segment based on local intensity extrema. This leads to a class of techniques that segment based on *edge* data. An edge in an image is defined as a contour whose pixels are local extrema of the absolute value of the intensity gradient. There is a large class of methods for computing edges in an image, known as an *edge map*. The most well-known is the Canny edge detector [9], but there are others that compute edge maps in related ways. A few examples of such methods are the Haralick edge detector [30,31], the Marr-Hildreth edge detector [41], and the SUSAN edge detector [61]. Each of these methods detects edges in an image and thereby segments an image into disjoint regions that are separated by the edges. A few such algorithms can be found in [33,43,75].

Another class of unrelated, edge-based segmentation techniques is composed of the original active contour models, also called *snakes*. These methods take a variational approach to edge detection by defining edges as extremal values of segmentation functionals [6,35,74]. Though these approaches are not mathematically similar to the edge detection methods mentioned above, the result is the same: a segmentation that separates regions of the input image based on edges.

Edge-based image segmentation is not an appropriate method for the leaf data, because the
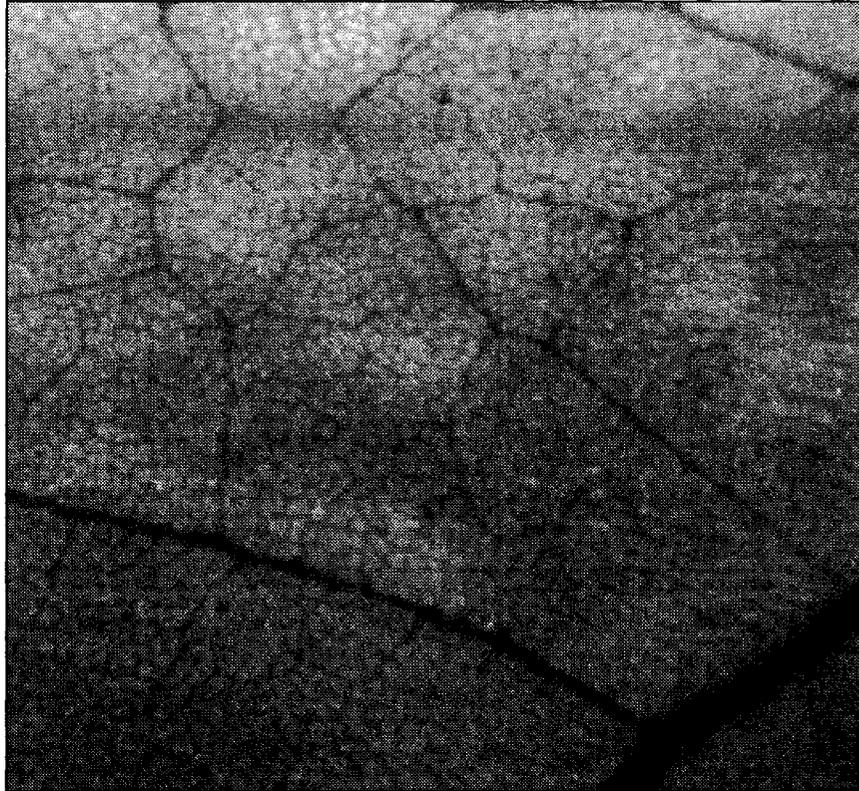
Figure 4.1: Edges Correspond to Veins in the Leaf Images

primary edges in the leaf images are the boundaries of the veins. This can be seen in Figure 4.1. The goal of this work is to segment the regions of the leaf that are actively fluorescing at a given time, but the boundaries of the fluorescing regions do not form edges that are as sharp as the edges corresponding to the vein boundaries. An edge-based technique would extract the vein boundaries and not yield information about the fluorescing regions.

4.1.3 Region-Based Techniques

The pixels in a region that is segmented from an image by an edge-based technique do not necessarily have any uniform properties. Such a region is defined only by a boundary that is determined by the edge map. For some data, however, it is desired to extract regions from an image based on local intensity properties. This leads to region-based segmentation techniques, the primary goal of which is to extract regions from an image that have relatively homogeneous intensity values. An example of a region-based technique that is similar to the variational technique that will be introduced in Chapter 7 can be found in [49], where the authors use a probabilistic method for developing a gradient descent-type minimization procedure for a region-finding functional. Another region-based technique is given in [34], where a variational approach is used to segment regions of stereo images for three-dimensional object reconstruction.

The primary drawback of these two methods in light of the leaf data - along with other region-based segmentation techniques - is that they attempt to segment an image based on groupings of pixels with similar intensities. This is not an appropriate method for segmenting the leaf data, since this would lead to a segmentation that is similar to the segmentation given by an edge-based technique. An edge-based technique computes the boundaries of the veins in each image, whereas a region-based technique would extract the entire vein regions. This is due to the fact that the only regions in the leaf images with relatively homogeneous intensity are the dark regions that make up the veins. Even the fluorescing regions, which appear bright to the eye, are not relatively homogeneous in intensity, particularly when they cross the veins. This is due to the fact that the veins in the leaf do not fluoresce in the same way that the inter-vein regions do, and veins remain dark even when they are in an actively fluorescing region. Moreover, since these veins are constant in each image in a leaf video sequence, there is no three-dimensional extension of one of these region-based methods that would overcome this difficulty.

Threshold-based, edge-based, and region-based techniques have also been combined to form a class of hybrid methods (see e.g. [66, 73]) and, though these work well in applications where edge-based and classical region-based information are relevant, they do not overcome the problems of these methods for the leaf data.

4.2 Background Subtraction and Motion Tracking

The natural approach to segmenting the regions of active fluorescence in the leaf video sequences is a background subtraction technique. This is a class of methods that generate a background model based on what the image would look like if there were no intensity dynamics. For a street intersection, a background model would be an image of the intersection when there are no cars present. A car could then be detected by comparing the scene with a car to the scene without a car. For the leaf data, the background model is an image of the given leaf when there is no active fluorescence. Since it is possible to develop such a background model for the leaf data directly from the experiment, this is a useful way of determining which regions in an image of a leaf video sequence are actively changing their intensities.

The three primary approaches to background-based video segmentation are rigid region or feature tracking, classical background subtraction, and blob or non-rigid motion tracking. The regions of active fluorescence in a leaf video sequence are constantly changing their size and shape in a non-rigid way, so rigid region and feature tracking techniques are not appropriate.

Background substraction has been studied a great deal, and there are many varying techniques. A review of these methods can be found in [44], but most involve subtracting a current frame from a dynamic background model that is based on individual pixel statistics or frame differencing. If the background model is based on pixel statistics, then each pixel is treated statistically independently. This is not appropriate for the leaf data, since the fluorescence measured at a pixel due to stomatal dynamics is strongly correlated with fluorescence

measured at neighboring pixels. If the background model is based on frame differencing, then a segmentation based on background subtraction will only yield results looking backward in time. With the leaf data, it is useful to take into account the full time series for each pixel - rather than looking only backward in time - as this allows one to compute the continuous deformation of the fluorescing regions over the entire video.

Another primary reason that traditional background subtraction techniques will not yield the desired segmentation for the leaf data is that the vein pixels never undergo intensity increases of the same magnitude as non-vein pixels, even when a region of fluorescence crosses the veins. Pixels corresponding to the veins are dark in the background model, but they are dark throughout each leaf video, even when they are in a region of active fluorescence. Thus the difference between vein pixels in an active frame and vein pixels in the background model would be small, indicating that the vein pixels are in the background even when in a fluorescing region.

Other classes of techniques, such as *Blob tracking* or model-based methods for non-rigid region tracking, also break down on the leaf data. Blob tracking methods such as the *Active Blobs* of Sclaroff and Isidoro [58] use texture information for tracking non-rigid blobs. In the case of the leaf data, texture is caused by the veins, which do not move with the fluorescing regions, so texture information cannot be used for extracting or tracking the regions. Methods for non-rigid motion tracking such as the *Motion of Disturbances* method proposed by Halevy and Weinshall in [29] or that developed for facial movements in [5] also are not applicable in this case as they use dynamic background models or parametric models that do not accurately model the non-rigid deformation of the fluorescence regions in the leaf data. Other methods for non-rigid motion tracking have similar drawbacks.

Due to the fact that none of these classical techniques is appropriate for segmenting the leaf videos in 3D, it is necessary to develop another approach that does not suffer similar drawbacks. Such a development is the motivation for the algorithm in Chapter 7. Before that

algorithm is described, however, we and discuss the necessary mathematical theory for our variational approach.

Chapter 5

The Surface Manifold and Its Tangent Space

The two primary mathematical tools used in the development of the video segmentation algorithm in the chapters that follow are the calculus of variations and differential geometry. In this chapter the necessary background for the differential geometry of surface functionals is detailed. The derivation here is based on that of Solem and Overgaard in [63, 64], but many of the results and definitions are more general results from differential geometry and can be found, for example, in [20, 32, 69].

5.1   Finite-dimensional Gradients and Gradients on Surfaces

Before discussing gradients of surface functionals, we present some definitions and theoretical results.

26

5.1.1 Gradient of Functions in $\mathbb{R}^n$ and Steepest Descent

A *local minimizer* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a point $\mathbf{x}' \in \mathbb{R}^n$ such that $f(\mathbf{x}') \leq f(\mathbf{x})$ for all $\mathbf{x}$ in an open neighborhood $V \subset \mathbb{R}^n$ of $\mathbf{x}'$. If $f \in C^1(\mathbb{R}^n)$ and $\mathbf{x}_0 \in \mathbb{R}^n$, then a method for computing a local minimizer of f is given by solving the initial-value problem

$$\frac{d}{dt}\mathbf{x}(t) = -\nabla f(\mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{5.1}$$

where ∇f, the *gradient of f*, is defined by

$$\nabla f(\mathbf{x}_0) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right) \Big|_{\mathbf{x}=\mathbf{x}_0}.$$

Minimizing f by solving (5.1) is called the *steepest descent method*. In practice the steepest descent method does not always yield a local minimizer of f, but we can say the following.

Theorem 5.1.1. *Let $\mathbf{x}(t)$ be a solution of (5.1). Then $f(\mathbf{x}(t))$ is a non-increasing function of t.*

Proof. Applying the chain rule for differentiation together with (5.1) yields

$$\frac{d}{dt}f(\mathbf{x}(t)) = \nabla f(\mathbf{x}(t)) \cdot \frac{d}{dt}\mathbf{x}(t) = \nabla f(\mathbf{x}(t)) \cdot (-\nabla f(\mathbf{x}(t))) = -\|\nabla f(\mathbf{x}(t))\|^2. \tag{5.2}$$

Since the time-derivative of f is non-positive, f is a non-increasing function of t. $\qquad\square$

If f is not bounded from below, then one cannot expect that $\mathbf{x}(t)$ will converge as $t \to \infty$. If $\mathbf{x}(t)$ does converge, however, then it must converge to a *critical point* of f, i.e. a point $\mathbf{x}'$ such that $\nabla f(\mathbf{x}') = 0$.

Theorem 5.1.2. *Let $\mathbf{x}(t)$ be the solution of (5.1), and suppose that $\mathbf{x}(t) \to \mathbf{x}'$ as $t \to \infty$. Then $\nabla f(\mathbf{x}') = 0$.*

Proof. Let $\{h_k\}_{k=1}^{\infty}$ be a sequence of positive real numbers such that $h_k \to 0$ monotonically as $k \to \infty$. Then, by the definition of the derivative of a function of a real variable,

$$\frac{\mathbf{x}(t + h_k) - \mathbf{x}(t)}{h_k} \to \frac{d\mathbf{x}(t)}{dt} \quad \text{as} \quad k \to \infty.$$

Let $\epsilon > 0$ be given. Since $\mathbf{x}(t) \to \mathbf{x}'$, for each $k \in \mathbb{N}$ there exists $t_k \in \mathbb{R}^+$ such that $|\mathbf{x}(t) - \mathbf{x}'| < h_k \epsilon$ for all $t > t_k$. Thus, for all $t > t_k$, $|\mathbf{x}(t + h_k) - \mathbf{x}'| < h_k \epsilon$, and

$$\left| \frac{d}{dt} \mathbf{x}(t) \right| = \left(\lim_{k \to \infty} \left| \frac{\mathbf{x}(t + h_k) - \mathbf{x}(t)}{h_k} \right| \right) \leq \lim_{k \to \infty} \left(\left| \frac{\mathbf{x}(t + h_k) - \mathbf{x}'}{h_k} \right| + \left| \frac{\mathbf{x}(t) - \mathbf{x}'}{h_k} \right| \right)$$

$$\leq \lim_{k \to \infty} \left(\frac{h_k \epsilon}{h_k} + \frac{h_k \epsilon}{h_k} \right) \leq \lim_{k \to \infty} (\epsilon + \epsilon) = 2\epsilon.$$

Since for any $\epsilon > 0$ there exists $t_k \in \mathbb{R}^+$ such that $\left| \frac{d}{dt} \mathbf{x}(t) \right| < 2\epsilon$ for all $t > t_k$, we must have that $\frac{d}{dt} \mathbf{x}(t) \to 0$ as $t \to \infty$. Thus

$$\nabla f(\mathbf{x}') = \lim_{t \to \infty} \nabla f(\mathbf{x}(t)) = \lim_{t \to \infty} \frac{d}{dt} \mathbf{x}(t) = 0.$$

$\square$

Before continuing we note that $\nabla f(\mathbf{x}') = 0$ is a necessary, but not generally sufficient, condition for $\mathbf{x}'$ to be a local minimizer of f.

5.2 Regular Surfaces in $\mathbb{R}^n$ and Level Set Representations

We now define the gradient of a function on a surface in $\mathbb{R}^n$ or a manifold. First we define a manifold using the definition of Hörmander in [32].

Definition 5.2.1. *A topological Hausdorff space M with a countable basis is called an m-dimensional manifold if for every point $\mathbf{x} \in M$ there exists an open neighborhood $U_{\mathbf{x}} \subset M$ of $\mathbf{x}$, an open set $V_{\mathbf{x}} \subset \mathbb{R}^m$, and a homeomorphism $f_{\mathbf{x}} : U_{\mathbf{x}} \to V_{\mathbf{x}}$.*

Precise definitions of topological terms can be found in [47], but, intuitively, this means that the local geometry of M is Euclidean and of the same dimension at every point. This Euclidean dimension is the dimension of the manifold. In the remainder of this chapter, M will denote an m-dimensional manifold, which is assumed to be a subset of $\mathbb{R}^n$ for some $n \geq m$.

Analogous to what was done in the previous section, our goal is to find a procedure for finding $\mathbf{x}' \in M$ such that $f(\mathbf{x}') \leq f(\mathbf{x})$ for all $\mathbf{x} \in U_{\mathbf{x}} \subset M$, where $U_{\mathbf{x}}$ is some open neighborhood of $\mathbf{x}'$. In order to generalize (5.1), it is necessary to define the gradient of a function f that is defined on a manifold M, and for this we define the tangent space to a manifold at a point.

Definition 5.2.2. *Let $\boldsymbol{\alpha}\colon (-\epsilon, \epsilon) \to M$, for some $\epsilon > 0$, be a differentiable curve such that $\boldsymbol{\alpha}(0) = \mathbf{x_0}$. Then the vector $\mathbf{v} = \frac{d}{dt}\boldsymbol{\alpha}(0)$ is called a tangent vector to M at $\mathbf{x_0}$. The tangent space to M at $\mathbf{x_0}$, denoted $T_{\mathbf{x_0}}M$, is the set of all tangent vectors to M at $\mathbf{x_0}$.*

Theorem 5.2.3. *Given $\mathbf{x_0} \in M$, $T_{\mathbf{x_0}}M$ is a vector space.*

Proof. First of all, it is clear that the zero vector is an element of $T_{\mathbf{x_0}}M$, since it is the velocity of the constant map, $\boldsymbol{\alpha}(t) = \mathbf{x_0}$ for all $t \in (-\epsilon, \epsilon)$.

Suppose that $\mathbf{v}$ is an element of $T_{\mathbf{x_0}}M$ corresponding to the differentiable curve $\boldsymbol{\alpha}$. Then $c\mathbf{v}$ for $c \in \mathbb{R}$ is also an element of the tangent space to M at $\mathbf{x_0}$. This follows from the facts that $\frac{d}{dt}(\boldsymbol{\alpha}(ct)) = c\frac{d}{dt}\boldsymbol{\alpha}(t) = c\mathbf{v}$ and that $\boldsymbol{\alpha}(ct)$ is also a differentiable curve on $(-\epsilon, \epsilon)$ with $\boldsymbol{\alpha}(c \cdot 0) = \boldsymbol{\alpha}(0) = \mathbf{x_0}$.

Lastly, suppose that $\mathbf{v_1}$ and $\mathbf{v_2}$ are tangent vectors at $\mathbf{x_0}$ associated with curves $\boldsymbol{\alpha}_1(t)$ and $\boldsymbol{\alpha}_2(t)$. Then $\frac{1}{2}(\boldsymbol{\alpha}_1(t) + \boldsymbol{\alpha}_2(t))$ is a differentiable curve on $(-\epsilon, \epsilon)$, $\frac{1}{2}(\boldsymbol{\alpha}_1(0) + \boldsymbol{\alpha}_2(0)) = \mathbf{x_0}$, and

$$\frac{d}{dt}\frac{1}{2}(\boldsymbol{\alpha}_1(t) + \boldsymbol{\alpha}_2(t))\bigg|_{t=0} = \frac{1}{2}\left(\frac{d}{dt}\boldsymbol{\alpha}_1(0) + \frac{d}{dt}\boldsymbol{\alpha}_2(0)\right) = \frac{1}{2}(\mathbf{v_1} + \mathbf{v_2}).$$

Thus $\frac{1}{2}(\mathbf{v_1} + \mathbf{v_2})$ is the velocity associated with the curve $\frac{1}{2}(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2)(t)$, and, since constant multiples of a tangent vector are tangent vectors, we get also that $\mathbf{v_1} + \mathbf{v_2}$ is an element of

the tangent space $T_{\mathbf{x_0}}M$.

Since differentiation is a linear mapping, the other properties of vector spaces follow immediately. □

Lemma 5.2.4. *The tangent space to a manifold at a point* $\mathbf{x}$, $T_{\mathbf{x}}M$, *is an inner product space with inner product defined by*

$$\langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{x}} = \sum_{i=1}^{n} v_i w_i, \quad \mathbf{v}, \mathbf{w} \in T_{\mathbf{x}}M.$$

It is clear that this is an inner product, since it is the Euclidean inner product on $\mathbb{R}^n$.

A function $f\colon M \to \mathbb{R}$ is called differentiable if there exists a differentiable function $\tilde{f}\colon \mathbb{R}^n \to \mathbb{R}$ such that $\tilde{f}(\mathbf{x})\big|_M = f(\mathbf{x})$. The function $\tilde{f}$ is called an *extension of* f to $\mathbb{R}^n$.

Definition 5.2.5. *Let* $f\colon M \to \mathbb{R}$ *be a differentiable function,* $\mathbf{x} \in M$, *and* $\mathbf{v} \in T_{\mathbf{x}}M$. *Suppose that* $\boldsymbol{\alpha}\colon (-\epsilon, \epsilon) \to M$ *is a differentiable curve such that* $\boldsymbol{\alpha}(0) = \mathbf{x}$ *and* $\frac{d}{dt}\boldsymbol{\alpha}(0) = \mathbf{v}$. *Then the differential of* f *at* $\mathbf{x}$ *is the mapping* $df(\mathbf{x})\colon T_{\mathbf{x}}M \to \mathbb{R}$ *given by*

$$df(\mathbf{x})\mathbf{v} = \frac{d}{dt}f(\boldsymbol{\alpha}(t))\bigg|_{t=0}. \tag{5.3}$$

The differential is well-defined, since this definition is independent of the choice of $\boldsymbol{\alpha}$.

Theorem 5.2.6. *Let* $\mathbf{x} \in M \subset \mathbb{R}^n$, *and suppose that* $\boldsymbol{\alpha}\colon (-\epsilon, \epsilon) \to M$ *and* $\boldsymbol{\beta}\colon (-\epsilon, \epsilon) \to M$ *are differentiable curves such that* $\boldsymbol{\alpha}(0) = \boldsymbol{\beta}(0) = \mathbf{x}$ *and* $\frac{d}{dt}\boldsymbol{\alpha}(0) = \frac{d}{dt}\boldsymbol{\beta}(0) = \mathbf{v}$. *Then*

$$\frac{d}{dt}f(\boldsymbol{\alpha}(t))\bigg|_{t=0} = \frac{d}{dt}f(\boldsymbol{\beta}(t))\bigg|_{t=0}.$$

Proof. Since $M \subset \mathbb{R}^n$, we can write $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$ and

$$\begin{aligned}
\boldsymbol{\beta}(t) &= (\beta_1(t), \beta_2(t), \ldots, \beta_n(t)), \\
\boldsymbol{\alpha}(t) &= (\alpha_1(t), \alpha_2(t), \ldots, \alpha_n(t)).
\end{aligned}$$

Let $\tilde{f}$ be a differentiable extension of f to $\mathbb{R}^n$. Then, invoking the chain rule for differentiation,

$$\begin{aligned}
\frac{d}{dt} f(\boldsymbol{\alpha}(t)) \Big|_{t=0} &= \frac{d}{dt} \tilde{f}((\alpha_1(t), \ldots, \alpha_n(t))) \Big|_{t=0} = \left(\frac{\partial \tilde{f}}{\partial \alpha_1} \frac{d\alpha_1}{dt} + \ldots + \frac{\partial \tilde{f}}{\partial \alpha_n} \frac{d\alpha_n}{dt} \right) \Big|_{t=0} \\
&= \left\langle \nabla \tilde{f}(\mathbf{x}), \left(\frac{d\alpha_1}{dt} \Big|_{t=0}, \ldots, \frac{d\alpha_n}{dt} \Big|_{t=0} \right) \right\rangle = \left\langle \nabla \tilde{f}(\mathbf{x}), \frac{d}{dt} \boldsymbol{\alpha}(0) \right\rangle = \left\langle \nabla \tilde{f}(\mathbf{x}), \mathbf{v} \right\rangle.
\end{aligned}$$

A similar argument shows that

$$\frac{d}{dt} f(\boldsymbol{\beta}(t)) \Big|_{t=0} = \left\langle \nabla \tilde{f}(\mathbf{x}), \mathbf{v} \right\rangle,$$

and hence, $\frac{d}{dt} f(\boldsymbol{\beta}(t)) \big|_{t=0} = \frac{d}{dt} f(\boldsymbol{\alpha}(t)) \big|_{t=0}$. $\qquad\square$

This definition for the differential of f at $\mathbf{x}$ satisfies the basic properties that we associate with the idea of a derivative.

Theorem 5.2.7. *The differential of f at $\mathbf{x}$, $df(\mathbf{x})$, is a linear map from $T_{\mathbf{x}}M$ to $\mathbb{R}$.*

Proof. This is clear from the proof of Theorem 5.2.6, since, letting $a, b \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in T_{\mathbf{x}}M$ yields

$$df(\mathbf{x})(a\mathbf{v} + b\mathbf{w}) = \left\langle \nabla \tilde{f}, a\mathbf{v} + b\mathbf{w} \right\rangle = a \left\langle \nabla \tilde{f}, \mathbf{v} \right\rangle + b \left\langle \nabla \tilde{f}, \mathbf{w} \right\rangle = a\, df(\mathbf{x})\mathbf{v} + b\, df(\mathbf{x})\mathbf{w}$$

for some fixed differentiable extension $\tilde{f}$ of f. This proves the linearity of the differential. $\qquad\square$

In order to prove the above result it was necessary to use the gradient of $\tilde{f}$ where $\tilde{f}$ is a

differentiable extension of f. This is, in fact, not necessary, as now we can define the gradient of a function on a manifold. The following definition is given in [63].

Definition 5.2.8. *Let* $f \colon M \to \mathbb{R}$ *be a differentiable function. Then the gradient of* f *at* $\mathbf{x}$ *is a vector* $\nabla_M f(\mathbf{x}) \in T_{\mathbf{x}} M$ *such that*

$$df(\mathbf{x})\mathbf{v} = \langle \nabla_M f(\mathbf{x}), \mathbf{v} \rangle_{\mathbf{x}}$$

for all $\mathbf{v} \in T_{\mathbf{x}} M$.

The map $df(\mathbf{x})$ is independent of the extension $\tilde{f}$, which can be seen from the following result.

Lemma 5.2.9. *The gradient of* f *at* $\mathbf{x}$, $\nabla_M f(\mathbf{x})$, *is unique when it exists.*

Proof. Suppose that $\mathbf{v}_1, \mathbf{v}_2 \in T_{\mathbf{x}} M$ are such that

$$df(\mathbf{x})\mathbf{v} = \langle \mathbf{v}_1, \mathbf{v} \rangle_{\mathbf{x}}, \quad \text{and} \quad df(\mathbf{x})\mathbf{v} = \langle \mathbf{v}_2, \mathbf{v} \rangle_{\mathbf{x}}$$

for all $\mathbf{v} \in T_{\mathbf{x}} M$. Then

$$\langle \mathbf{v}_1 - \mathbf{v}_2, \mathbf{v} \rangle_{\mathbf{x}} = 0$$

for all $\mathbf{v} \in T_{\mathbf{x}} M$. Since $\mathbf{v}_1 - \mathbf{v}_2 \in T_{\mathbf{x}} M$, this yields

$$\langle \mathbf{v}_1 - \mathbf{v}_2, \mathbf{v}_1 - \mathbf{v}_2 \rangle_{\mathbf{x}} = \| \mathbf{v}_1 - \mathbf{v}_2 \|^2 = 0.$$

Thus $\mathbf{v}_1 = \mathbf{v}_2$. $\qquad\square$

Given the above definitions, one can construct a gradient descent algorithm for minimizing a function f on a manifold M. Analogous to equation (5.1), this is given by

$$\frac{d}{dt}\mathbf{x}(t) = -\nabla_M f(\mathbf{x}(t)), \quad \mathbf{x}(0) = \mathbf{x}_0. \tag{5.4}$$

Minimizing f by solving (5.4) is called the *steepest descent method* for a function f on M, and, following the proof of Theorem 5.1.2, it can be shown that, if $\lim_{t \to \infty} \mathbf{x}(t)$ exists and equals $\mathbf{x}'$, then $\nabla_M f(\mathbf{x}') = 0$.

It can be shown that the gradient of f on M at $\mathbf{x}$ is the projection of the Euclidean gradient of $\tilde{f}$ at $\mathbf{x}$ onto the tangent space of M at $\mathbf{x}$. If the manifold M has an implicit representation of the form $M = \{\mathbf{x} \in \mathbb{R}^n \mid \Phi(\mathbf{x}) = 0\}$ for some function $\Phi \colon \mathbb{R}^n \to \mathbb{R}$ such that $\nabla \Phi(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in M$, then

$$\nabla_M f(\mathbf{x}) = \nabla f(\mathbf{x}) - \frac{\langle \nabla f(\mathbf{x}), \nabla \Phi(\mathbf{x}) \rangle}{|\nabla \Phi(\mathbf{x})|^2} \nabla \Phi(\mathbf{x}).$$
(5.5)

For this reason, (5.4) is often also referred to as the *gradient projection method*.

We now restrict to a special class of manifolds, the regular surfaces.

Definition 5.2.10 (Thorpe [69], p. 16). *An m-dimensional manifold $M \subset \mathbb{R}^{m+1}$ is called a* regular surface *if there exists a function* $\Phi \colon \mathbb{R}^{m+1} \to \mathbb{R}$ *that satisfies the following properties:*

1. $\Phi(\mathbf{x}) = 0$ if and only if $\mathbf{x} \in M$,

2. $\Phi \in C(\mathbb{R}^{m+1})$,

3. For each $\mathbf{x} \in M$ there exists a neighborhood $U_{\mathbf{x}} \subset \mathbb{R}^{m+1}$ of $\mathbf{x}$ such that $\Phi \in C^1(U_{\mathbf{x}})$,

4. $\nabla \Phi(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in M$.

We denote the set of all regular surfaces in $\mathbb{R}^n$ by $\mathfrak{M}(\mathbb{R}^n)$.

A function Φ that represents a regular surface M is called a *level set representation* of M or a *level set function* for M, because the surface M is the zero level set of Φ. Level set representations of regular surfaces are not unique, but each level set function corresponds to exactly one regular surface via its zero level set.

Note that a regular surface M is a closed set, since it is the inverse image of the closed set $\{0\} \subset \mathbb{R}$ under its level set representation Φ, which is continuous. The types of surfaces that arise in our application are not, in general, connected, but each connectedness component of these surfaces is bounded. Thus, by the Heine-Borel Theorem (cf. [54]), each of the connectedness components of M is a compact set.

Since a regular surface is a manifold, its tangent space is defined in Definition 5.2.2. Along with the tangent space at a point for a regular surface, it is useful to define a surface normal.

Definition 5.2.11. *Let* $M \subset \mathbb{R}^n$ *be a regular surface, then a* normal vector to the surface at the point $\mathbf{x} \in M$ *is a non-zero vector,* $\hat{\mathbf{n}}(\mathbf{x}) \in \mathbb{R}^n$, *such that*

$$\langle \hat{\mathbf{n}}(\mathbf{x}), \mathbf{v} \rangle = 0$$

for all $\mathbf{v} \in T_{\mathbf{x}} M$.

A regular surface in $\mathbb{R}^n$ admits exactly two normal vectors of unit length at each point $\mathbf{x} \in M$ - the inward and outward normals. These differ only by a sign.

Since a regular surface is topologically closed in $\mathbb{R}^n$, we can define the Euclidean distance from a point to a regular surface. This is given by

$$\rho(\mathbf{x}, M) = \min_{\mathbf{v} \in M} \|\mathbf{x} - \mathbf{v}\| \tag{5.6}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^n$. Since each connectedness component of M is compact, the minimum is attained, and hence (5.6) is a well-defined mapping. Thus we can define the canonical level set representation for a regular surface.

Definition 5.2.12. *Let* $M \subset \mathbb{R}^n$ *be a regular surface with level set representation* Ψ. *Then*

the signed distance function *to M is the function*

$$\Phi(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}, M) & \Psi(\mathbf{x}) \geq 0 \\ \\ -\rho(\mathbf{x}, M) & \Psi(\mathbf{x}) < 0. \end{cases} \tag{5.7}$$

Note that the sign of the signed distance function is defined in terms of a given level set representation of the regular surface M. If another level set representation Υ is chosen with opposite orientation, i.e. $\{\mathbf{x} \in \mathbb{R}^n \mid \Upsilon(\mathbf{x}) < 0\} = \{\mathbf{x} \in \mathbb{R}^n \mid \Psi(\mathbf{x}) > 0\}$, then the signed distance function associated with Υ is just the negative of the signed distance function associated with Ψ. These two signed distance functions have the same continuity and differentiability properties, so we will not distinguish between them.

A lemma is required before it can be shown that the signed distance function is an appropriate level set representation of a regular surface M. Denote by $D_{\mathbf{v}}\Psi(\mathbf{x})$ the directional derivative of Ψ at $\mathbf{x}$ in the direction $\mathbf{v}$.

Lemma 5.2.13. *Let $M \subset \mathbb{R}^n$ be a regular surface with level set representation Ψ, and let $\mathbf{x} \in M$. If $\mathbf{v} \in \mathbb{R}^n$ is a vector with unit length such that $D_{\mathbf{v}}\Psi(\mathbf{x}) \neq 0$, then there exists $h_0 > 0$ such that $\Psi(\mathbf{x} + h\mathbf{v}) < 0$ or $\Psi(\mathbf{x} + h\mathbf{v}) > 0$ for all $0 < h < h_0$.*

Proof. Let $\mathbf{v} \in \mathbb{R}^n$ be a vector of unit length. Since a level set representation of a regular surface M is differentiable on a neighborhood of each point of M, $D_{\mathbf{v}}\Psi(\mathbf{x})$ exists for each $\mathbf{x} \in M$, and we assume that $D_{\mathbf{v}}\Psi(\mathbf{x}) \neq 0$.

Suppose that there exists $h_0 > 0$ such that $\Psi(\mathbf{x} + h\mathbf{v}) = 0$ for all $0 < h < h_0$. Then

$$D_{\mathbf{v}}\Psi(\mathbf{x}) = \lim_{h \to 0} \frac{\Psi(\mathbf{x} + h\mathbf{v}) - \Psi(\mathbf{x})}{h} = \lim_{h \to 0} \frac{0}{h} = 0,$$

contrary to hypothesis.

Now suppose that there exist two sequences of positive real numbers, $\{h_k\}$ and $\{\eta_k\}$, such that $h_k \to 0$, $\eta_k \to 0$, $\Psi(\mathbf{x} + h_k\mathbf{v}) > 0$ for all $k \in \mathbb{N}$, and $\Psi(\mathbf{x} + \eta_k\mathbf{v}) < 0$ for all $k \in \mathbb{N}$. Since the directional derivative exists,

$$D_\mathbf{v}\Psi(\mathbf{x}) = \lim_{k\to\infty} \frac{\Psi(\mathbf{x} + h_k\mathbf{v}) - \Psi(\mathbf{x})}{h_k} = \lim_{k\to\infty} \frac{\Psi(\mathbf{x} + h_k\mathbf{v})}{h_k} \geq 0$$

$$D_\mathbf{v}\Psi(\mathbf{x}) = \lim_{k\to\infty} \frac{\Psi(\mathbf{x} + \eta_k\mathbf{v}) - \Psi(\mathbf{x})}{\eta_k} = \lim_{k\to\infty} \frac{\Psi(\mathbf{x} + \eta_k\mathbf{v})}{\eta_k} \leq 0.$$

Thus $D_\mathbf{v}\Psi(\mathbf{x}) = 0$, contrary to hypothesis.

Since no two such sequences $\{h_k\}$ and $\{\eta_k\}$ can exist, every sequence $\{h_k\}$ tending to 0 must satisfy either $\Psi(\mathbf{x} + h_k\mathbf{v}) < 0$ or $\Psi(\mathbf{x} + h_k\mathbf{v}) > 0$ for all sufficiently large k. This is equivalent to the claim. $\square$

Lemma 5.2.14. *Let M be a regular surface with level set representation Ψ and Φ be the signed distance function given by* (5.7). *Then, for every $\mathbf{x} \in M$, $\nabla\Phi(\mathbf{x})$ exists and $|\nabla\Phi(\mathbf{x})| = 1$.*

Proof. It will be shown that the directional derivative at $\mathbf{x}$ in the direction $\mathbf{v}$ exists for an arbitrary unit vector $\mathbf{v}$ and every $\mathbf{x} \in M$, but first note that Φ is continuous on $\mathbb{R}^n$.

Let $\mathbf{v} \in \mathbb{R}^n$ be a vector of unit length, and let $\mathbf{x} \in M$. Suppose that $D_\mathbf{v}\Psi(\mathbf{x}) = 0$. Then either $\Psi(\mathbf{x})$ is locally constant along the line in the direction of $\mathbf{v}$ through $\mathbf{x}$ or $\Psi(\mathbf{x})$ is a local extremum along the line in the direction of $\mathbf{v}$ through $\mathbf{x}$. If $\Psi(\mathbf{x})$ is a local extremum of Ψ along the line through $\mathbf{x}$ in the direction of $\mathbf{v}$, then, since $\Psi(\mathbf{x}) = 0$, we have that there exists $h_0 > 0$ such that $\Psi(\mathbf{x} - h\mathbf{v})$ is increasing (or decreasing) to 0 as $h \searrow 0$ ($h < h_0$) and $\Psi(\mathbf{x} + h\mathbf{v})$ is decreasing (or increasing) as $h \searrow 0$ ($h < h_0$). Thus Ψ is either positive on both sides of the surface or negative on both sides of the surface, contradicting the definition of a level set representation of a regular surface.

If $\Psi(\mathbf{x})$ is locally constant along the line through $\mathbf{x}$ in the direction of $\mathbf{v}$, then there exists $h_0 > 0$ such that $\Psi(\mathbf{x} + h\mathbf{v}) = 0$ for all $0 < h < h_0$. In this case, $\Phi(\mathbf{x} + h\mathbf{v}) = 0$ for all

$0 < h < h_0$, so $D_{\mathbf{v}}\Phi(\mathbf{x})$ exists and is equal to 0.

Now assume that $D_{\mathbf{v}}\Psi(\mathbf{x}) \neq 0$. Then, by Lemma 5.2.13, there exists $h > 0$ such that either $\Psi(\mathbf{x} + h\mathbf{v}) > 0$ or $\Psi(\mathbf{x} + h\mathbf{v}) < 0$ for all $0 < h < h_0$. Without loss of generality we assume that $\Psi(\mathbf{x} + h\mathbf{v}) > 0$, and note that Ψ and Φ have the same sign by construction.

Then, applying the generalized triangle inequality for the distance between points and closed sets, we have

$$D_{\mathbf{v}}\Phi(\mathbf{x}) = \lim_{h \to 0} \frac{\Phi(\mathbf{x} + h\mathbf{v}) - \Phi(\mathbf{x})}{h} = \lim_{h \to 0} \frac{1}{h} \left(\rho(\mathbf{x} + h\mathbf{v}, M) - \rho(\mathbf{x}, M) \right)$$

$$= \lim_{h \to 0} \frac{1}{h} \rho(\mathbf{x} + h\mathbf{v}, M) \leq \lim_{h \to 0} \frac{1}{h} \|\mathbf{x} - \mathbf{x} + h\mathbf{v}\| = \lim_{h \to 0} \frac{1}{h} h \|\mathbf{v}\| = 1$$

Thus at each point on the surface M, the directional derivative in the direction of an arbitrary unit vector exists and is less than or equal to 1.

Since all the directional derivatives at $\mathbf{x}$ exist, $\nabla\Phi(\mathbf{x})$ exists. Furthermore, taking $\mathbf{v} = \nabla\Phi/|\nabla\Phi|$ yields $|D_{\mathbf{v}}\Phi(\mathbf{x})| = |\nabla\Phi|$, and hence $|\nabla\Phi| \leq 1$. Thus, since $|\nabla\Phi(\mathbf{x})| = \max_{\mathbf{v}} |D_{\mathbf{v}}\Phi(\mathbf{x})|$ (cf. [67], Theorem 16.28), it remains to show that there exists a vector $\mathbf{v}$ such that $|D_{\mathbf{v}}\Phi(\mathbf{x})| \geq 1$.

Let $\hat{\mathbf{n}}$ be the unit normal to the surface M at $\mathbf{x}$. Since the unit normal is orthogonal to the tangent space of M at $\mathbf{x}$, we have that $\rho(\mathbf{x}, \mathbf{x} + h\hat{\mathbf{n}}) = h$ for all sufficiently small h. Therefore

$$|D_{\hat{\mathbf{n}}}\Phi(\mathbf{x})| = \left| \lim_{h \to 0} \frac{\Phi(\mathbf{x} + h\hat{\mathbf{n}}) - \Phi(\mathbf{x})}{h} \right| = \left| \lim_{h \to 0} \frac{1}{h} \left(\rho(\mathbf{x} + h\hat{\mathbf{n}}, M) - \rho(\mathbf{x}, M) \right) \right|$$

$$= \left| \lim_{h \to 0} \frac{1}{h} \rho(\mathbf{x} + h\hat{\mathbf{n}}, M) \right| = \left| \lim_{h \to 0} \frac{1}{h} \rho(\mathbf{x}, \mathbf{x} + h\hat{\mathbf{n}}) \right| = \lim_{h \to 0} \frac{1}{h} h = 1,$$

and thus $|\nabla\Phi(\mathbf{x})| = 1$. $\qquad \square$

Theorem 5.2.15. *Let $M \subset \mathbb{R}^n$ be a regular surface with level set representation $\Psi(\mathbf{x})$. Then the signed distance function Φ given in (5.7) is also a level set representation for M.*

This comes directly from the definition of the signed distance function, Lemma 5.2.14, and Definition 5.2.10.

5.3 Regular Surface Evolutions and Level Set Geometry

In this section we define a regular surface evolution using a level set representation and introduce the geometric concepts associated with level set functions.

Definition 5.3.1 (Solem, Overgaard, [63]). *Let M_0 be a regular surface, and let $\Phi(\mathbf{x}, t): \mathbb{R}^n \times (-\epsilon, \epsilon) \to \mathbb{R}$, for some $\epsilon > 0$, be a family of level set functions such that $\Phi(\mathbf{x}, t)$ is differentiable in t for all fixed $\mathbf{x} \in \mathbb{R}^n$, $\nabla\Phi(\mathbf{x}, t) \neq 0$ for all $\mathbf{x} \in \{\mathbf{y} \in \mathbb{R}^n \mid \Phi(\mathbf{y}, t) = 0\}$, and $\Phi(\mathbf{x}, 0)$ is a level set representation for M_0. Then the regular surface evolution through M_0 corresponding to $\Phi(\mathbf{x}, t)$ is the family of surfaces given by*

$$M(t) = \{\mathbf{x} \in \mathbb{R}^n \mid \Phi(\mathbf{x}, t) = 0\}.$$

Note that these surfaces are uniquely defined by Φ, since each level set function $\Phi(\cdot, t)$ uniquely defines a regular surface by its zero level set. Thus we will often associate the family of surfaces $\{M(t)\}$ directly with the family of level set functions.

Given the definition of a regular surface evolution, it will be useful to introduce some of the geometric concepts of level set functions that will be used in proving some of following results.

Theorem 5.3.2. *Let M be a regular surface with level set representation $\Phi(\mathbf{x})$. Then the unit surface normals at $\mathbf{x} \in M$, $\hat{\mathbf{n}}(\mathbf{x})$, are given by*

$$\hat{\mathbf{n}}(\mathbf{x}) = \pm \frac{\nabla\Phi(\mathbf{x})}{|\nabla\Phi(\mathbf{x})|}. \tag{5.8}$$

Proof. It is clear that the definition in (5.8) yields unit vectors, so it is only left to show that

$\nabla\Phi \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in T_{\mathbf{x}}M$.

What follows is based on the proof given by Thorpe, Lemma 1 of Section 3 [69] (see also Section 5, Theorem 2). Let $\mathbf{x} \in M$, and let $T_{\mathbf{x}}M$ be the tangent space to M at $\mathbf{x}$. Suppose that $\mathbf{v} \in T_{\mathbf{x}}M$, then $\mathbf{v}$ has the form $\mathbf{v} = \dot{\boldsymbol{\alpha}}(0)$, where $\dot{\boldsymbol{\alpha}} = d\boldsymbol{\alpha}/dt$, for some $\epsilon > 0$ and some differentiable function $\boldsymbol{\alpha}\colon (-\epsilon, \epsilon) \to M$ with $\boldsymbol{\alpha}(0) = \mathbf{x}$. Differentiating both sides of $\Phi(\boldsymbol{\alpha}(t)) = 0$ with respect to t at $t = 0$ yields

$$0 = \frac{d}{dt}\left(\Phi \circ \boldsymbol{\alpha}\right)(0) = \nabla\Phi(\boldsymbol{\alpha}(0)) \cdot \dot{\boldsymbol{\alpha}}(0) = \nabla\Phi(\mathbf{x}) \cdot \mathbf{v}.$$

Thus $\nabla\Phi(\mathbf{x}) \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in T_{\mathbf{x}}M$, which is what was to be shown. $\qquad\square$

Suppose $\Phi(\mathbf{x}, t)$ is a regular surface evolution, and let $\boldsymbol{\alpha}\colon (-\epsilon, \epsilon) \to \mathbb{R}^n$ be a differentiable curve such that $\boldsymbol{\alpha}(t) \in M(t)$ for all t. Then the equation $\Phi(\boldsymbol{\alpha}(t), t) = 0$ holds for all t. Differentiating with respect to t yields

$$0 = \frac{\partial\Phi}{\partial\boldsymbol{\alpha}} \cdot \frac{d\boldsymbol{\alpha}}{dt} + \frac{\partial\Phi}{\partial t} = \frac{\partial\Phi}{\partial t} + \dot{\boldsymbol{\alpha}} \cdot \nabla\Phi. \tag{5.9}$$

This is called the *level set equation*. Denote $v_n = \dot{\boldsymbol{\alpha}} \cdot \hat{\mathbf{n}}$, then the evolution given by equation (5.9) can be described by an evolution in the direction normal to the surface by

$$\frac{\partial\Phi}{\partial t} + \dot{\boldsymbol{\alpha}} \cdot \nabla\Phi = \frac{\partial\Phi}{\partial t} + \dot{\boldsymbol{\alpha}}|\nabla\Phi| \cdot \frac{\nabla\Phi}{|\nabla\Phi|} = \frac{\partial\Phi}{\partial t} + (\dot{\boldsymbol{\alpha}} \cdot \hat{\mathbf{n}})|\nabla\Phi| = \frac{\partial\Phi}{\partial t} + v_n|\nabla\Phi|,$$

which yields the normal evolution of equation (5.9) reformulated as

$$\frac{\partial\Phi}{\partial t} + v_n|\nabla\Phi| = 0. \tag{5.10}$$

The quantity v_n is called the *normal velocity*. The tangential components of a level set evolution correspond only to reparameterizations of each surface (or a different family of level set functions representing the same regular surface evolution), so, even though evolutions

as defined in (5.10) have velocity only in the normal direction, they are actually completely general.

The fundamental property of a regular surface evolution, $\{M(t)\}$, is that it is independent of the level set function $\Phi(\mathbf{x}, 0)$ chosen to represent $M(0)$.

Theorem 5.3.3 (Solem, Overgaard, [63] Lemma 1). *Let $M(t)$ be a regular surface evolution with level set representation $\Phi(\mathbf{x}, t)$, and let $\boldsymbol{\alpha} \colon (-\epsilon, \epsilon) \to \mathbb{R}^n$ be a differentiable path such that $\boldsymbol{\alpha}(0) = \mathbf{x}_0 \in M(0)$ and $\boldsymbol{\alpha}(t) \in M(t)$ for all $t \in (-\epsilon, \epsilon)$. Then the normal velocity at $\mathbf{x}_0$, $v_n(\mathbf{x}_0)$, is independent of $\boldsymbol{\alpha}$ and of the level set representation of the evolution, in particular it is independent of $\Phi(\cdot, 0)$.*

Proof. The normal velocity at $\mathbf{x}_0$ is given by $v_n = \dot{\boldsymbol{\alpha}}(0) \cdot \hat{\mathbf{n}}(\mathbf{x}_0)$, which is clearly independent of Φ.

On the other hand, equation (5.10) yields

$$v_n = -\frac{\Phi_t(\mathbf{x}_0, 0)}{|\nabla \Phi(\mathbf{x}_0, 0)|}, \tag{5.11}$$

which is independent of the curve $\boldsymbol{\alpha}$. Note that $\nabla \Phi(\mathbf{x}_0, 0) \neq 0$ by the definition of a level set representation of a regular surface evolution, so the division on the right-hand side is allowed. $\qquad \square$

There are two reasons this result is important. First of all, it shows that the normal velocity of a regular surface evolution is an *intrinsic* property of the evolution itself and not a property of any of the representations of the evolution. Secondly, it allows us to choose any level set representation for a given evolution. In particular, given an initial surface $M(0)$, the regular surface evolution $M(t)$ can be represented by $\Phi(\mathbf{x}, t)$ where $\Phi(\mathbf{x}, 0)$ is the signed distance function from $M(0)$. This will be useful in what follows.

5.4 The Tangent Space to the Surface Manifold

In this section it will be shown that to each regular surface one can associate an entire family of regular surfaces and that a formal gradient can be defined on this family of regular surfaces.

Definition 5.4.1 (Solem, Overgaard, [63] section 4). *Let M_0 be a regular surface, then the surface manifold, $\mathcal{M}_0$, associated with M_0 is the set of all regular surfaces that can be obtained from M_0 by a regular surface evolution.*

The surface manifold is closely related to the families of level set functions that can be evolved from the original level set representation of M_0.

The surface manifold $\mathcal{M}_0$ is not actually a manifold in the sense of Definition 5.2.1. Nonetheless it is possible to define a tangent space to a surface manifold. The first step in proving this is to show that the surface manifold associated with M_0 induces an equivalence relation.

Theorem 5.4.2. *Write $M \sim M_0$ if M can be produced from M_0 by a regular surface evolution, i.e. if $M \in \mathcal{M}_0$. Then $\sim$ is an equivalence relation on $\mathfrak{M}(\mathbb{R}^n)$.*

This result is stated, but not proven, in [63]. We prove it here.

Proof. First of all, it is clear that $M_0 \sim M_0$, since M_0 can be produced from itself by the constant regular surface evolution, i.e. $\Phi(\mathbf{x}, t) = \Phi(\mathbf{x}, 0)$ for all $t \in (-\epsilon, \epsilon)$. Thus $\sim$ is a reflexive relation.

Now suppose that $M \sim M_0$ and that $\Phi(\mathbf{x}, t)$ is the regular surface evolution which yields M from M_0 at time T. Then $\Phi(\mathbf{x}, T - t)$ is a regular surface evolution which produces M_0 from M. This demonstrates that $\sim$ is a symmetric relation.

Lastly, it must be shown that $\sim$ is a transitive relation. Suppose that M_0, M_1, and M_2 are regular surfaces such that $M_0 \sim M_1$ and $M_1 \sim M_2$. Then, without loss of generality, there

exist regular surface evolutions given by the level set functions $\Phi(\mathbf{x}, t)$ and $\Psi(\mathbf{x}, t)$ such that

$$M_0 = \{\mathbf{x} \in \mathbb{R}^n \mid \Phi(\mathbf{x}, 0) = 0\}$$

$$M_1 = \left\{\mathbf{x} \in \mathbb{R}^n \mid \Phi\left(\mathbf{x}, \frac{1}{2}\right) = 0\right\} = \left\{\mathbf{x} \in \mathbb{R}^n \mid \Psi\left(\mathbf{x}, \frac{1}{2}\right) = 0\right\}$$

$$M_2 = \{\mathbf{x} \in \mathbb{R}^n \mid \Psi(\mathbf{x}, 1) = 0\}.$$

Define

$$\Upsilon(\mathbf{x}, t) = \begin{cases} \Phi\left(\mathbf{x}, -2t^2 + 2t\right) & t \in \left[0, \frac{1}{2}\right] \\[2mm] \Psi\left(\mathbf{x}, 2t^2 - 2t + 1\right) & t \in \left[\frac{1}{2}, 1\right], \end{cases} \tag{5.12}$$

then

$$\begin{aligned} \Upsilon(\mathbf{x}, 0) &= \Phi(\mathbf{x}, -2(0) + 2(0)) &= \Phi(\mathbf{x}, 0) \\[1mm] \Upsilon\left(\mathbf{x}, \tfrac{1}{2}\right) &= \Phi\left(\mathbf{x}, -2\left(\tfrac{1}{2}\right)^2 + 2\left(\tfrac{1}{2}\right)\right) &= \Phi\left(\mathbf{x}, \tfrac{1}{2}\right) \\[1mm] \Upsilon\left(\mathbf{x}, \tfrac{1}{2}\right) &= \Psi\left(\mathbf{x}, 2\left(\tfrac{1}{2}\right)^2 - 2\left(\tfrac{1}{2}\right) + 1\right) &= \Psi\left(\mathbf{x}, \tfrac{1}{2}\right) \\[1mm] \Upsilon\left(\mathbf{x}, 1\right) &= \Psi(\mathbf{x}, 2(1)^2 - 2(1) + 1) &= \Psi(\mathbf{x}, 1). \end{aligned}$$

Thus Υ continuously evolves M_0 to M_2.

It is only left to be shown that Υ is differentiable with respect to t. Since Φ and Ψ are differentiable,

$$\Upsilon_t(\mathbf{x}, t) = \begin{cases} (-4t + 2)\, \Phi_t(\mathbf{x}, -2t^2 + 2t) & t \in \left[0, \frac{1}{2}\right] \\[2mm] (4t - 2)\, \Psi_t(\mathbf{x}, 2t^2 - 2t + 1) & t \in \left[\frac{1}{2}, 1\right]. \end{cases}$$

This function is clearly well-defined for $t \in \left[0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right]$. It is also well-defined for $t = 1/2$, since

$$\begin{aligned} \lim_{t \to 1/2^-} \Upsilon_t(\mathbf{x}, t) &= \left(-4\left(\tfrac{1}{2}\right) + 2\right) \Phi_t\left(\mathbf{x}, -2\left(\tfrac{1}{2}\right)^2 + 2\left(\tfrac{1}{2}\right)\right) &= 0 \\[1mm] \lim_{t \to 1/2^+} \Upsilon_t(\mathbf{x}, t) &= \left(4\left(\tfrac{1}{2}\right) - 2\right) \Psi_t\left(\mathbf{x}, 2\left(\tfrac{1}{2}\right)^2 - 2\left(\tfrac{1}{2}\right) + 1\right) &= 0 \end{aligned}$$

Applying L'Hospital's Rule, it can be shown that this suffices for Υ to be time-differentiable. Thus Υ is a regular surface evolution from M_0 to M_2, which shows that $\sim$ is transitive.

As $\sim$ is reflexive, symmetric, and transitive, it is an equivalence relation. $\square$

Since M_0 is any arbitrary regular surface in $\mathcal{M}_0$, we will simply write $\mathcal{M}$ to be a surface manifold and not specifically associate it with any one of its elements.

Before defining the tangent space to a surface manifold, we first define the normal velocity for a surface evolution. The next three definitions are all due to Solem and Overgaard [63,64].

Definition 5.4.3. *Let* $t \mapsto M(t)$ *be a regular surface evolution with level set representation* $\Phi(\mathbf{x}, t)$. *Then the normal velocity of the surface evolution, denoted* $\dot{M}$, *is given by*

$$\dot{M}(t) = -\left. \frac{\Phi_t(\mathbf{x}, t)}{|\nabla \Phi(\mathbf{x}, t)|} \right|_{\mathbf{x} \in M(t)}. \tag{5.13}$$

By Theorem 5.3.3 and its proof, this is a well-defined quantity that is independent of Φ. Since $\Phi(\mathbf{x}, t)$ is continuously differentiable with respect to t and is continuously differentiable with respect to $\mathbf{x}$ on M, $\dot{M}(t)$ is a continuous function of the *spatial* variable $\mathbf{x}$ on $M(t)$ for each t.

We now formally define a tangent vector to a surface manifold.

Definition 5.4.4. *Let* $\mathcal{M}$ *be a surface manifold, and let* $M_0 \in \mathcal{M}$. *A tangent vector to the surface manifold* $\mathcal{M}$ *at* M_0 *is a normal velocity* $\dot{M}$ *associated with any regular surface evolution* $M(t)$ *such that* $M(0) = M_0$.

As noted above, tangent vectors to $\mathcal{M}$ at M are continuous functions on M. Also, the condition that $M(0) \in \mathcal{M}$ is equivalent to the condition that $M(t) \in \mathcal{M}$ for any surface in the evolution, but $\dot{M}$ is a tangent vector to $\mathcal{M}$ at the "point" $M(0)$.

Definition 5.4.5. *The tangent space to a surface manifold* $\mathcal{M}$ *at a point* M, *denoted* $T_M\mathcal{M}$, *is the set of all tangent vectors to* $\mathcal{M}$ *at* M.

A natural question to ask is whether or not there exist any tangent vectors to an arbitrary

surface manifold $\mathcal{M}$ at an arbitrary point M. It turns out that the existence of many tangent vectors can be guaranteed. The following theorem can be found in [63], and we present a simplified variation of the proof given there.

Theorem 5.4.6. *Let $\mathcal{M}$ be a surface manifold, and let $M \in \mathcal{M}$. If $v \in C^2(M)$, then v is a tangent vector to $\mathcal{M}$ at M.*

Proof. Let $\tilde{v} \in C^2(\mathbb{R}^n)$ be an extension of $v \in C^2(M)$, and let $\Phi_0(\mathbf{x})$ be the signed distance function to M. Define

$$\Phi(\mathbf{x}, t) = \Phi_0(\mathbf{x}) - t\tilde{v}(\mathbf{x})|\nabla\Phi_0(\mathbf{x})| = \Phi_0(\mathbf{x}) - t\tilde{v}(\mathbf{x}),$$

and let $M(t) = \{\mathbf{x} \in \mathbb{R}^n \mid \Phi(\mathbf{x}, t) = 0\}$. Recall that $|\nabla\Phi_0| = 1$. Since $\Phi_t(\mathbf{x}, t) = -\tilde{v}(\mathbf{x})$, there exists $\epsilon > 0$ such that $\Phi(\mathbf{x}, t) \in C^1(-\epsilon, \epsilon)$ for each $\mathbf{x} \in \mathbb{R}^n$. Now

$$\nabla\Phi(\mathbf{x}, t) = \nabla\Phi_0(\mathbf{x}) - t\nabla\tilde{v}(\mathbf{x}).$$

Since $\nabla\Phi_0(\mathbf{x}, 0) \neq 0$ for $\mathbf{x} \in M$ and $\nabla\tilde{v} \in C^1(\mathbb{R}^n)$, there exists $\epsilon > 0$ such that $\nabla\Phi(\mathbf{x}, t) \neq 0$ for $\mathbf{x} \in M(t)$ for all $t \in (-\epsilon, \epsilon)$, which means that $M(t)$ is a regular evolution for $t \in (-\epsilon, \epsilon)$ with $M(0) = M$. Furthermore

$$\dot{M}(0) = -\frac{\Phi_t(\mathbf{x}, 0)}{|\nabla\Phi(\mathbf{x}, 0)|} = -\Phi_t(\mathbf{x}, 0) = \tilde{v}\,|_{\mathbf{x} \in M} = v.$$

Thus v is a tangent vector. $\qquad\square$

This shows, in particular, that $C^2(M) \subset T_M\mathcal{M} \subset C(M)$, as it was already noted (see comments following Def. 5.4.3) that normal velocities are continuous functions on M.

Now we can show that the tangent space to $\mathcal{M}$ at M is a vector space.

Theorem 5.4.7. *Let $\mathcal{M}$ be a surface manifold, and let $M \in \mathcal{M}$. Then $T_M\mathcal{M}$ is a vector*

space.

This is stated in [63] without proof, so we prove it here.

Proof. The proof follows that of Theorem 5.2.3. Let $\dot{M}_1, \dot{M}_2 \in T_M\mathcal{M}$. It must be shown that $\dot{M}_1 + \dot{M}_2 \in T_M\mathcal{M}$, that $c\dot{M}_1 \in T_M\mathcal{M}$ for all $c \in \mathbb{R}$, and finally that the zero vector is an element of $T_M\mathcal{M}$. All other properties of vector spaces will follow immediately.

First of all, it is clear that the function $\dot{M} \equiv 0$ is an element of $T_M\mathcal{M}$, since this is the velocity associated with the constant surface evolution.

Now suppose that $\dot{M}(0) \in T_M\mathcal{M}$ is the normal velocity for some regular surface evolution $M(t)$ at $t = 0$ and that $\boldsymbol{\alpha}\colon (-\epsilon, \epsilon) \to \mathbb{R}^n$ is a differentiable path such that $\boldsymbol{\alpha}(0) = \mathbf{x} \in M(0)$ and $\boldsymbol{\alpha}(t) \in M(t)$ for all t. Then, by (5.13),

$$\dot{M}(0) = -\frac{\Phi_t(\mathbf{x}, 0)}{|\nabla\Phi(\mathbf{x}, 0)|} = \dot{\boldsymbol{\alpha}}(0) \cdot \hat{\mathbf{n}}(\mathbf{x}), \tag{5.14}$$

which implies that

$$c\dot{M}(0) = c\dot{\boldsymbol{\alpha}}(0) \cdot \hat{\mathbf{n}}(\mathbf{x}) = \frac{d}{dt}\boldsymbol{\alpha}(ct)\bigg|_{t=0} \cdot \hat{\mathbf{n}}(\mathbf{x}).$$

Since $\boldsymbol{\alpha}(ct)$ is a differentiable curve such that $\boldsymbol{\alpha}(c(0)) = \mathbf{x}$, $c\dot{M}(0)$ is also a normal velocity for the regular surface evolution associated with $\boldsymbol{\alpha}(ct)$. Thus $c\dot{M}(0) \in T_M\mathcal{M}$.

Similarly, suppose that $M_1(t)$ and $M_2(t)$ are regular surface evolutions in $\mathcal{M}$ such that $M_1(0) = M_2(0)$ and that $\dot{M}_1(0)$ and $\dot{M}_2(0)$ are tangent vectors to $\mathcal{M}$ at M_1 and M_2, respectively. Let $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\colon (-\epsilon, \epsilon) \to \mathbb{R}^n$ be differentiable paths such that $\boldsymbol{\alpha}_1(0) = \boldsymbol{\alpha}_2(0) = \mathbf{x} \in M_1(0) = M_2(0)$ and such that $\boldsymbol{\alpha}_1(t) \in M_1(t)$ and $\boldsymbol{\alpha}_2(t) \in M_2(t)$ for all $t \in (-\epsilon, \epsilon)$. Then

$$\frac{1}{2}\left(\dot{M}_1(0) + \dot{M}_2(0)\right) = \frac{1}{2}\left(\dot{\boldsymbol{\alpha}}_1(0) \cdot \hat{\mathbf{n}}(\mathbf{x}) + \dot{\boldsymbol{\alpha}}_2(0) \cdot \hat{\mathbf{n}}(\mathbf{x})\right) = \frac{d}{dt}\frac{1}{2}\left(\boldsymbol{\alpha}_1(t) + \boldsymbol{\alpha}_2(t)\right)\bigg|_{t=0} \cdot \hat{\mathbf{n}}(\mathbf{x}).$$

Since $\frac{1}{2}(\alpha_1(t) + \alpha_2(t))$ is a differentiable curve such that $\frac{1}{2}(\alpha_1(0) + \alpha_2(0)) = \mathbf{x}$, we have $\frac{1}{2}\left(\dot{M}_1(0) + \dot{M}_2(0)\right) \in T_M\mathcal{M}$. Since constant multiples of elements of $T_M\mathcal{M}$ are in $T_M\mathcal{M}$, $\dot{M}_1(0) + \dot{M}_2(0) \in T_M\mathcal{M}$, which is what was to be shown. $\qquad\square$

Since each connectedness component of each surface is assumed to be bounded, there is an inner product on the tangent space to a surface manifold, $\mathcal{M}$, at each point $M \in \mathcal{M}$. In what follows we assume that M is bounded.

Lemma 5.4.8. *Let $\mathcal{M}$ be a surface manifold. Then $T_M\mathcal{M}$ is an inner product space with inner product given by*

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_M = \int_M \mathbf{v}_1(\mathbf{x})\mathbf{v}_2(\mathbf{x}) \, dS \qquad \mathbf{v}_1, \mathbf{v}_2 \in T_M\mathcal{M},$$

where dS is the Lebesgue surface measure on M.

Since $\mathbf{v}_1, \mathbf{v}_2$ are continuous functions and M is a bounded set, it is clear that the integral exists. Thus for each surface manifold $\mathcal{M}$, the tangent space at each point $M \in \mathcal{M}$ is a subspace of $L^2(M)$.

The Lebesgue surface measure, dS, on a regular surface is related to the Lebesgue volume measure, $d\mathbf{x}$, by the next result.

Theorem 5.4.9 ([32], see Theorem 6.1.5). *Let $M \subset \mathbb{R}^n$ be a regular surface. Then*

$$\int_M dS = \int_{\mathbb{R}^n} |\nabla\Phi|\delta(\Phi(\mathbf{x})) \, d\mathbf{x}.$$

Here $\delta(x)$ is the Dirac delta distribution, and $\delta(\Phi(\mathbf{x}))$ will often be denoted simply as $\delta(\Phi)$.

Thus we have the following corollary to Lemma 5.4.8.

Corollary 5.4.10. *Let $\mathcal{M}$ be a surface manifold and $M \in \mathcal{M}$ be a subset of $\mathbb{R}^n$. Suppose*

that $\mathbf{v}_1, \mathbf{v}_2 \in T_M \mathcal{M}$ and that $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2$ are continuous extensions of $\mathbf{v}_1, \mathbf{v}_2$, respectively, to all of $\mathbb{R}^n$. Then

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_M = \int_{\mathbb{R}^n} \tilde{\mathbf{v}}_1(\mathbf{x}) \tilde{\mathbf{v}}_2(\mathbf{x}) |\nabla \Phi| \delta(\Phi) \, d\mathbf{x}. \tag{5.15}$$

This is a useful representation of the inner product on $T_M \mathcal{M}$ as will be seen in the next chapter.

Chapter 6

Gradient Descent for Surface Functionals

In this chapter we introduce the results from the calculus of variations that are required for the development of the video segmentation algorithm. These results are then combined with the results from the previous chapter to develop a gradient descent algorithm for computing local minimizers of surface functionals.

6.1 Surface Functionals

First we define a surface functional.

Definition 6.1.1. *Let $\mathcal{M}$ be a surface manifold, $M \in \mathcal{M}$ a bounded regular surface, and Φ a level set representation for M. Suppose that g is an integrable function of $\mathbf{x}$, Φ, and $\nabla \Phi$ defined on M and that f is an integrable function of $\mathbf{x}$, Φ, and $\nabla \Phi$ defined on $\mathbb{R}^n$. Then a*

48

surface functional *is a mapping* $E \colon \mathcal{M} \to \mathbb{R}$ *of one of the following forms:*

$$E(M) = \int_M g(\mathbf{x}, \Phi, \nabla\Phi) \ dS, \tag{6.1}$$

$$E(M) = \int_{\mathbb{R}^n} f(\mathbf{x}, \Phi, \nabla\Phi) \ d\mathbf{x}. \tag{6.2}$$

Such functionals will often be written $E(\Phi)$ for some fixed level set representation of the surface M.

Noting again (see Theorem 5.4.9) that the Lebesgue measure, dS, on a surface M with level set representation Φ, is given by $|\nabla\Phi|\delta(\Phi)d\mathbf{x}$, a surface functional of type (6.1) can be rewritten as a surface functional of type (6.2) by

$$E(\Phi) = \int_M g(\mathbf{x}, \Phi, \nabla\Phi) \ dS = \int_{\mathbb{R}^n} \tilde{g}(\mathbf{x}, \Phi, \nabla\Phi)|\nabla\Phi|\delta(\Phi)d\mathbf{x}, \tag{6.3}$$

where $\tilde{g}$ is an arbitrary continuous extension of g to $\mathbb{R}^n$. This demonstrates that a surface functional against the Lebesgue surface measure on M can be written as a surface functional against the Lebesgue area (volume) measure on $\mathbb{R}^n$.

6.2 Differentiability and the Gradient of Surface Functionals

We now define differentiability for a surface functional. The next two definitions are found in [63].

Definition 6.2.1. *Let $E \colon \mathcal{M} \to \mathbb{R}$ be a surface functional. Then we say that E is differentiable at M if, for every regular surface evolution $M(t)$ such that $M(0) = M$, the mapping $t \mapsto E(M(t))$ is differentiable at $t = 0$, i.e.*

$$\lim_{h \to 0} \frac{E\left(M(h)\right) - E(M(0))}{h}$$

exists and is finite.

We now define the differential of E.

Definition 6.2.2. *Let* $E \colon \mathcal{M} \to \mathbb{R}$ *be a differentiable surface functional, and let* $M \in \mathcal{M}$. *If* $M(t)$ *is a regular surface evolution such that* $M(0) = M$ *and* $\mathbf{v} = \dot{M}(t)$, *then the differential of* E *at* M *is the mapping* $dE(M) \colon T_M \mathcal{M} \to \mathbb{R}$ *given by*

$$dE(M)\mathbf{v} = \frac{d}{dt} E(M(t)) \bigg|_{t=0}.$$

This definition is the direct analogue of the definition of the differential of a function on a manifold as is given in (5.3), but it does not necessarily endow the differential with the properties we associate with derivatives.

Theorem 6.2.3. *Let* $E \colon \mathcal{M} \to \mathbb{R}$ *be a differentiable surface functional at* $M \in \mathcal{M}$. *Then* $dE(M)$ *is a homogeneous mapping of degree 1.*

Proof. Let $c \in \mathbb{R}$. Then $M(ct)$ is a regular surface evolution such that $M(ct)\big|_{t=0} = M$ and $\dot{M}(ct) = c\dot{M}(t)$. Thus $dE(M)$ is homogenous, since

$$dE(M)(c\mathbf{v}) = \frac{d}{dt} E(M(ct)) \bigg|_{t=0} = c \frac{d}{dt} E(M(t)) \bigg|_{t=0} = c\, dE(M)\mathbf{v}.$$

This is what was to be shown. □

Though the differential is homogeneous, it is not necessarily additive and therefore not necessarily linear. Next we define the gradient of a surface functional, and, when the gradient of a surface functional exists, it will be shown that the differential of the surface functional is a linear map.

Definition 6.2.4. *Let* $E \colon \mathcal{M} \to \mathbb{R}$ *be a surface functional that is differentiable at the point*

$M \in \mathcal{M}$, and suppose there exists $\mathbf{w} \in T_M \mathcal{M}$ such that

$$dE(M)\mathbf{v} = \langle \mathbf{w}, \mathbf{v} \rangle_M = \int_M \mathbf{w}(\mathbf{x})\mathbf{v}(\mathbf{x}) \, dS \tag{6.4}$$

for all $\mathbf{v} \in T_M \mathcal{M}$. *Then* $\mathbf{w}$ *is called the* gradient of E at M *and is denoted by* $\nabla E(M)$.

In the definition of gradient given in 5.2.8, the gradient vector was always an element of the tangent space of M at $\mathbf{x}$. The analogous result does not hold in this case, as a vector $\mathbf{w}$ that satisfies (6.4) may exist in $L^2(M)$ but not in $T_M \mathcal{M}$. We will refer to $\mathbf{w}$ as the gradient only when it exists and is an element of $T_M \mathcal{M}$.

It is clear that the differential is linear when the gradient exists, since it is given by an inner product.

Lemma 6.2.5. *Let* $\mathcal{M}$ *be a surface manifold, and* $M \in \mathcal{M}$. *Suppose that* $E \colon \mathcal{M} \to \mathbb{R}$ *is a surface functional such that* $\nabla E(M)$ *exists and is an element of* $T_M \mathcal{M}$. *Then* $\nabla E(M)$ *is unique.*

Proof. This follows the argument given in the proof of 5.2.9.

Suppose that $\mathbf{w}_1, \mathbf{w}_2 \in T_M \mathcal{M}$ are such that

$$\langle \mathbf{w}_1, \mathbf{v} \rangle_M - \langle \mathbf{w}_2, \mathbf{v} \rangle_M = 0$$

for all $\mathbf{v} \in T_M \mathcal{M}$. Then, since $\mathbf{w}_1 - \mathbf{w}_2 \in T_M \mathcal{M}$, we have

$$\langle \mathbf{w}_1 - \mathbf{w}_2, \mathbf{w}_1 - \mathbf{w}_2 \rangle_M = \|\mathbf{w}_1 - \mathbf{w}_2\|_{L^2(M)}^2 = 0.$$

Thus $\mathbf{w}_1 - \mathbf{w}_2 = 0$, which implies that $\mathbf{w}_1 = \mathbf{w}_2$. $\qquad\square$

The primary question that remains is whether or not the gradient of *any* surface functional

exists. It will be shown in Section 6.4 that the gradient does exist for the surface functionals in which we are interested and that the differential of such functionals is a linear map.

6.3 Gradient Flow for Minimizing a Surface Functional

Our primary task is to define a minimization procedure for surface functionals analogous to the steepest descent method that was presented in Chapter 5 for functions on manifolds. A *gradient flow* (following the terminology of Evans [23]) for a surface functional is an evolution equation whose steady-state solution is a local minimizer of the surface functional. Such a method is often referred to as a *gradient descent*, and the two terms will be used interchangeably here.

Definition 6.3.1. *Let $E \colon \mathcal{M} \to \mathbb{R}$ be a surface functional, and $M(t) \in \mathcal{M}$ a regular surface evolution on $(-\epsilon, \epsilon)$ such that $\nabla E(M(t))$ exists for all $t \in (-\epsilon, \epsilon)$. Then $M(t)$ is called a* gradient flow *for E if it is a solution of*

$$\dot{M}(t) = -\nabla E(M(t)) \tag{6.5}$$

for all $t \in (-\epsilon, \epsilon)$.

The next two theorems are analogues of Theorems 5.1.1 and 5.1.2 and are the primary results that will be used in the chapters that follow.

Theorem 6.3.2. *Let $M(t)$ be a gradient flow for a surface functional $E \colon \mathcal{M} \to \mathbb{R}$. Then $t \mapsto E(M(t))$ is a non-increasing function of time.*

Proof. Invoking Definition 6.2.2 and (6.5) yields

$$\frac{d}{dt}E(M(t)) = dE(M(t))\dot{M}(t) = \left\langle \nabla E(M(t)), \dot{M}(t) \right\rangle_{M(t)}$$

$$= -\left\langle \nabla E(M(t)), \nabla E(M(t)) \right\rangle_{M(t)} \leq 0.$$

Thus $t \mapsto E(M(t))$ is a non-increasing function of time.   □

Note that using this proof, analogous to what is often done with minimization algorithms for functions of real variables, one can define a descent direction in general for a functional.

Definition 6.3.3. *Let E be a surface functional and $\mathcal{M}$ a surface manifold. Then a* descent direction *for E at $M \in \mathcal{M}$ is any vector $\mathbf{v} \in T_M\mathcal{M}$ such that*

$$dE(M)\mathbf{v} = \langle \nabla E, \mathbf{v} \rangle_M \leq 0. \tag{6.6}$$

Note that this allows one to develop a large number of minimization algorithms for a differentiable surface functional by choosing different descent directions. Moreover, it is easy to determine whether or not a vector $\mathbf{v} \in T_M\mathcal{M}$ is a descent direction by checking whether or not it satisfies (6.6).

Theorem 6.3.4. *Let $M(t)$ be a gradient flow such that $M(t) \to M_0$ as $t \to \infty$. Then $\nabla E(M_0) = 0$.*

Proof. Let $\mathbf{x}_0 \in M(0)$ and $\boldsymbol{\alpha} \colon [0,\infty) \to \mathbb{R}^n$ be a differentiable curve such that $\boldsymbol{\alpha}(0) = \mathbf{x}_0$ and $\boldsymbol{\alpha}(t) \in M(t)$ for all t. Since $M(t) \to M_0$, we have that $\boldsymbol{\alpha}(t) \to \mathbf{x}' \in M_0$ as $t \to \infty$. By the proof of Theorem 5.1.2, $\dot{\boldsymbol{\alpha}}(t) = \frac{d}{dt}\boldsymbol{\alpha}(t) \to 0$ as $t \to \infty$. Then by (5.14),

$$\dot{M}(t) = \dot{\boldsymbol{\alpha}}(t) \cdot \hat{\mathbf{n}}(\boldsymbol{\alpha}(t)) \to 0 \quad \text{as } t \to \infty,$$

which implies, by (6.5), that $\nabla E(M(t)) \to \nabla E(M_0) = 0$.   □

Let $\Phi(\mathbf{x}, t)$ be a level set representation of the gradient flow $M(t)$. Due to the fact that $\dot{M}(t) = -\Phi_t(\mathbf{x}, t)/|\nabla\Phi(\mathbf{x}, t)|$ (see (5.13)), the gradient flow equation can be written

$$\frac{\partial\Phi}{\partial t} = \nabla E|\nabla\Phi|. \tag{6.7}$$

The important thing to note here, in contrast to gradient descents for functions on manifolds, is that, if $M(0) \in \mathcal{M}$ then an evolution given by (6.7) yields a regular surface evolution $M(t)$ such that $M(t) \in \mathcal{M}$ for all t, i.e. the gradient flow produces a continuous time evolution that remains on the surface manifold associated with $M(0)$ for all time. This is a direct result of the definitions of regular surface evolution and gradient flow.

6.4 Computing the Gradient of Surface Functionals

In this section we compute the gradient of three classes of surface functionals by computing the Gâteaux derivative of each of them. First, we define a variation of a level set representation of a regular surface evolution.

Definition 6.4.1. *Let* Φ *be a level set representation of a regular surface* $M \in \mathcal{M}$. *Then a* variation *of* $\Phi(\mathbf{x})$, $\Psi(\mathbf{x})$, *is another level set representation of* M, *and the* variational regular surface evolution *of* Φ *along* Ψ *is the regular surface evolution given by* $t \rightarrow \Phi + t\Psi$ *for* $t \in (-\epsilon, \epsilon)$.

The next definition is a general definition from the calculus of variations for a large class of functionals, but we include it here only for surface functionals.

Definition 6.4.2. *Let* E *be a differentiable surface functional,* M *be a surface with level set representation* Φ, *and* Ψ *be a variation of* Φ. *The* Gâteux derivative *of* E *along* Ψ *is given by*

$$\left.\frac{d}{dt}E(\Phi + t\Psi)\right|_{t=0}.$$

If this derivative exists for all variations Ψ, then we say that E is Gâteux differentiable at the point $M = \{\mathbf{x} \in \mathbb{R}^n \mid \Phi(\mathbf{x}, 0) = 0\}$ of the surface manifold $\mathcal{M}$.

Using this definition, the Gâteux derivative is just the first variation of a surface functional as defined using the calculus of variations (cf. [17, 26]).

The Gâteux derivative of a surface functional is useful to us, because, in many cases, it is the differential of the functional.

Another point that should be made and will be used in what follows is the next result.

Theorem 6.4.3. *Let $\Phi(\mathbf{x})$ be a level set representation of a regular surface M, and let $\Psi(\mathbf{x})$ be a variation of Φ. Then the normal velocity of the variational regular surface evolution of Φ along Ψ, $t \to \Phi(\mathbf{x}) + t\Psi(\mathbf{x})$, is*

$$\mathbf{v}_n = \frac{-\Psi}{|\nabla\Phi|} \tag{6.8}$$

at $t = 0$.

Proof. According to equation (5.13), the normal velocity for the regular surface evolution $t \mapsto \Phi + t\Psi$ is given by

$$\mathbf{v}_n = -\frac{\partial(\Phi(\mathbf{x}) + t\Psi(\mathbf{x}))/\partial t}{|\nabla(\Phi(\mathbf{x}) + t\Psi(\mathbf{x}))|}\bigg|_{t=0} = \frac{1}{|\nabla\Phi(\mathbf{x})|}\left(-\frac{\partial\Phi(\mathbf{x})}{\partial t} - \frac{\partial(t\Psi(\mathbf{x}))}{\partial t}\right)\bigg|_{t=0} = \frac{-\Psi(\mathbf{x})}{|\nabla\Phi(\mathbf{x})|}.$$

This completes the proof. $\qquad\square$

Let $H(x)$ denote the Heaviside function,

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x \le 0. \end{cases}$$

Since this function is not continuous, it is not differentiable in the classical sense. It is

differentiable in the sense of distributions, however. The distribution derivative of H (see Section 5.4 of [21]) is the function H' such that

$$\langle f, H'(\Phi(\mathbf{x})) \rangle = - \int_{\mathbb{R}^n} H(\Phi(\mathbf{x})) f'(\mathbf{x}) \, d\mathbf{x} = - \langle f'(\mathbf{x}), H(\Phi(\mathbf{x})) \rangle$$

for all $f \in C_0^\infty(\mathbb{R}^n)$. It can be shown that the Dirac delta distribution is the unique distribution that satisfies this relationship. Thus

$$- \langle f'(\mathbf{x}), H(\Phi(\mathbf{x})) \rangle = \langle f(\mathbf{x}), \delta(\Phi(\mathbf{x})) \rangle$$

for all $f \in C_0^\infty(\mathbb{R}^n)$. In this sense we can "differentiate" the Heaviside function inside an integral, i.e.

$$\frac{d}{d\Phi} \int_{\mathbb{R}^n} H(\Phi) \, d\mathbf{x} = \int_{\mathbb{R}^n} \delta(\Phi) \, d\mathbf{x}.$$

Theorem 6.4.4. *Let $f \in L^2(\mathbb{R}^n)$, M be a regular surface with level set representation Φ, and E be a surface functional given by*

$$E(M) = \int_\Omega f(\mathbf{x}) \, d\mathbf{x},$$

where Ω is the interior of M, i.e. $\Omega = \{\mathbf{x} \in \mathbb{R}^n \mid \Phi(\mathbf{x}) < 0\}$. Then $\nabla E = f(\mathbf{x})$.

Proof. To see this, we compute the Gâteux derivative of E.

$$E(M) = \int_\Omega f(\mathbf{x}) \, d\mathbf{x} = \int_{\mathbb{R}^n} f(\mathbf{x}) \, (1 - H(\Phi)) \, d\mathbf{x}, \tag{6.9}$$

so

$$
\begin{aligned}
\frac{d}{ds}E(\Phi + s\Psi)\Big|_{s=0} &= \frac{d}{ds}\int_{\mathbb{R}^n} f(\mathbf{x})\left(1 - H(\Phi + s\Psi)\right)\,d\mathbf{x}\Big|_{s=0} \\
&= \int_{\mathbb{R}^n} -\frac{d}{ds}H(\Phi(\mathbf{x}) + s\Psi(\mathbf{x}))\Big|_{s=0} f(\mathbf{x})\,d\mathbf{x} \\
&= \int_{\mathbb{R}^n} (-\Psi(\mathbf{x}))\,\delta(\Phi(\mathbf{x}))f(\mathbf{x})\,d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{-\Psi}{|\nabla\Phi|}f(\mathbf{x})\,|\nabla\Phi|\delta(\Phi)\,d\mathbf{x} = \int_M \mathbf{v}_n(\mathbf{x})f(\mathbf{x})\,dS = \langle \mathbf{v}_n, f\rangle_M,
\end{aligned}
$$

where $\mathbf{v}_n = -\Psi/|\nabla\Phi|$ is the normal velocity of the variational regular surface evolution of Φ along Ψ. Since any function $\mathbf{v}_n \in T_M\mathcal{M}$ can be written in the form $-\Psi/|\nabla\Phi|$ for some variation $\Psi \in L^2(M)$ of Φ, Definition 6.2.4 yields that $\nabla E = f(\mathbf{x})$ and, moreover, that the Gâteux derivative of E at M is the differential of E at M applied to the normal velocity $-\Psi/|\nabla\Phi|$. $\qquad\square$

Applying (6.7), a gradient flow for (6.9) is given by

$$
\frac{\partial\Phi}{\partial t} = f(\mathbf{x})|\nabla\Phi|. \tag{6.10}
$$

This is a well-known result that can be found, for example, in [36, 49]. The proofs are given there are in two dimensions and do not scale to higher dimensions. The proof given here is independent of dimension.

An immediate corollary of Theorem 6.4.4 is that if E is defined by

$$
E(M) = \int_{\mathbb{R}^n\backslash\Omega} f(\mathbf{x})\,d\mathbf{x} = \int_{\mathbb{R}^n} f(\mathbf{x})H(\Phi(\mathbf{x}))\,d\mathbf{x},
$$

then $\nabla E = -f(\mathbf{x})$, and a gradient flow for $E(M)$ is given by

$$
\frac{\partial\Phi}{\partial t} = -f(\mathbf{x})|\nabla\Phi|.
$$

Combining these two results yields that if E is a surface functional of the form

$$E(M) = \int_{\mathbb{R}^n} f(\mathbf{x}) \, d\mathbf{x} = \int_{\mathbb{R}^n} f_1(\mathbf{x}) \left(1 - H(\Phi(\mathbf{x}))\right) \, d\mathbf{x} + \int_{\mathbb{R}^n} f_2(\mathbf{x}) H(\Phi(\mathbf{x})) \, d\mathbf{x},$$

where $f_1, f_2 \in L^2(\mathbb{R}^n)$ are such that $f_1(\mathbf{x}) = f(\mathbf{x})$ on Ω and $f_2(\mathbf{x}) = f(\mathbf{x})$ on $\mathbb{R}^n \setminus \Omega$, then $\nabla E = f_1 - f_2$ and the gradient flow for E is given by

$$\frac{\partial \Phi}{\partial t} = (f_1(\mathbf{x}) - f_2(\mathbf{x})) \, |\nabla \Phi|. \tag{6.11}$$

Before computing the gradient of the next surface functional, we require some notation.

Definition 6.4.5. *Let* $\mathbf{g} \colon \mathbb{R}^n \to \mathbb{R}^n$ *be a differentiable function with components*

$$\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_n(\mathbf{x})).$$

Then the divergence *of* $\mathbf{g}$ *is given by*

$$\mathrm{div}(\mathbf{g}) = \nabla \cdot \mathbf{g} = \sum_{i=1}^{n} \frac{\partial g_i}{\partial x_i}.$$

If $g \colon M \to \mathbb{R}$, define $g_{\hat{\mathbf{n}}}(\mathbf{x}) \colon M \to \mathbb{R}^{m+1}$ by

$$g_{\hat{\mathbf{n}}}(\mathbf{x}) = \nabla g(\mathbf{x}) - \langle \hat{\mathbf{n}}(\mathbf{x}), \nabla g(\mathbf{x}) \rangle \, \hat{\mathbf{n}}(\mathbf{x}), \tag{6.12}$$

where $\hat{\mathbf{n}}(\mathbf{x})$ is unit normal to M at $\mathbf{x}$. Note that $\hat{\mathbf{n}}(\mathbf{x}) \colon M \to S^m$, where

$$S^m = \{\mathbf{x} \in \mathbb{R}^{m+1} \; : \; \|\mathbf{x}\| = 1\},$$

and, hence, following (5.5), $g_{\hat{\mathbf{n}}}(\mathbf{x}) = \nabla_{S^m} g$.

Theorem 6.4.6. *Let* $M \subset \mathbb{R}^{m+1}$ *be a regular surface,* $g \colon M \times S^m \to \mathbb{R}$ *be integrable on* M,

and let E be a surface functional of the form

$$E(M) = \int_M g(\mathbf{x}, \hat{\mathbf{n}}(\mathbf{x})) \ dS, \tag{6.13}$$

where $\hat{\mathbf{n}}$ is normal to the surface at $\mathbf{x}$. Suppose further that g is continuously differentiable with respect to $\mathbf{x}$ and twice continuously differentiable with respect to $\hat{\mathbf{n}}$, then

$$\nabla E = \nabla \cdot (g_{\hat{\mathbf{n}}} + g\hat{\mathbf{n}}) \tag{6.14}$$

where $g_{\hat{\mathbf{n}}}$ is given by (6.12).

This result is well-known, and and a proof in two dimensions is given in [28]. That proof does not scale, so we give a scalable proof here. The results of the four following lemmas are stated without proof in [63]. We include the statements and proofs.

We must compute the Gâteux derivative of (6.13). First note that

$$E(M) = \int_M g(\mathbf{x}, \hat{\mathbf{n}}) \ dS = \int_{\mathbb{R}^n} g\left(\mathbf{x}, \frac{\nabla\Phi}{|\nabla\Phi|}\right) |\nabla\Phi|\delta(\Phi) \ d\mathbf{x},$$

where Φ is a level set representation of the surface M. The following short sequence of results will be combined to prove the theorem.

Lemma 6.4.7. *Let Ψ be a variation of Φ. Then*

$$\frac{d}{ds} |\nabla(\Phi + s\Psi)| \bigg|_{s=0} = \frac{\nabla\Phi \cdot \nabla\Psi}{|\nabla\Phi|} = \frac{\nabla\Phi}{|\nabla\Phi|} \cdot \nabla\Psi. \tag{6.15}$$

Proof. Invoking the quotient rule for differentiation yields

$$\frac{d}{ds}\left|\nabla(\Phi + s\Psi)\right|\bigg|_{s=0} = \frac{d}{ds}\left|\nabla\Phi + s\nabla\Psi\right|\bigg|_{s=0}$$

$$= \frac{d}{ds}\frac{\nabla(\Phi + s\Psi)\cdot\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\bigg|_{s=0}$$

$$= \frac{d}{ds}\left(\frac{|\nabla\Phi|^2}{|\nabla(\Phi + s\Psi)|} + \frac{2s\nabla\Phi\cdot\nabla\Psi}{|\nabla(\Phi + s\Psi)|} + \frac{s^2|\nabla\Psi|^2}{|\nabla(\Phi + s\Psi)|}\right)\bigg|_{s=0}$$

$$= \frac{|\nabla(\Phi + s\Psi)|(0) - |\nabla\Phi|^2\frac{d}{ds}|\nabla(\Phi + s\Psi)|}{|\nabla(\Phi + s\Psi)|^2}\bigg|_{s=0} +$$

$$\frac{2|\nabla(\Phi + s\Psi)|\nabla\Phi\cdot\nabla\Psi - 2s\nabla\Phi\cdot\nabla\Psi\frac{d}{ds}|\nabla(\Phi + s\Psi)|}{|\nabla(\Phi + s\Psi)|^2}\bigg|_{s=0} +$$

$$\frac{2s|\nabla(\Phi + s\Psi)||\nabla\Psi|^2 - s^2|\nabla\Psi|^2\frac{d}{ds}|\nabla(\Phi + s\Psi)|}{|\nabla(\Phi + s\Psi)|^2}\bigg|_{s=0}$$

$$= -\frac{d}{ds}|\nabla(\Phi + s\Psi)|\bigg|_{s=0} + 2\frac{\nabla\Phi\cdot\nabla\Psi}{|\nabla\Phi|}.$$

Thus

$$\frac{d}{ds}\left|\nabla(\Phi + s\Psi)\right|\bigg|_{s=0} = \frac{\nabla\Phi\cdot\nabla\Psi}{|\nabla\Phi|} = \frac{\nabla\Phi}{|\nabla\Phi|}\cdot\nabla\Psi.$$

□

Lemma 6.4.8. *Again let* Ψ *be a variation of* Φ. *Then*

$$\frac{d}{ds}\left(\frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\right)\bigg|_{s=0} = \frac{\nabla\Psi}{|\nabla\Phi|} - \frac{\nabla\Phi}{|\nabla\Phi|}\left(\frac{\nabla\Phi}{|\nabla\Phi|}\cdot\frac{\nabla\Psi}{|\nabla\Phi|}\right) \qquad (6.16)$$

Proof. A direct computation, along with (6.15), yields

$$\frac{d}{ds}\left(\frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\right)\bigg|_{s=0} = \frac{|\nabla(\Phi + s\Psi)|\frac{d}{ds}\nabla(\Phi + s\Psi) - \nabla(\Phi + s\Psi)\frac{d}{ds}|\nabla(\Phi + s\Psi)|}{|\nabla(\Phi + s\Psi)|^2}\bigg|_{s=0}$$

$$= \frac{\nabla\Psi|\nabla\Phi|}{|\nabla\Phi|^2} - \frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|^2}\left(\frac{\nabla\Phi}{|\nabla\Phi|}\cdot\nabla\Psi\right)\bigg|_{s=0}$$

$$= \frac{\nabla\Psi}{|\nabla\Phi|} - \frac{\nabla\Phi}{|\nabla\Phi|}\left(\frac{\nabla\Phi}{|\nabla\Phi|}\cdot\frac{\nabla\Psi}{|\nabla\Phi|}\right).$$

$\square$

Lemma 6.4.9. *The Gâteux derivative of g is given by*

$$\frac{d}{ds}g\left(\mathbf{x}, \frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\right)\bigg|_{s=0} = g_{\hat{\mathbf{n}}}\cdot\frac{\nabla\Psi}{|\nabla\Phi|} \tag{6.17}$$

Proof. Let $\hat{\mathbf{n}}^s = \frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}$, then

$$\frac{d}{ds}g\left(\mathbf{x}, \frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\right)\bigg|_{s=0} = \frac{\partial g}{\partial\mathbf{x}}\cdot\frac{\partial\mathbf{x}}{\partial s} + g_{\hat{\mathbf{n}}^s}\cdot\frac{d\hat{\mathbf{n}}^s}{ds}\bigg|_{s=0} = g_{\hat{\mathbf{n}}^s}\cdot\frac{d\hat{\mathbf{n}}^s}{ds}\bigg|_{s=0}.$$

Now

$$g_{\hat{\mathbf{n}}^s}\cdot\hat{\mathbf{n}}^s = [\nabla g - \langle\nabla g, \hat{\mathbf{n}}^s\rangle\hat{\mathbf{n}}^s]\cdot\hat{\mathbf{n}}^s = \langle\nabla g, \hat{\mathbf{n}}^s\rangle(1 - \|\hat{\mathbf{n}}^s\|^2) = 0,$$

so invoking equation (6.16) yields

$$g_{\hat{\mathbf{n}}^s}\cdot\frac{d\hat{\mathbf{n}}^s}{ds}\bigg|_{s=0} = g_{\hat{\mathbf{n}}^s}\cdot\frac{d}{ds}\frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\bigg|_{s=0}$$

$$= g_{\hat{\mathbf{n}}}\cdot\left[\frac{\nabla\Psi}{|\nabla\Phi|} - \frac{\nabla\Phi}{|\nabla\Phi|}\left(\frac{\nabla\Phi}{|\nabla\Phi|}\cdot\frac{\nabla\Psi}{|\nabla\Phi|}\right)\right]$$

$$= g_{\hat{\mathbf{n}}}\cdot\frac{\nabla\Psi}{|\nabla\Phi|} - (g_{\hat{\mathbf{n}}}\cdot\hat{\mathbf{n}})\left(\frac{\nabla\Phi}{|\nabla\Phi|}\cdot\frac{\nabla\Psi}{|\nabla\Phi|}\right) = g_{\hat{\mathbf{n}}}\cdot\frac{\nabla\Psi}{|\nabla\Phi|}.$$

This is what was to be shown. $\square$

Lemma 6.4.10. *The Gâteux derivative of* $\delta(\Phi)$ *is*

$$\frac{d}{ds}\delta(\Phi + s\Psi)\bigg|_{s=0} = \delta'(\Phi)\Psi, \tag{6.18}$$

where $\delta'(\Phi)$ *is the distribution derivative of the Dirac delta at* Φ.

Proof. The classical chain rule for differentiation, which also holds for distribution derivatives, yields

$$\frac{d}{ds}\delta(\Phi + s\Psi)\bigg|_{s=0} = \delta'(\Phi + s\Psi)\Psi\big|_{s=0} = \delta'(\Phi)\Psi.$$

$\square$

These four results will be combined to prove the final result.

Proof of Theorem 6.4.6. The goal is to compute the gradient of the surface functional E by computing the Gâteux derivative. Recalling the change of variables in (6.3) and the functional in (6.13), we have

$$\frac{d}{ds}E(\Phi + s\Psi)\bigg|_{s=0} = \int_{\mathbb{R}^n} \frac{d}{ds}g\left(\mathbf{x}, \frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\right) |\nabla(\Phi + s\Psi)|\delta(\Psi + s\Psi)\bigg|_{s=0} d\mathbf{x}.$$

Computing the derivative of the integrand by invoking the above lemmas yields

$$\frac{d}{ds}\left[g\left(\mathbf{x}, \frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\right) |\nabla(\Phi + s\Psi)|\delta(\Psi + s\Psi)\right]\bigg|_{s=0}$$

$$= g\frac{d}{ds}\left(|\nabla(\Phi + s\Psi)|\delta(\Phi + s\Psi)\right)\bigg|_{s=0} + |\nabla(\Phi)|\delta(\Phi)\frac{d}{ds}\left[g\left(\mathbf{x}, \frac{\nabla(\Phi + s\Psi)}{|\nabla(\Phi + s\Psi)|}\right)\right]\bigg|_{s=0}$$

$$= g\left[|\nabla(\Phi)|\delta'(\Phi)\Psi + \delta(\Phi)\left(\frac{\nabla\Phi}{|\nabla\Phi|}\cdot\nabla\Psi\right)\right] + |\nabla(\Phi)|\delta(\Phi)\left(g_{\hat{n}}\cdot\frac{\nabla\Psi}{|\nabla\Phi|}\right)$$

$$= g|\nabla\Phi|\delta'(\Phi)\Psi + \delta(\Phi)\left(g_{\hat{n}} + g\hat{n}\right)\cdot\nabla\Psi.$$

Therefore

$$\frac{d}{ds}E(\Phi + s\Psi)\Big|_{s=0} = \int_{\mathbb{R}^n} g|\nabla\Phi|\delta'(\Phi)\Psi + \delta(\Phi)\left(g_{\hat{\mathbf{n}}} + g\hat{\mathbf{n}}\right) \cdot \nabla\Psi \; d\mathbf{x}$$

$$= \int_{\mathbb{R}^n} g|\nabla\Phi|\delta'(\Phi)\Psi \; d\mathbf{x} + \int_{\mathbb{R}^n} \left(g_{\hat{\mathbf{n}}} \cdot \nabla\Psi\right)\delta(\Phi) \; d\mathbf{x}$$

$$+ \int_{\mathbb{R}^n} g(\hat{\mathbf{n}} \cdot \nabla\Psi)\delta(\Phi) \; d\mathbf{x}.$$

Let $F = g_{\hat{\mathbf{n}}}\delta(\Phi)\Psi$, then applying the product rule for the divergence of a vector field gives

$$\text{div}F = g_{\hat{\mathbf{n}}}\delta(\Phi) \cdot \nabla\Psi + \Psi\left(\nabla \cdot g_{\hat{\mathbf{n}}}\delta(\Phi)\right).$$

By Gauss' Divergence Theorem (cf. [42], p. 506),

$$\int_{\mathbb{R}^n} \text{div}F \; d\mathbf{x} = \int_M (F \cdot \hat{\mathbf{n}}) \; dS = \int_M \delta(\Phi)\Psi(\mathbf{x})\left(g_{\hat{\mathbf{n}}} \cdot \frac{\nabla\Phi}{|\nabla\Phi|}\right) \; dS = 0,$$

since $\Psi(\mathbf{x}) = 0$ for all $\mathbf{x} \in M$. Thus

$$\int_{\mathbb{R}^n} \text{div}F \; d\mathbf{x} = \int_{\mathbb{R}^n} g_{\hat{\mathbf{n}}}\delta(\Phi) \cdot \nabla\Psi + \Psi\left(\nabla \cdot g_{\hat{\mathbf{n}}}\delta(\Phi)\right) \; d\mathbf{x} = 0,$$

and therefore

$$\int_{\mathbb{R}^n} \left(g_{\hat{\mathbf{n}}} \cdot \nabla\Psi\right)\delta(\Phi) \; d\mathbf{x} = \int_{\mathbb{R}^n} (-\Psi)\left(\nabla \cdot g_{\hat{\mathbf{n}}}\delta(\Phi)\right) \; d\mathbf{x}.$$

Applying similar arguments yields

$$\int_{\mathbb{R}^n} \left(g\delta(\Phi)\frac{\nabla\Phi}{|\nabla\Phi|}\right) \cdot \nabla\Psi \; d\mathbf{x} = \int_{\mathbb{R}^n} (-\Psi)\left[\nabla \cdot \left(g\delta(\Phi)\frac{\nabla\Phi}{|\nabla\Phi|}\right)\right] \; d\mathbf{x}.$$

Thus

$$\frac{d}{ds}E(\Phi + s\Psi)\Big|_{s=0} = -\int_{\mathbb{R}^n} (-\Psi)\, g|\nabla\Phi|\delta'(\Phi)\, d\mathbf{x} + \int_{\mathbb{R}^n} (-\Psi)\, [\nabla\cdot(g_{\hat{\mathbf{n}}}\delta(\Phi))]\, d\mathbf{x}$$

$$+ \int_{\mathbb{R}^n} (-\Psi)\left[\nabla\cdot\left(g\delta(\Phi)\frac{\nabla\Phi}{|\nabla\Phi|}\right)\right]\, d\mathbf{x}$$

$$= \int_{\mathbb{R}^n} (-\Psi)\left[\nabla\cdot(g_{\hat{\mathbf{n}}}\delta(\Phi)) + \nabla\cdot\left(g\delta(\Phi)\frac{\nabla\Phi}{|\nabla\Phi|}\right) - g|\nabla\Phi|\delta'(\Phi)\right]\, d\mathbf{x}$$

$$= \int_{\mathbb{R}^n} (-\Psi)\left[(g_{\hat{\mathbf{n}}}\cdot\nabla\Phi)\delta'(\Phi) + (\nabla\cdot g_{\hat{\mathbf{n}}})\delta(\Phi) + \left(\nabla\cdot g\frac{\nabla\Phi}{|\nabla\Phi|}\right)\delta(\Phi)\right.$$

$$\left. + g\frac{\nabla\Phi}{|\nabla\Phi|}\cdot\nabla\Phi\delta'(\Phi) - g|\nabla\Phi|\delta'(\Phi)\right]\, d\mathbf{x}$$

As was noted above, $g_{\hat{\mathbf{n}}}\cdot\hat{\mathbf{n}} = 0$, which implies that $g_{\hat{\mathbf{n}}}\cdot\nabla\Phi = 0$. Also $g\frac{\nabla\Phi}{|\nabla\Phi|}\cdot\nabla\Phi\delta'(\Phi) = g|\nabla\Phi|\delta'(\Phi)$, so the last two terms cancel. Therefore

$$\frac{d}{ds}E(\Phi + s\Psi)\Big|_{s=0} = \int_{\mathbb{R}^n} (-\Psi)\left[(\nabla\cdot g_{\hat{\mathbf{n}}})\delta(\Phi) + \left(\nabla\cdot g\frac{\nabla\Phi}{|\nabla\Phi|}\right)\delta(\Phi)\right]\, d\mathbf{x}$$

$$= \int_{\mathbb{R}^n}\left(\frac{-\Psi}{|\nabla\Phi|}\right)\nabla\cdot\left[g_{\hat{\mathbf{n}}} + g\frac{\nabla\Phi}{|\nabla\Phi|}\right]|\nabla\Phi|\delta(\Phi)\, d\mathbf{x}$$

$$= \int_M\left(\frac{-\Psi}{|\nabla\Phi|}\right)\nabla\cdot\left[g_{\hat{\mathbf{n}}} + g\frac{\nabla\Phi}{|\nabla\Phi|}\right]\, dS$$

$$= \left\langle\frac{-\Psi}{|\nabla\Phi|}, \nabla\cdot\left[g_{\hat{\mathbf{n}}} + g\frac{\nabla\Phi}{|\nabla\Phi|}\right]\right\rangle_M,$$

and hence the gradient of the surface functional E is given by $\nabla E = \nabla\cdot(g_{\hat{\mathbf{n}}} + g\hat{\mathbf{n}})$. $\qquad\square$

This implies, following (6.7), that a gradient flow for E as given in (6.13) is

$$\frac{\partial\Phi}{\partial t} = \nabla\cdot\left[g_{\hat{\mathbf{n}}} + g\frac{\nabla\Phi}{|\nabla\Phi|}\right]|\nabla\Phi|. \tag{6.19}$$

Corollary 6.4.11. *The gradient of the surface functional*

$$E(M) = \int_M g(\mathbf{x})\, dS, \tag{6.20}$$

for $g \in C^1(\mathbb{R}^n)$, is given by $\nabla E = \nabla\cdot(g\hat{\mathbf{n}}) = \nabla\cdot\left(g\frac{\nabla\Phi}{|\nabla\Phi|}\right)$.

This is an immediate consequence of the preceding theorem, since g is no longer a function of the normal $\hat{n}$. A gradient flow for (6.20) is therefore

$$\frac{\partial \Phi}{\partial t} = \nabla \cdot \left(g \frac{\nabla \Phi}{|\nabla \Phi|} \right) |\nabla \Phi|. \tag{6.21}$$

Chapter 7

Leaf Video Segmentation

Given the theoretical developments in Chapters 5 and 6, we now return to the problem at hand. In Chapter 3, the preprocessing steps for a leaf video were discussed. Now we can address the actual segmentation of a preprocessed leaf video, i.e. the extraction of the regions of active fluorescence from each frame.

We follow the general approach of Mumford and Shah in [46], which was reformulated within a level set framework by Chan and Vese in [15] and Kimmel in [36, 37].

7.1 The Background Model

The goal is to extract the regions in each frame of a leaf video that are bright due to fluorescence dynamics. This is not possible directly from the intensity values of the preprocessed video, $I(\mathbf{x})$. The vein regions in each frame are dark throughout the leaf video, so all pixels that do not correspond to veins would be treated as bright. These pixels are not necessarily bright due to dynamics and are therefore not to be extracted. Since the goal is to extract the

66

regions that are fluorescing, another approach is required.

It is possible to record several frames of the leaf before the dynamics are initiated. This results in images that reflect the light intensity of the leaf when there is no fluorescence. A background model, $B(x, y)$, is then generated by averaging the intensity of each pixel over these initial frames. In the experiments seen here, between 20 and 60 images were averaged to generate the background model. Since this background model reflects the leaf intensities when there is no fluorescence, it is appropriate to compare each frame in the entire video sequence to this model, which depends only on the x and y coordinates.

It is possible to use intensity-based region segmentation techniques for extracting the fluorescence regions by separating

$$D(\mathbf{x}) = |I(\mathbf{x}) - B(x, y)|, \qquad (\; \mathbf{x} = (x, y, z) \;) \qquad (7.1)$$

into regions of high intensity and regions of low intensity. Regions of a leaf video that correspond to high values of $D(\mathbf{x})$ are those whose intensities at a given time are different from the corresponding intensities in the background model. These are the regions of active fluorescence.

The function $D(\mathbf{x})$ treats regions that are bright due to dynamics equally to regions that are dark due to dynamics. This is not a problem in our experiments, because dynamics are always initiated by causing the stomata to close. Thus the stomatal dynamics cause only bright regions in each experimental video.

In order to enhance the segmentation, $D(\mathbf{x})$ is first linearly scaled to take on the full range of values in $[0, 1]$. It is then scaled non-linearly and replaced with $D^*(\mathbf{x})$ where

$$D^*(\mathbf{x}) = \frac{1}{\pi} \arctan \left(s_1 \left(D(\mathbf{x}) - s_2 \right) \right) + 0.5. \qquad (7.2)$$

Here s_1 is a scale factor that determines the extent to which the data is scaled and s_2 is a parameter that is a lower bound for the values to segmented. This scaling pushes values greater than or equal s_2 above 0.5 and values less than s_2 below 0.5. A similar scaling can be done with any 'S'-curve, such as a cumulative probability density function. After the nonlinear scaling, $D(\mathbf{x})$ is again scaled linearly to take on values in the full range $[0,1]$. Though the original video data was discrete, taking integer values in the interval $[0,255]$, the preprocessed data - and therefore also the background model - are continuous in the sense that they take non-integer values. Therefore the scaled $D(\mathbf{x})$ can take any value in $[0,1]$, not just the values $1/n$ for $n \in \{0,1,2,\ldots,255\}$.

Images of $D(\mathbf{x})$ for a particular experiment can be seen in Figure 7.1. The "bright" regions in this data correspond to the regions that are actively fluorescing at each point in time.

7.2 Segmentation Functional

The data in Figure 7.1 is not binary; it takes values in the full range $[0,1]$ on the bounded video domain Ω. There are, however, distinguishable bright regions. The goal is, therefore, to approximate $D(\mathbf{x})$ by a binary function, i.e. a piecewise-constant function that takes on exactly two values. This formulation is called a *shape reconstruction* problem. There are several techniques for shape reconstruction problems. In [2, 48], the authors use level set methods based on shape derivatives. It has also been shown that shape reconstruction problems can be solved within a variational level set framework (cf. [7,8,56]). An overview of both types of methods can be found in [68]. The approach based on regular surface evolutions used here is a hybrid of methods based on variational and shape derivative techniques.

Let $u \colon \mathbb{R}^3 \to \mathbb{R}$ be a piecewise-constant function that takes the two values c_1 and c_2, and let

$$\Omega_1 = \{\mathbf{x} \in \Omega \mid u(\mathbf{x}) = c_1\}, \quad \Omega_2 = \{\mathbf{x} \in \Omega \mid u(\mathbf{x}) = c_2\}.$$
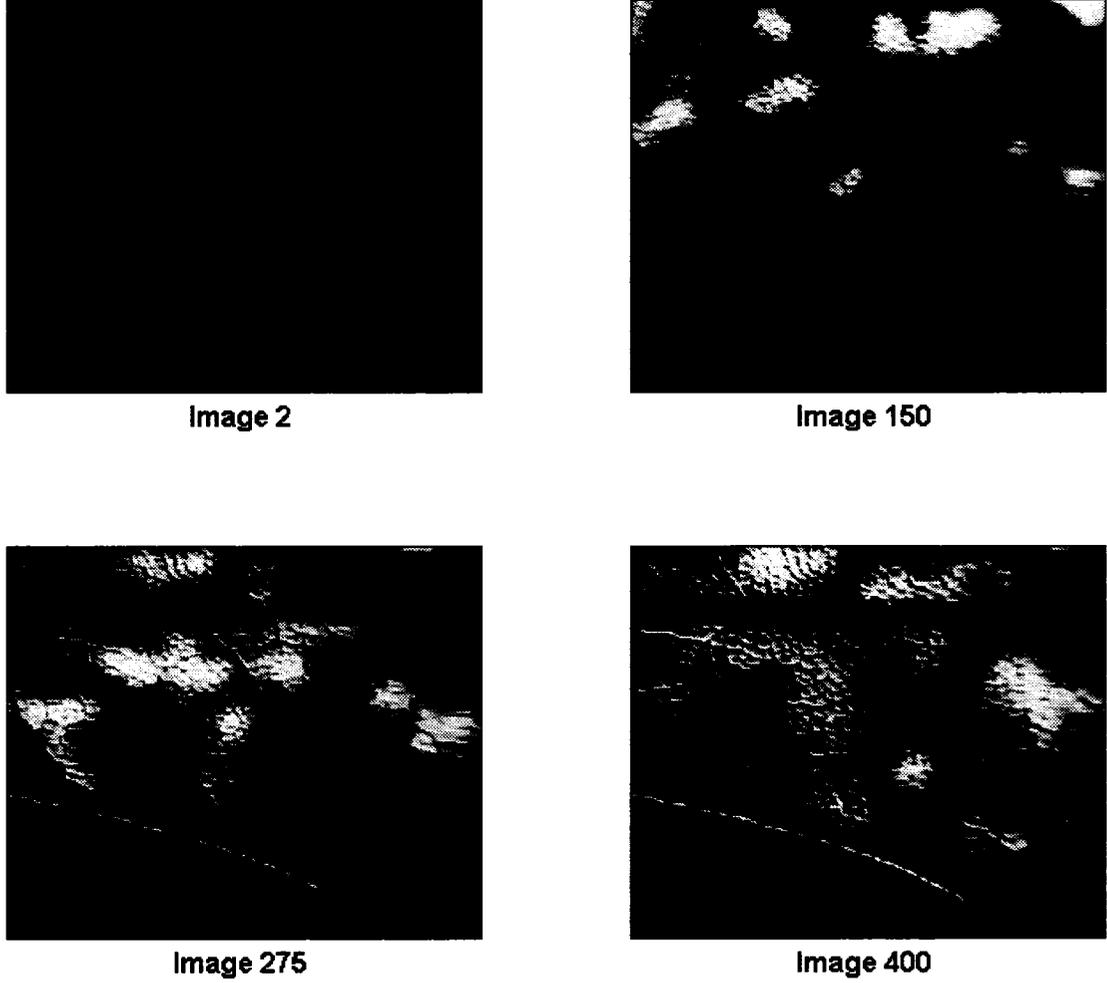
Image 2

Image 150

Image 275

Image 400

Figure 7.1: Non-linearly Scaled $D(\mathbf{x})$ with $s_1 = 15$ and $s_2 = 0.33$

We define the functional

$$E(u, c_1, c_2) = \|D(\mathbf{x}) - u(\mathbf{x})\|_{L^2(\Omega)}^2 = \int_\Omega (D(\mathbf{x}) - u(\mathbf{x}))^2 \, d\mathbf{x}$$

$$= \int_{\Omega_1} (D(\mathbf{x}) - u(\mathbf{x}))^2 \, d\mathbf{x} + \int_{\Omega_2} (D(\mathbf{x}) - u(\mathbf{x}))^2 \, d\mathbf{x}$$

$$= \int_{\Omega_1} (D(\mathbf{x}) - c_1)^2 \, d\mathbf{x} + \int_{\Omega_2} (D(\mathbf{x}) - c_2)^2 \, d\mathbf{x}. \tag{7.3}$$

Minimizing this functional with respect to u, c_1, and c_2 allows us to compute the sets Ω_1 and Ω_2 that best segment Ω into bright and dark regions with respect to D. This is called

a *minimum partition problem* and was first described by Mumford and Shah in [46]. Similar methods have been used by Chan, Vese, and Kimmel in [13–15,36,37].

First we compute the optimal values of c_1 and c_2.

Lemma 7.2.1. *Let E be as in* (7.3). *Then, for fixed u, the minimum of $E(u, c_1, c_2)$ occurs at*

$$c_1 = \frac{\int_{\Omega_1} D(\mathbf{x}) \, d\mathbf{x}}{\int_{\Omega_1} d\mathbf{x}}, \quad and \quad c_2 = \frac{\int_{\Omega_2} D(\mathbf{x}) \, d\mathbf{x}}{\int_{\Omega_2} d\mathbf{x}}. \tag{7.4}$$

Thus the optimal values for c_1 and c_2 are the mean values of $D(\mathbf{x})$ on Ω_1 and Ω_2, respectively.

Proof. Assume that u is fixed. In order to compute the minimum of E with respect to c_1, we compute the first derivative

$$\frac{d}{dc_1} E(c_1, c_2) = \frac{d}{dc_1} \int_{\Omega_1} (D(\mathbf{x}) - c_1)^2 \, d\mathbf{x} + \frac{d}{dc_1} \int_{\Omega_2} (D(\mathbf{x}) - c_2)^2 \, d\mathbf{x}$$

$$= \int_{\Omega} \frac{d}{dc_1} (D(\mathbf{x}) - c_1)^2 \, d\mathbf{x} = \int_{\Omega_1} 2(D(\mathbf{x}) - c_1) \, d\mathbf{x}.$$

Since $D - c_1$ and $D - c_2$ are uniformly bounded and Ω_1 and Ω_2 are bounded sets, interchanging the order of integration and differentiation is justified.

A local minimizer or maximizer occurs where the first derivative with respect to c_1 is 0, and

$$\int_{\Omega_1} 2(D(\mathbf{x}) - c_1) \, d\mathbf{x} = 0 \quad \Rightarrow \quad c_1 = \frac{\int_{\Omega_1} D(\mathbf{x}) \, d\mathbf{x}}{\int_{\Omega_1} d\mathbf{x}}.$$

This is the only extremum of E with respect to c_1 for fixed u. It must be a minimizer, since E is bounded from below with respect to c_1 but is unbounded from above.

A similar argument shows that the minimum of E with respect to c_2 occurs at

$$c_2 = \frac{\int_{\Omega_2} D(\mathbf{x}) \, d\mathbf{x}}{\int_{\Omega_2} d\mathbf{x}}.$$

The formulae in (7.4) assume that the measure of the regions Ω_1 and Ω_2 are finite and nonzero. Since they are contained in Ω, which is bounded, it is clear that the measure of these sets is finite. If the measure of one these sets is zero, then there is no constraint on the corresponding constant. □

Lemma 7.2.1 allows us to write E only in terms of u, since c_1 and c_2 have explicit representations in terms of u. Thus the segmentation functional is

$$E(u) = \int_{\Omega_1} (D(\mathbf{x}) - c_1)^2 \ d\mathbf{x} + \int_{\Omega_2} (D(\mathbf{x}) - c_2)^2 \ d\mathbf{x}. \tag{7.5}$$

Minimizing (7.5) is not a well-posed problem in that neither the existence nor the uniqueness of minimizers is guaranteed for $D \in L^2(\Omega)$.

7.3 Level Set Formulation of Segmentation Functional

Minimizing (7.5) with respect to u is useful, because the set Ω_1 corresponds to the regions of the leaf video that are fluorescing due to stomatal dynamics. Equation (7.5) cannot be minimized directly, as we are unaware of any minimization procedure for minimizing a functional over a set of piecewise-constant functions that take on exactly two values. The video domain Ω is a subset of $\mathbb{R}^3$. Thus, if the boundary of Ω_1 is a regular surface, we can recast (7.5) using a level set formulation.

Assume that $\partial\Omega_1$ is a regular surface in $\mathbb{R}^3$ such that $\partial\Omega_1$ has Lebesgue measure 0, and let Φ be a level set representation of $\partial\Omega_1$ such that

$$\Omega_1 = \{\mathbf{x} \in \Omega \mid u(\mathbf{x}) = c_1\} = \{\mathbf{x} \in \mathbb{R}^3 \mid \Phi(\mathbf{x}) < 0\},$$
$$\Omega_2 = \{\mathbf{x} \in \Omega \mid u(\mathbf{x}) = c_2\} = \{\mathbf{x} \in \mathbb{R}^3 \mid \Phi(\mathbf{x}) > 0\},$$

where set equality is assumed to be up to a set of Lebesgue measure 0. Then (7.5) can be rewritten

$$E(\Phi) = \int_{\Omega_1} (D(\mathbf{x}) - c_1)^2 \, d\mathbf{x} + \int_{\Omega_2} (D(\mathbf{x}) - c_2)^2 \, d\mathbf{x}, \qquad (7.6)$$

where

$$c_1 = \text{mean} \{ D(\mathbf{x}) \mid \mathbf{x} \in \Omega, \ \Phi(\mathbf{x}) < 0 \}$$

$$c_2 = \text{mean} \{ D(\mathbf{x}) \mid \mathbf{x} \in \Omega, \ \Phi(\mathbf{x}) > 0 \}.$$

The goal is to compute a local minimizer of $E(\Phi)$ with respect to Φ. Since E is a surface functional of the form in Theorem 6.4.4, a gradient flow for (7.6) is given by (6.11). Given an initial regular surface Ω_1^* on the surface manifold $\mathcal{M}$ with level set representation $\Phi_0(\mathbf{x})$, we seek the steady-state solution of the initial-value problem

$$\frac{\partial \Phi}{\partial t} = \left[(D(\mathbf{x}) - c_1)^2 - (D(\mathbf{x}) - c_2)^2 \right] |\nabla \Phi|$$

$$= \left[(c_2 - c_1) \left(D(\mathbf{x}) + \frac{1}{2} (c_1 + c_2) \right) \right] |\nabla \Phi|, \quad \Phi(\mathbf{x}, 0) = \Phi_0(\mathbf{x}). \qquad (7.7)$$

If this gradient flow converges, then the steady-state solution is not necessarily a local minimizer to (7.6) over all functions u that are piecewise-constant taking on exactly two values. The zero level set of the solution is, however, a local minimizer on $\mathcal{M}$.

7.4   Regularization of the Segmentation Functional

Since the existence of minimizers for (7.5) is not guaranteed, it is also not guaranteed that the gradient flow in (7.7) will converge. It is therefore necessary to regularize the segmentation functional so that the resulting gradient flow will converge.

7.4.1 Geodesic Active Contours

In [46] (see Theorem 5.1), it is proven that there exists a piecewise-constant function $f^* \in L^2(\Omega)$ and a closed set $\Omega_1 \subset \Omega$ that are minimizers of the functional

$$E(f) = \int_\Omega (D(\mathbf{x}) - f(\mathbf{x}))^2 \, d\mathbf{x} + \alpha \int_{\partial\Omega_1} dS, \tag{7.8}$$

under the assumption that D is continuous. The first term can be viewed as a fidelity term ensuring that the solution f approximates D in $L^2(\Omega)$; the second term provides regularization in that it ensures the existence of minimizers of E and also ensures a regularity of the boundary of Ω_1; and α is the regularization parameter. Existence of minimizers of (7.8) is also proven in [18].

More generally, it is desirable to use a regularization term that is data-dependent. We follow [36] and regularize (7.6) by adding a *geodesic active contour* (GAC) term [10]. This is given by

$$E_{\text{GAC}}(\Phi) = \alpha \int_{\partial\Omega_1} g(\mathbf{x}) \, dS. \tag{7.9}$$

Here $g \colon \Omega \to \mathbb{R}$ is depends on the measured data; we use the standard edge-indicator

$$g(x, y, z) = \frac{1}{1 + I_x(x, y, z)^2 + I_y(x, y, z)^2}, \tag{7.10}$$

where I is the preprocessed video sequence. An edge-indicator function should only yield edge information, and edges are two-dimensional features. Thus g is computed using only the partial derivatives of I with respect to x and y.

Though the function in (7.10) has been used for regularization of segmentation functionals similar to (7.6), this particular usage is different. In [36], the GAC term is used to regularize a segmentation functional that segments based directly on the measured data. Thus both the segmentation term of the functional and the regularization term act on the input image. Here

the regularization term acts on the preprocessed input video, but the segmentation term acts on background-subtracted data.

The functional in (7.9) is of the form in Theorem 6.4.6. Thus a gradient flow for this functional is given by

$$\frac{\partial \Phi}{\partial t} = \alpha \mathrm{div}\left(g(\mathbf{x})\frac{\nabla \Phi}{|\nabla \Phi|}\right)|\nabla \Phi|. \tag{7.11}$$

The corresponding regularized segmentation functional is therefore given by

$$E(\Phi) = \beta \int_{\Omega_1} (D(\mathbf{x}) - c_1)^2 \, d\mathbf{x} + \beta \int_{\Omega_2} (D(\mathbf{x}) - c_2)^2 \, d\mathbf{x} + \alpha \int_{\partial \Omega_1} g(\mathbf{x}) \, dS. \tag{7.12}$$

This is a variation of the segmentation functional found in [36], where the segmentation is performed directly on the measured data and does not involve a background model.

The parameter β in (7.12) is a weighting factor for the segmentation terms of the functional. It is added so that the weights of the segmentation and regularization can be controlled individually. Such a weighting factor is not necessary, because β can be normalized to 1. Having it explicitly represented in the functional, however, eases the numerical approximation developed in the next chapter.

The functional in (7.12) is a generalization of the Mumford-Shah functional in (7.8), where $g \equiv 1$ (see also [14,15]). Problems that involve minimizing functionals that are sums of volume energies and surface energies, such as (7.12), are called *free discontinuity problems* [19].

The existence of minimizers of this functional follows from the existence of minimizers of the Mumford-Shah functional. This is shown in [3,18], assuming that $g \in L^\infty(\Omega)$, which clearly holds for g as in (7.10).

Combining (7.7) and (7.11), the gradient flow for (7.12) is

$$\frac{\partial \Phi}{\partial t} = \left(\alpha \mathrm{div} \left(g \frac{\nabla \Phi}{|\nabla \Phi|} \right) + \beta(c_2 - c_1) \left(D - \frac{1}{2}(c_1 + c_2) \right) \right) |\nabla \Phi|, \quad \Phi(\mathbf{x}, 0) = \Phi_0(\mathbf{x}). \quad (7.13)$$

Extracting the zero level set of the steady-state solution of (7.13) yields the boundary of the set Ω_1 that we seek. The decomposition of Ω into the sets Ω_1 and Ω_2 is then the desired the segmentation of the video sequence.

Chapter 8

Numerical Methods

Here we address the numerical techniques required for computing an approximate solution to the equation

$$\frac{\partial \Phi}{\partial t} = \left(\alpha \text{div} \left(g \frac{\nabla \Phi}{|\nabla \Phi|} \right) + \beta (c_2 - c_1) \left(D - \frac{1}{2}(c_1 + c_2) \right) \right) |\nabla \Phi|, \quad \Phi(\mathbf{x}, 0) = \Phi_0(\mathbf{x}), \quad (8.1)$$

presented in the previous chapter. This will be referred to as the hyperbolic form of the segmentation PDE. The video domain, Ω, is discretized as a uniform cubic grid with points (x_i, y_j, z_k) for $i = 1, \ldots, n_i$, $j = 1, \ldots, n_j$, and $k = 1, \ldots, n_k$. We denote by $\Phi_{i,j,k}^n$ the approximation to Φ at the nth numerical iteration at the spatial point (x_i, y_j, z_k).

8.1 An Explicit Method

The most direct method for computing a numerical approximation to the solution of (8.1) is to discretize the equation explicitly in time and to approximate the spatial derivatives using a *conservative numerical flux* function for hyperbolic conservation laws and Hamilton-Jacobi

76

equations. Conservative schemes are generalizations of upwind schemes for approximating spatial derivatives that numerically account for the formation of shock waves along the characteristic curves of the PDE.

8.1.1 Explicit Time Discretization

The explicit discretization of the time derivative is given by

$$\frac{\partial \Phi}{\partial t} \approx \frac{\Phi_{i,j,k}^{n+1} - \Phi_{i,j,k}^{n}}{\Delta t}.$$

Next we approximate $\beta(c_2 - c_1)\left(D - \frac{1}{2}(c_1 + c_2)\right)|\nabla \Phi|$ and define

$$F_{i,j,k}^{n} = \beta(c_2^{n} - c_1^{n})(D_{i,j,k} - \frac{1}{2}(c_1^{n} + c_2^{n})), \tag{8.2}$$

recalling that c_1 and c_2 depend on Φ. Following the notation in [59], the Engquist-Osher [22] conservative numerical approximation to $F_{i,j,k}^{n}|\nabla \Phi^{n}|$ is given by

$$F_{i,j,k}^{n}|\nabla \Phi^{n}| \approx \max(F_{i,j,k}^{n}, 0)\nabla^{+} + \min(F_{i,j,k}^{n}, 0)\nabla^{-},$$
$$F_{i,j,k}^{n}|\nabla \Phi^{n}| \approx \max(F_{i,j,k}^{n}, 0)\nabla^{+} + \min(F_{i,j,k}^{n}, 0)\nabla^{-},$$

where

$$\nabla^{+} = [\,\max(D_x^{-}\Phi_{i,j,k}^{n}, 0)^2 + \min(D_x^{+}\Phi_{i,j,k}^{n}, 0)^2$$

$$\max(D_y^{-}\Phi_{i,j,k}^{n}, 0)^2 + \min(D_y^{+}\Phi_{i,j,k}^{n}, 0)^2$$

$$\max(D_z^{-}\Phi_{i,j,k}^{n}, 0)^2 + \min(D_z^{+}\Phi_{i,j,k}^{n}, 0)^2]^{1/2},$$

and

$$\nabla^- = [\min(D_x^- \Phi_{i,j,k}^n, 0)^2 + \max(D_x^+ \Phi_{i,j,k}^n, 0)^2$$
$$\min(D_y^- \Phi_{i,j,k}^n, 0)^2 + \max(D_y^+ \Phi_{i,j,k}^n, 0)^2$$
$$\min(D_z^- \Phi_{i,j,k}^n, 0)^2 + \max(D_z^+ \Phi_{i,j,k}^n, 0)^2]^{1/2}.$$

Recall that D_x^+ and D_x^- are the forward and backward finite difference approximations to the partial derivative in the x direction (see (3.9)), likewise $D_y^+, D_y^-, D_z^+, D_z^-$ are the forward and backward finite difference approximations to the y and z partial derivatives. The theory of conservative numerical schemes for hyperbolic conservation laws, along with other conservative numerical schemes, can be found in [53].

Finally we approximate the regularization term of (8.1),

$$\mathrm{div}\left(g\frac{\nabla\Phi}{|\nabla\Phi|}\right)|\nabla\Phi| = \left(g\mathrm{div}\left(\frac{\nabla\Phi}{|\nabla\Phi|}\right) + \nabla g \cdot \frac{\nabla\Phi}{|\nabla\Phi|}\right)|\nabla\Phi|$$
$$= g|\nabla\Phi|\kappa + \nabla g \cdot \nabla\Phi. \tag{8.3}$$

There are two choices for the curvature, κ, in three dimensions: the mean curvature and the Gaussian curvature. We choose the mean curvature here. Curvature is a geometric quantity that can be approximated using centered differences. The centered difference approximations to the first and second derivatives in the x direction are given by

$$D_x^0 \Phi_{i,j,k}^n \approx \frac{\Phi_{i,j+1,k}^n - \Phi_{i,j-1,k}^n}{2\Delta x} \tag{8.4}$$
$$D_{xx} \Phi_{i,j,k}^n \approx \frac{1}{\Delta x^2}\left(\Phi_{i,j+1,k}^n - 2\Phi_{i,j,k}^n + \Phi_{i,j-1,k}^n\right).$$

Similar formulas hold for the y and z derivatives.

Mean curvature is given by

$$\kappa = \frac{\kappa_1 + \kappa_2}{\left(\Phi_x^2 + \Phi_y^2 + \Phi_z^2\right)^{3/2}},$$

where

$$\kappa_1 = (\Phi_{yy} + \Phi_{zz})\Phi_x^2 + (\Phi_{xx} + \Phi_{zz})\Phi_y^2 + (\Phi_{yy} + \Phi_{yy})\Phi_z^2,$$

$$\kappa_2 = -2\left(\Phi_x\Phi_y\Phi_{xy} + \Phi_x\Phi_z\Phi_{xz} + \Phi_y\Phi_z\Phi_{yz}\right).$$

The norm of the gradient, $|\nabla\Phi|$, in (8.3) can also be approximated using centered differences.

The second term in (8.3), $\nabla g \cdot \nabla\Phi$, is not a geometric term, and $\nabla\Phi$ cannot be approximated using centered differences. Since g is a time-independent function, its gradient can be approximated with centered differences, and we define $\nabla_0 g$, the centered difference approximation to the spatial gradient of g, by

$$\nabla_0 g_{i,j,k} = \left(D_x^0 g_{i,j,k}, \ D_y^0 g_{i,j,k}, \ D_z^0 g_{i,j,k}\right). \tag{8.5}$$

The inner product of ∇g and $\nabla\Phi$ is a passive advection field, which does not require a conservative scheme, so we approximate $\nabla g \cdot \nabla\Phi$ using the upwind scheme

$$\nabla_0 g_{i,j,k} \cdot \nabla\Phi_{i,j,k}^n \approx \max(D_x^0 g_{i,j,k}, 0)D_x^-\Phi_{i,j,k}^n + \min(D_x^0 g_{i,j,k}, 0)D_x^+\Phi_{i,j,k}^n$$

$$+ \max(D_y^0 g_{i,j,k}, 0)D_y^-\Phi_{i,j,k}^n + \min(D_y^0 g_{i,j,k}, 0)D_y^+\Phi_{i,j,k}^n$$

$$+ \max(D_z^0 g_{i,j,k}, 0)D_z^-\Phi_{i,j,k}^n + \min(D_z^0 g_{i,j,k}, 0)D_z^+\Phi_{i,j,k}^n.$$

Thus we have the complete discretization of (8.1) using an explicit time discretization.

8.1.2   The CFL Condition for the Explicit Scheme

Solving (8.1) using the numerical approach outlined in this section yields a complicated numerical scheme, since three different approximations to the gradient are required for the three different terms. This makes the scheme computationally intensive, but not prohibitively so. In order to ensure numerical stability of the scheme, however, the time step, Δt, and the spatial discretization steps, $\Delta x = \Delta y = \Delta z = 1$, must satisfy a *Courant-Friedrichs-Lewy (CFL) condition* [52].

Let $h = \Delta x = \Delta y = \Delta z$, then it can be shown (cf. [59]) that the CFL condition for (8.1) is given by

$$\max_{\Omega} \left[\alpha \mathrm{div} \left(g \frac{\nabla \Phi}{|\nabla \Phi|} \right) + \beta(c_2 - c_1) \left(D - \frac{1}{2}(c_1 + c_2) \right) \right] \Delta t \leq h. \tag{8.6}$$

Since this maximum can be quite large and $h = 1$ is fixed for this problem, Δt is required to be small. Moreover, the left-hand side of (8.6) is time-dependent, so Δt must be updated at each numerical iteration to ensure that (8.6) still holds. Table 8.1 shows the required time steps

Iteration	1	2	3	4	5	6	7	8	9	10
Required Δt $(\times 10^{-6})$	7.91	10.38	7.65	5.55	4.0	3.07	2.47	2.05	1.74	1.51
Iteration time (seconds)	131	131	132	131	131	132	131	132	132	131

Table 8.1: Required Time Step for First 10 Iterations of Explicit Scheme

for the first ten iterations of the explicit segmentation method for the data seen in Figure 7.1 along with the actual time required to compute each iteration. Since the correct segmentation occurs for $t \approx 2$, the explicit scheme requires on the order of 10^6 iterations which, at 131 seconds per iteration, would take almost 4 years to complete.

Because the gradient requires multiple approximations at each numerical iteration and Δt must be very small and updated at each numerical iteration, solving (8.1) numerically is an untractable problem using an explicit time discretization.

8.2 Eliminating the Norm of the Gradient

Since solving (8.1) numerically is not feasible using an explicit time discretization, it is necessary to develop methods that are computationally simpler and that will not require restrictive time steps for numerical stability.

The first approximation that is often made is that $|\nabla \Phi| = 1$. Note that this is not a problem for the initial level set function, since $\Phi(\mathbf{x}, 0)$ can be chosen to be the signed distance function from the initial surface (cf. Theorem 5.3.3). At each further numerical iteration, however, this constraint will not hold and must be enforced.

Under the constraint that $|\nabla \Phi| = 1$, the resulting form of (8.1) is

$$\frac{\partial \Phi}{\partial t} = \alpha \operatorname{div}(g \nabla \Phi) + \beta(c_2 - c_1)\left(D - \frac{1}{2}(c_1 + c_2)\right), \quad \Phi(\mathbf{x}, 0) = \Phi_0(\mathbf{x}). \tag{8.7}$$

8.2.1 Reinitialization

In order to ensure that $|\nabla \Phi| = 1$ at each numerical iteration, it is necessary to reinitialize Φ to be the signed distance function to its zero level set. This is done by replacing Φ^n with ψ where ψ is the steady-state solution of

$$\psi_t = \operatorname{sign}(\Phi^n)\left(1 - |\nabla \psi|\right), \quad \psi(\mathbf{x}, 0) = \Phi^n. \tag{8.8}$$

Using an explicit discretization in time and a variation of the Engquist-Osher conservative approximation to the spatial gradient yields

$$\psi_{i,j,k}^{\eta+1} = \psi_{i,j,k}^{\eta} - \Delta\tau \operatorname{sign}(\Phi_{i,j,k}^n) G(\psi_{i,j,k}^{\eta})$$

where $\Delta\tau$ is the time step and η is the iteration count (cf. [15]). Let

$$
\begin{aligned}
A &= \max(D_x^- \psi_{i,j,k}^\eta, 0)^2, & A' &= \min(D_x^- \psi_{i,j,k}^\eta, 0)^2, \\
B &= \min(D_x^+ \psi_{i,j,k}^\eta, 0)^2, & B' &= \max(D_x^+ \psi_{i,j,k}^\eta, 0)^2, \\
C &= \max(D_y^- \psi_{i,j,k}^\eta, 0)^2, & C' &= \min(D_y^- \psi_{i,j,k}^\eta, 0)^2, \\
D &= \min(D_y^+ \psi_{i,j,k}^\eta, 0)^2, & D' &= \max(D_y^+ \psi_{i,j,k}^\eta, 0)^2, \\
E &= \max(D_z^- \psi_{i,j,k}^\eta, 0)^2, & E' &= \min(D_z^- \psi_{i,j,k}^\eta, 0)^2, \\
F &= \min(D_z^+ \psi_{i,j,k}^\eta, 0)^2, & F' &= \max(D_z^+ \psi_{i,j,k}^\eta, 0)^2,
\end{aligned}
$$

then

$$
G(\psi_{i,j,k}^\eta) = \begin{cases}
(\max(A,B) + \max(C,D) + \max(E,F))^{1/2} - 1 & \Phi_{i,j,k}^n > 0 \\[2mm]
(\max(A',B') + \max(C',D') + \max(E',F'))^{1/2} - 1 & \Phi_{i,j,k}^n < 0 \\[2mm]
0, & \text{otherwise.}
\end{cases}
$$

In practice it is not necessary to perform a large number of iterations, since it is only necessary to maintain accuracy near the zero level set of Φ^n. A stopping procedure (cf. [76]) for the reinitialization is given by computing

$$
Q = \frac{\sum_{|\Phi_{i,j,k}^n| < 1} |\psi_{i,j,k}^\eta - \Phi_{i,j,k}^n|}{M},
$$

where M is the number grid points such that $|\Phi_{i,j,k}^n| < 1$. If $Q \leq \Delta t$, then the solution is considered stationary and the reinitialization procedure is stopped.

Thus the algorithm for solving (8.1) is given by iterating the following two steps:

1. given Φ^n, compute Φ^{n+1} using (8.7);

2. reinitialize Φ^{n+1} to be the signed distance function from its zero level set by approximately solving (8.8) using the above procedure.

In all examples shown here, $\Phi_0(\mathbf{x})$ is chosen to be the signed distance function to a sphere contained entirely inside the video domain, Ω, and the orientation is chosen so that $\Phi_0(\mathbf{x})$ is negative on the interior of the sphere.

8.3 Discretizing the Divergence Operator

There are no derivatives in the term $\beta(c_2 - c_1)\left(D - \frac{1}{2}(c_1 + c_2)\right)$, so it is only necessary to develop a spatial discretization for

$$\text{div}\,(g(\mathbf{x})\nabla\Phi(\mathbf{x}, t)) \tag{8.9}$$

at the grid point (x_i, y_j, z_k). We follow the argument given in Morton and Mayers ([45], Section 6.3). Suppose that we want to solve the equation $\nabla \cdot (g\nabla\Phi) = 0$. Rather than discretizing this equation directly, it is possible to discretize an integral form.

Suppose that V is the cube given by the solid lines in Figure 8.1. This is the cube whose faces are parallel to the coordinate axes and whose volume is $1/8$ the volume of the cube centered at (x_i, y_j, z_k) whose vertices are the neighboring grid points (the cube given by the dashed lines in Figure 8.1). Each solid black dot is a point on the discretization grid for this problem. The points marked by an "×" are the half-grid points along the coordinate directions, each of which is the center point of a face of V.

Given this picture, solving $\nabla \cdot (g\nabla\Phi) = 0$ at (x_i, y_j, z_k) is equivalent to solving the integral form $\int_V \nabla \cdot (g\nabla\Phi)\,d\mathbf{x} = 0$ at (x_i, y_j, z_k). By Gauss' divergence theorem (cf. [42], Section 8.4)

$$\int_V \nabla \cdot (g\nabla\Phi)\,d\mathbf{x} = \int_{\partial V} g(\mathbf{x})\left(\frac{\partial\Phi}{\partial\hat{\mathbf{n}}}\right)\,dS, \tag{8.10}$$

where $\partial\Phi/\partial\hat{\mathbf{n}}$ is the normal derivative of Φ. This is a particularly useful expression given the volume element V, since the normal directions to the boundary of V are the coordinate
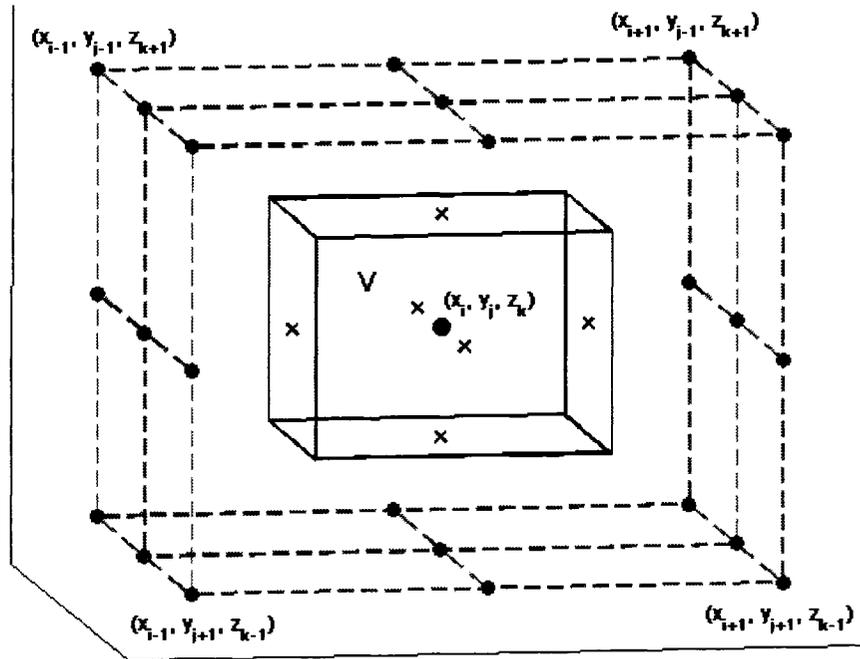
Figure 8.1: Volume Element for Discretizing the Divergence Operator

vectors. This means that the normal derivatives of Φ along the faces of V are just the regular partial derivatives, i.e. the normal derivative on the top face of V is Φ_z, the normal derivative on a vertical face parallel to the yz plane is Φ_x and the normal derivative on a vertical face parallel to the xz plane is Φ_y.

Recalling that $\Delta x = \Delta y = \Delta z = 1$ and noting that the area of each face of V is 1, we approximate

$$\int_{\partial V} g(\mathbf{x}) \left(\frac{\partial \Phi}{\partial \hat{\mathbf{n}}} \right) \, dS = 0 \tag{8.11}$$

by

$$g_{i+1/2,j,k} \left(\Phi_{i+1,j,k} - \Phi_{i,j,k} \right) - g_{i-1/2,j,k} \left(\Phi_{i,j,k} - \Phi_{i-1,j,k} \right)$$

$$+ g_{i,j+1/2,k} \left(\Phi_{i,j+1,k} - \Phi_{i,j,k} \right) - g_{i,j-1/2,k} \left(\Phi_{i,j,k} - \Phi_{i,j-1,k} \right)$$

$$+ g_{i,j,k+1/2} \left(\Phi_{i,j,k+1} - \Phi_{i,j,k} \right) - g_{i,j,k-1/2} \left(\Phi_{i,j,k} - \Phi_{i,j,k-1} \right) = 0 \qquad (8.12)$$

The finite difference approximation $\Phi_{i,j+1,k} - \Phi_{i,j,k}$ is the centered difference approximation to the x partial derivative at the point $\Phi_{i,j+1/2,k}$. If we approximate the value of $g_{i,j+1/2,k}$ by linearly interpolating in the x direction, then

$$g_{i,j+1/2,k} \approx \frac{g_{i,j+1,k} + g_{i,j,k}}{2}.$$

The approximation given by (8.12) can then be rewritten

$$\frac{1}{2} \sum_{l \in \{i+1,i-1\}} \left(g_{l,j,k} + g_{i,j,k} \right) \left(\Phi_{l,j,k} - \Phi_{i,j,k} \right) + \frac{1}{2} \sum_{l \in \{j+1,j-1\}} \left(g_{i,l,k} + g_{i,j,k} \right) \left(\Phi_{i,l,k} - \Phi_{i,j,k} \right)$$

$$\frac{1}{2} \sum_{l \in \{k+1,k-1\}} \left(g_{i,j,l} + g_{i,j,k} \right) \left(\Phi_{i,j,l} - \Phi_{i,j,k} \right) = 0. \qquad (8.13)$$

The discretization technique on the left-hand side of (8.13) is called a *finite volume method*, and we will use it to approximate the divergence term in of (8.7).

8.4 A Semi-implicit Scheme with Operator Splitting

It was noted above that there are two primary drawbacks to using the hyperbolic form of the segmentation PDE (8.1). The first is that the gradient must be approximated in three different ways. We have overcome this difficulty by enforcing $|\nabla \Phi| = 1$ at each numerical iteration. The second difficulty is that the CFL condition for the equation requires an extremely small

time step. This is prohibitive, since, in our case, a large number of iterations are required.

8.4.1 Semi-Implicit Scheme

Rather than discretizing (8.7) explicitly in time, it is necessary to develop a more stable technique. One such scheme is the *semi-implicit* time discretization of (8.7). This scheme is given by

$$\Phi_{i,j,k}^{n+1} - \alpha \Delta t \mathrm{div} \left(g_{i,j,k} \nabla \Phi_{i,j,k}^{n+1} \right) = \Phi_{i,j,k}^n + \beta \Delta t (c_2^n - c_1^n) \left(D_{i,j,k} - \frac{1}{2}(c_1^n + c_2^n) \right). \qquad (8.14)$$

Using the theory of scale-spaces, it is shown in [70, 71] that this scheme is numerically stable for all $\Delta t > 0$.

Let

$$A_s = \frac{\partial}{\partial s} g \frac{\partial}{\partial s}, \qquad s \in \{x, y, z\}. \qquad (8.15)$$

Expanding the divergence on the left-hand side of (8.14) yields

$$\Phi_{i,j,k}^{n+1} - \alpha \Delta t \left(A_x + A_y + A_z \right) \Phi_{i,j,k}^{n+1} = \Phi_{i,j,k}^n + \beta \Delta t (c_2^n - c_1^n) \left(D_{i,j,k} - \frac{1}{2}(c_1^n + c_2^n) \right). \qquad (8.16)$$

Thus the time update for Φ^{n+1} is given by

$$\Phi^{n+1} = (\mathcal{I} - \alpha \Delta t \left(A_x + A_y + A_z \right))^{-1} \left(\Phi^n + \beta \Delta t (c_2^n - c_1^n) \left(D - \frac{1}{2}(c_1^n + c_2^n) \right) \right), \qquad (8.17)$$

where $\mathcal{I}$ is the identity operator.

8.4.2 Additive Operator Splitting

The inverse operator required in (8.17) can be computed directly using a multigrid approach, but it is also possible to use operator splitting. Two common forms of operator splitting are the multiplicative splitting methods ADI (cf. [45]) and the *locally one-dimensional* (LOD) scheme (cf. [36]). The LOD method uses the approximation

$$(\mathcal{I} - \alpha\Delta t\,(A_x + A_y + A_z))^{-1} \approx (\mathcal{I} - \alpha\Delta t A_x)^{-1}\,(\mathcal{I} - \alpha\Delta t A_y)^{-1}\,(\mathcal{I} - \alpha\Delta t A_z)^{-1}. \qquad (8.18)$$

The operators on the right-hand side do not commute, so the order in which they are applied affects the solution. This introduces an axial bias, which is undesirable. Other multiplicative splitting schemes suffer from the same problem.

We use, instead, *additive operator splitting* (AOS) (cf. [39, 40, 70, 71]), where the following approximation is used:

$$(\mathcal{I} - \alpha\Delta t\,(A_x + A_y + A_z))^{-1} \approx \frac{1}{3} \sum_{s\in\{x,y,z\}} (\mathcal{I} - 3\alpha\Delta t A_s)^{-1}. \qquad (8.19)$$

This operator splitting yields an $\mathcal{O}(\Delta t^2)$ approximation to the semi-implicit scheme as can be seen from the Taylor series expansions

$$(\mathcal{I} - \alpha\Delta t(A_x + A_y + A_z))^{-1}\,v = v + \alpha\Delta t(A_x + A_y + A_z)v + \mathcal{O}(\Delta t^2),$$

and

$$\frac{1}{3}(\mathcal{I} - 3\alpha\Delta t A_x)^{-1}v + \frac{1}{3}(\mathcal{I} - 3\alpha\Delta t A_y)^{-1}v + \frac{1}{3}(\mathcal{I} - 3\alpha\Delta t A_z)^{-1}v$$

$$= \frac{1}{3}(v + 3\alpha\Delta t A_x v) + \frac{1}{3}(v + 3\alpha\Delta t A_y v) + \frac{1}{3}(v + 3\alpha\Delta t A_z v) + \mathcal{O}(\Delta t^2)$$

$$= v + \alpha\Delta t(A_x + A_y + A_z)v + \mathcal{O}(\Delta t^2).$$

Thus, since (8.17) is already an $\mathcal{O}(\Delta t^2)$ scheme, the global error does not change when AOS is used.

To compute the approximation to A_x, we use a one-dimensional version of the finite-volume discretization in (8.13). Let m be the number of elements in row i of frame k from the leaf video sequence, and $P_{i,k} \in \mathbb{R}^m$ be the vector whose l^{th} element is $\Phi_{i,l,k}$. Then $A_x \colon \mathbb{R}^m \to \mathbb{R}^m$ is given by

$$(A_x P_{i,k})_l \approx \frac{1}{2}\left(g_{i,l+1,k} + g_{i,l,k}\right)\left(\Phi_{i,l+1,k} - \Phi_{i,l,k}\right) + \frac{1}{2}\left(g_{i,l-1,k} + g_{i,l,k}\right)\left(\Phi_{i,l-1,k} - \Phi_{i,l,k}\right),$$

for $l \in \{2, \ldots, m - 1\}$. For row i in frame k, the matrix corresponding to A_x is then the tridiagonal

$$(A_x)_{j,l} = \begin{cases} \left(g_{i,l,k} + g_{i,j,k}\right)/2, & l \in \{j + 1, j - 1\}, \\[2ex] -\sum_{n \in \{j-1, j+1\}} \left(g_{i,n,k} + g_{i,j,k}\right)/2, & l = j, \\[2ex] 0, & \text{otherwise.} \end{cases} \tag{8.20}$$

Boundary pixels are assumed to have only one neighbor, so the first row of A_x is

$$\left(-\frac{1}{2}(g_{i,1,k} + g_{i,2,k}), \frac{1}{2}\left(g_{i,1,k} + g_{i,2,k}\right), 0, \ldots, 0\right). \tag{8.21}$$

The last row is computed analogously. The operator A_x is computed individually and applied independently to each row in each frame of the video sequence.

Similarly, for a fixed column j in a fixed frame k, the A_y approximation has the form

$$(A_y)_{i,l} = \begin{cases} (g_{l,j,k} + g_{i,j,k})/2, & l \in \{i+1, i-1\}, \\ \\ -\sum_{n \in \{i-1,i+1\}} (g_{n,j,k} + g_{i,j,k})/2, & l = i, \\ \\ 0, & \text{otherwise.} \end{cases}$$
(8.22)

The A_y operators are computed individually and applied independently to each column in each frame of the video sequence.

Finally, for a fixed row i and fixed column j, the A_z approximation is given by

$$(A_z)_{k,l} = \begin{cases} (g_{i,j,l} + g_{i,j,k})/2, & l \in \{k+1, k-1\}, \\ \\ -\sum_{n \in \{k-1,k+1\}} (g_{i,j,n} + g_{i,j,k})/2, & l = k, \\ \\ 0, & \text{otherwise.} \end{cases}$$
(8.23)

The A_z operators are computed individually and applied independently to the time series for each pixel in the video sequence.

The boundary conditions for the operators A_y and A_z are analogous to those for A_x, following (8.21). Given these operators, the time update for the AOS scheme is

$$\Phi^{n+1} = \frac{1}{3} \sum_{s \in \{x,y,z\}} (\mathcal{I} - 3\alpha \Delta t A_s)^{-1} \left(\Phi^n + \beta \Delta t (c_2^n - c_1^n) \left(D - \frac{1}{2}(c_1^n + c_2^n) \right) \right).$$
(8.24)

8.4.3 Stability of Linear Evolution Equations

We define numerical stability for a finite-difference or finite-volume approximation to a parabolic partial differential equation, but first we classify the segmentation PDE.

Definition 8.4.1. *A partial differential equation of the form*

$$\frac{\partial \Phi(\mathbf{x},t)}{\partial t} = A(\mathbf{x})\Phi(\mathbf{x},t) + F(\mathbf{x},t), \tag{8.25}$$

for some linear differential operator A, is called a linear parabolic evolution equation *(see* [23], *Chapter 7).*

The segmentation PDE given by (8.7) is not quite linear, because the function

$$F^n = \beta(c_2^n - c_1^n)\left(D - \frac{1}{2}(c_1^n + c_2^n)\right)$$

implicitly depends on Φ. We can treat (8.7) as a linear equation, however, since F is uniformly bounded for all time independent of Φ, which implies that the dependence of F^n on Φ^n does not affect the amplification of numerical errors. Even though c_1 and c_2 depend on Φ, their values are uniformly bounded above and below by $\min D(\mathbf{x}) \leq c_1, c_2 \leq \max D(\mathbf{x})$. Thus

$$-\beta \max D(\mathbf{x})^2 \leq F^n \leq \beta \max D(\mathbf{x})^2$$

for all time.

Now we define stability of a numerical approximation to a linear parabolic PDE (see [45], Section 5.5).

Definition 8.4.2. *Suppose that B_1 and B_0 are $m \times m$ matrices and that*

$$B_1 \Phi^{n+1} = B_0 \Phi^n + F^n \tag{8.26}$$

is a finite-difference or finite-volume approximation to the solution of (8.25) *such that* B_1 *is of the form*

$$B_1 = \frac{1}{\Delta t} + T, \tag{8.27}$$

for some linear operator T *that is independent of* Δt. *Then* (8.26) *is numerically stable for all* $\Delta t > 0$ *if*

$$\left\| B_1^{-1} \right\| \leq C_1 \Delta t, \tag{8.28}$$

for some constant $C_1 > 0$, *and*

$$\left\| \left(B_1^{-1} B_0 \right)^n \right\| \leq C_2 \tag{8.29}$$

for all $n \in \mathbb{N}$ *and fixed* $C_2 > 0$.

Here $\| \cdot \|$ is the spectral matrix norm, i.e.

$$\|A\| = \sup_{\mathbf{x} \neq 0} \frac{\langle A\mathbf{x}, A\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

8.4.4 Stability of AOS

The unconditional stability of (8.24) is proven in [71] using scale-space theory, but it can be proven directly using Definition 8.4.2.

We define

$$\Phi_x^{n+1} = \left(\mathcal{I} - 3\alpha \Delta t A_x \right)^{-1} \left(\Phi^n + \Delta t\, F^n \right) \tag{8.30}$$

$$\Phi_y^{n+1} = \left(\mathcal{I} - 3\alpha \Delta t A_y \right)^{-1} \left(\Phi^n + \Delta t\, F^n \right) \tag{8.31}$$

$$\Phi_z^{n+1} = \left(\mathcal{I} - 3\alpha \Delta t A_z \right)^{-1} \left(\Phi^n + \Delta t\, F^n \right), \tag{8.32}$$

and note that $\Phi^{n+1} = \frac{1}{3} \left(\Phi_x^{n+1} + \Phi_y^{n+1} + \Phi_z^{n+1} \right)$. It will be shown that the numerical schemes given in (8.30)-(8.32) are unconditionally stable, so their mean is also stable.

Equation (8.30) can be rewritten

$$\left(\frac{1}{\Delta t}\mathcal{I} - 3\alpha A_x\right)\Phi_x^{n+1} = \frac{1}{\Delta t}\Phi^n + F^n. \tag{8.33}$$

Let $B_1 = \frac{1}{\Delta t}\mathcal{I} - 3\alpha A_x$, and note that this satisfies (8.27). Let $B_0 = \frac{1}{\Delta t}\mathcal{I}$. We must show that B_1 and B_0 as defined here satisfy (8.28) and (8.29).

First we show that (8.28) is satisfied.

Lemma 8.4.3. *The matrix operators* $(-3\alpha A_s) : \mathbb{R}^m \to \mathbb{R}^m$, *for* $s \in \{x, y, z\}$, *are symmetric and positive semi-definite.*

Proof. By the definitions given in (8.20), (8.22), and (8.23), it is clear that the operators A_x, A_y, and A_z are symmetric, so therefore $-3\alpha A_x$, $-3\alpha A_y$, and $-3\alpha A_z$ are also symmetric. It is left to be shown that they are positive semi-definite.

Let $\mathbf{x} \in \mathbb{R}^m$ be nonzero. Then $-3\alpha A_x$ is positive semi-definite if $\mathbf{x}^T (-3\alpha A_x) \mathbf{x} \geq 0$. Write $\mathbf{x} = (x_1, x_2, \ldots, x_m)^T$. Then

$$(-3\alpha A_x)\mathbf{x} = (-3\alpha A_x)(x_1, x_2, \ldots, x_m)^T$$

$$= \begin{pmatrix} \frac{3}{2}\alpha\,(g_{i,1,k} + g_{i,2,k})\,(x_1 - x_2) \\ \frac{3}{2}\alpha\,[(g_{i,1,k} + g_{i,2,k})\,(x_2 - x_1) + (g_{i,2,k} + g_{i,3,k})\,(x_2 - x_3)] \\ \vdots \\ \frac{3}{2}\alpha\,[(g_{i,m-2,k} + g_{i,m-1,k})\,(x_{m-1} - x_{m-2}) + (g_{i,m-1,k} + g_{i,m,k})\,(x_{m-1} - x_m)] \\ \frac{3}{2}\alpha\,(g_{i,m,k} + g_{i,m-1,k})\,(x_m - x_{m-1}) \end{pmatrix},$$

so, since g is non-negative,

$$\mathbf{x}^{\mathrm{T}}\left(-3\alpha A_x\right)\mathbf{x} = x_1 \left(\frac{3}{2}\alpha\left(g_{i,1,k} + g_{i,2,k}\right)\left(x_1 - x_2\right)\right)$$

$$+ x_2 \left(\frac{3}{2}\alpha\left[\left(g_{i,1,k} + g_{i,2,k}\right)\left(x_2 - x_1\right) + \left(g_{i,2,k} + g_{i,3,k}\right)\left(x_2 - x_3\right)\right]\right)$$

$$+ \ldots + x_m \left(\frac{3}{2}\alpha\left(g_{i,m,k} + g_{i,m-1,k}\right)\left(x_m - x_{m-1}\right)\right)$$

$$= \frac{3}{2}\alpha\left(g_{i,1,k} + g_{i,2,k}\right)\left(x_1 - x_2\right)^2 + \frac{3}{2}\alpha\left(g_{i,2,k} + g_{i,3,k}\right)\left(x_2 - x_3\right)^2$$

$$+ \ldots + \frac{3}{2}\alpha\left(g_{i,m-1,k} + g_{i,m,k}\right)\left(x_{m-1} - x_m\right)^2$$

$$\geq 0.$$

Thus $-3\alpha A_x$ is symmetric and positive semi-definite. A similar argument shows that $-3\alpha A_y$ and $-3\alpha A_z$ are also positive semi-definite. $\qquad\square$

Since $-3\alpha A_s$ is symmetric and positive semi-definite, its eigenvalues are real and non-negative. Thus $B_1 = \frac{1}{\Delta t}\mathcal{I} - 3\alpha A_s$ is symmetric and positive-definite, and its eigenvalues have the form

$$\frac{1}{\Delta t} + \lambda = \frac{1 + \lambda\Delta t}{\Delta t},$$

where λ is an eigenvalue of $-3\alpha A_x$. Thus the eigenvalues of B_1^{-1} satisfy

$$\left(\frac{1 + \lambda\Delta t}{\Delta t}\right)^{-1} = \frac{\Delta t}{1 + \lambda\Delta t} \leq \Delta t.$$

Since B_1^{-1} is symmetric and all of its eigenvalues are real numbers less than or equal to Δt, $\left\|B_1^{-1}\right\| \leq \Delta t$. This shows that B_1 satisfies (8.28) with $C_1 = 1$.

Now we show that (8.29) is also satisfied. This follows from

$$\left\|\left(B_1^{-1}B_0\right)^n\right\| \leq \left\|B_1^{-1}\right\|^n \|B_0\|^n \leq (\Delta t)^n \left(\frac{1}{\Delta t}\right)^n = 1.$$

Thus (8.29) is satisfied with $C_2 = 1$.

Since (8.28) and (8.29) are both satisfied, the numerical scheme in (8.30) is stable for all $\Delta t > 0$. A similar argument holds for the schemes in (8.31) and (8.32). Since all three of these schemes are stable, their mean is also stable. Thus (8.19) is a stable numerical scheme for approximating the solution to (8.7) for all $\Delta t > 0$.

8.4.5   Convergence of AOS

Here we show that the AOS scheme in (8.24) converges to the solution of (8.7) as $\Delta t \to 0$ for a fixed ratio $\Delta t / \max\{\Delta x, \Delta y, \Delta z\}$. In order to show this, we first require a definition.

Definition 8.4.4. *Let u be the exact solution to a linear parabolic evolution equation as in (8.25), and let B_1 and B_0 be as in (8.26). The truncation error, T^n, of the numerical scheme is given by*

$$T^n = B_1 u^{n+1} - [B_0 u^n + F^n],$$

where $u^n = u(t_n)$. The numerical scheme determined by B_1 and B_0 is called consistent *if $T^n \to 0$ as $\Delta t \to 0$ for a fixed ratio $\Delta t / \max\{\Delta x, \Delta y, \Delta z\}$.*

We continue with a special case of a classical theorem of numerical analysis, the Lax-Equivalence Theorem ([45], Theorem 5.1).

Theorem 8.4.5. *If the numerical scheme given by (8.26) is consistent, satisfies $\left\| B_1^{-1} \right\| \leq 1$, and is stable, then it is convergent.*

It has already been shown that the AOS scheme in (8.24) is stable and satisfies $\left\| B_1^{-1} \right\| \leq 1$. Thus, in order to show that the AOS scheme is convergent, it is only left to be shown that this scheme is consistent. Again, we will show that each of the schemes in (8.30)-(8.32) are

consistent and that their sum is also consistent. First of all, noting (8.15),

$$A_x \to \frac{\partial}{\partial x} g \frac{\partial}{\partial x} \quad \text{as} \quad \Delta x \to 0.$$

Thus, for a fixed ratio $\Delta t / \Delta x$,

$$\left(\frac{1}{\Delta t} - 3\alpha A_x \right) u^{n+1} - \frac{1}{\Delta t} u^n - F^n \to \frac{\partial u}{\partial t} - 3\alpha \left(\frac{\partial}{\partial x} g \frac{\partial}{\partial x} \right) u - F^n.$$

A similar results holds for (8.31) and (8.32). Thus, since u is the exact solution for (8.7),

$$\frac{3}{\Delta t} \left(u^{n+1} - u^n \right) - 3\alpha \left(A_x + A_y + A_z \right) u^n - 3F^n \to 3\frac{\partial u}{\partial t} - 3\alpha \mathrm{div} \left(g\nabla u \right) - 3F = 0.$$

Thus the AOS scheme is consistent, and, therefore, by Theorem 8.4.5, convergent.

8.5 Advantages of AOS

There are two reasons that the AOS scheme in (8.24) is computationally advantageous over the semi-implicit scheme (8.17). Since the operators $\mathcal{I} - 3\alpha\Delta t A_s$ are strictly diagonally dominant and tridiagonal, the well-known Thomas algorithm - a special form of Gaussian elimination for strictly diagonally dominant, tridiagonal systems - can be used to solve the linear systems

$$(\mathcal{I} - 3\alpha\Delta t A_s) \Phi^{n+1} = \Phi^n + \beta\Delta t(c_2^n - c_1^n) \left(D - \frac{1}{2}(c_1^n + c_2^n) \right)$$

for Φ^{n+1}. This is a fast method for linear system backsolves, since the computational complexity of the method grows only linearly with the number of grid points. Also, each of the linear systems to be solved at each numerical iteration is independent of each of the others. Thus each of the backsolves can be performed in parallel. The operator A_x acts independently on each row of each frame in the data set. Likewise A_y acts on each column independently, and A_z acts on each pixel's time series independently. This means each of the linear systems

to be solved is small, but there are a large number of linear systems to be solved at each numerical iteration (708608 for the first example in the next chapter). Such a large number of linear system solves would be computationally problematic if they could not be done in parallel. As it is, they can be solved in parallel, and the method is computationally efficient.

The second primary advantage to the AOS scheme is that it is unconditionally stable with respect to Δt. Thus Δt can be chosen to be quite large, so that few iterations are required. Whereas each numerical iteration of the explicit method is much faster than a numerical iteration of the AOS method, using AOS requires on the order of 10 numerical iterations in order to compute the steady-state solution of (8.7) as opposed to on the order of 10^6 iterations required for the explicit method.

8.6 Implementation Details

The numerical method described in this chapter was implemented using ANSI C/C++ with no third-party libraries. Computations were performed on an SGI Origin 2000, administered by the University of Montana *Molecular Computational Core Facility*, with 32 processors (300 Mhz) and 32 gigabytes of RAM. Parallelization was achieved using the OpenMP shared-memory parallelization library. Testing was performed using Matlab and C/C++ on a PC running SUSE Linux with a 1 Ghz AMD Opteron-64 and 4 gigabytes of RAM.

8.7 Model Validation

In order to demonstrate that the segmentation algorithm and numerical scheme here presented segment gray-scale data correctly, we demonstrate the algorithm on binary data. The data used is generated using a two-dimensional cellular neural network (CNN) density classifier [16]. The initial state, i.e. the first image in the video sequence, is a random pattern of 0's

and 1's, and the evolution in time is determined by the density classifier. The final state has all 0's. Each image in the video sequence is 200 × 200 pixels, and there are 375 images in the sequence. Though the patterns produced are not identical to the patterns generated by stomatal patches, this data is useful for demonstrating the power of the segmentation algorithm. Since the data is binary, it is assumed that the background model is already subtracted. Thus only the segmentation component of our algorithm is demonstrated and not the background model generation. Example images for the CNN segmentation can be seen in Figure 8.2. Note that the boundary separating the black and gray regions is captured very well. For this segmentation $\alpha = 1$, $\beta = 2$, and $\Delta t = 0.15$. Only 3 iterations of the numerical scheme were required for this data, so the example images are shown at $t = 0.45$.

Image 50

Image 100

Image 150

Image 200

Image 250

Image 300
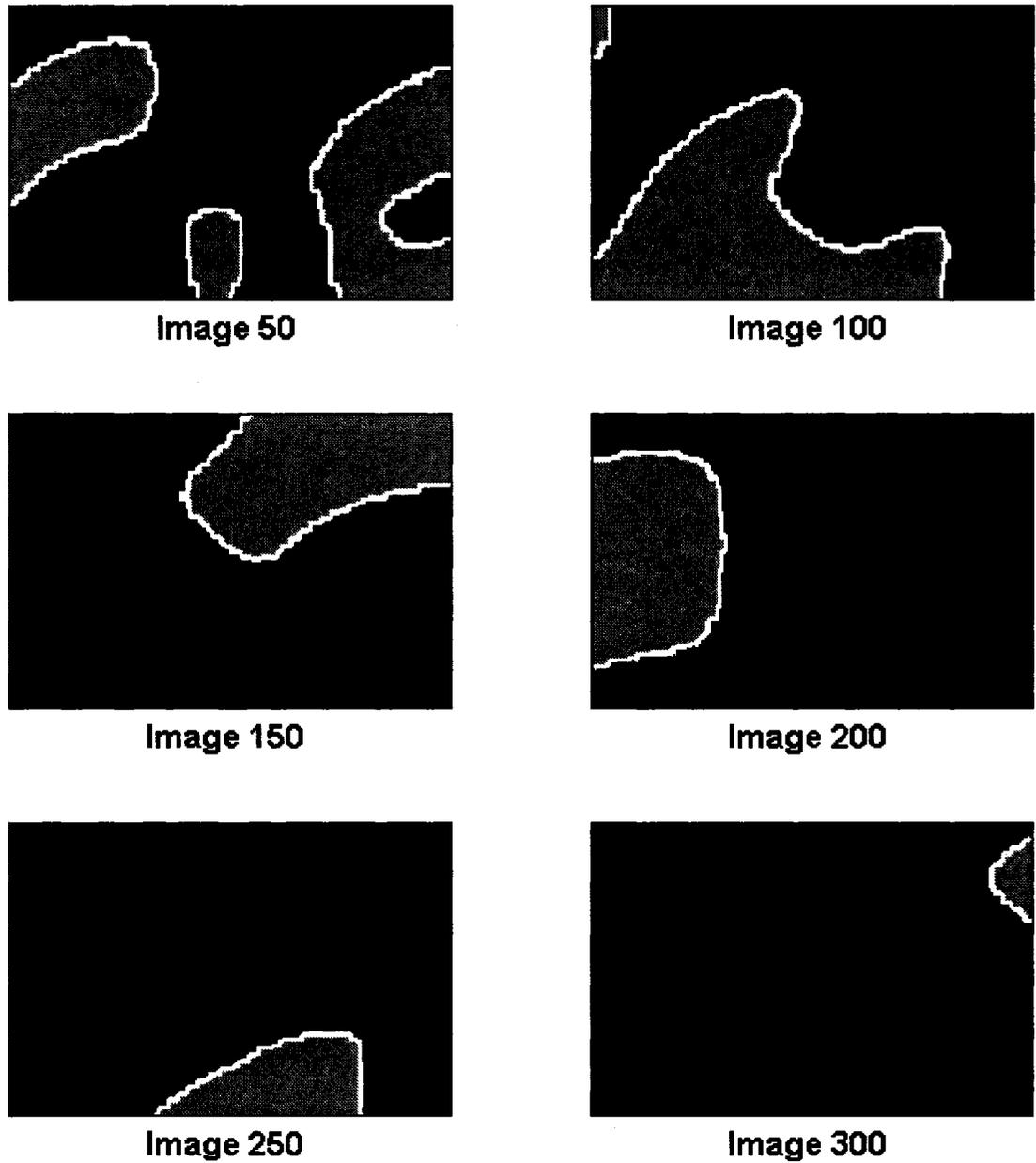
Figure 8.2: Segmentation of CNN Images

Chapter 9

Results and Future Work

In this chapter we present the results of the video segmentation algorithm. We also present some preliminary pattern analysis results and outline future work.

9.1 Results

We present the results of the video segmentation in two formats. First, we view the segmentation by overlaying the solution surface directly on the data $D(\mathbf{x})$ and $I(\mathbf{x})$. The segmentation contour in each frame is the cross-section of the solution surface for that time. The primary goal of the segmentation is to analyze the pattern dynamics of the fluorescing regions, so we also compute the binary patterns that the segmentation yields.

9.1.1 Contour Segmentation

The result of the video segmentation algorithm is a function Φ that is the steady-state solution to the parabolic form of the segmentation PDE (8.7). The zero level set of Φ is the surface

99

that segments the video domain into fluorescing regions and non-fluorescing regions. First, we view the zero level set overlayed directly on the data $D(\mathbf{x})$. Figures 9.1 through 9.3 show the segmentation contours overlayed on $D(\mathbf{x})$. The results were obtained using $\alpha = 1.0$, $\beta = 2.0$, and $\Delta t = 0.15$, and the steady-state was computed after 15 iterations ($t = 2.25$). In these images, the data is scaled so that the bright regions appear gray. The segmenting contours are in white. These example images come from a sequence of 436 frames that are 512×512 pixels. This particular sequence is indicative of the stomatal patchiness being investigated. The contours can also be viewed directly on the video data $I(\mathbf{x})$, as can be seen in Figures 9.4 through 9.6.

In image 400, a bright region is segmented along the largest vein in the lower left of the image. This is not an error of the segmentation but an artifact caused by errors in the background model. In order to eliminate this region - since it is not actually fluorescing - it is necessary to adjust the background model around the large veins. That is not done here.

In Figure 9.7, we show the same segmentation for Image 2. Note that there is no segmentation contour on this image. If there were a contour, that would indicate that there are fluorescing regions in that image. Since there is no contour, the video segmentation algorithm has correctly computed that there are no fluorescing regions, as there are no dynamics in the early frames. If the same segmentation algorithm were applied to each frame in the video sequence independently in two dimensions, an incorrect result would be obtained for Image 2. It is necessary for the algorithm to compare the scales of fluorescence across the entire video sequence, so the solution must be computed in three dimensions on the entire video sequence.

Results from a second leaf video sequence of 500 images can be seen in Figures 9.10 and 9.11, again using the parameter values $s_1 = 15$, $s_2 = 0.4$, $\alpha = 1.0$, $\beta = 2.0$, and $\Delta t = 0.15$. This sequence demonstrates initial stomatal patchiness, but the steady-state where all stomatal apertures are uniformly adjusted is attained after 250 frames.
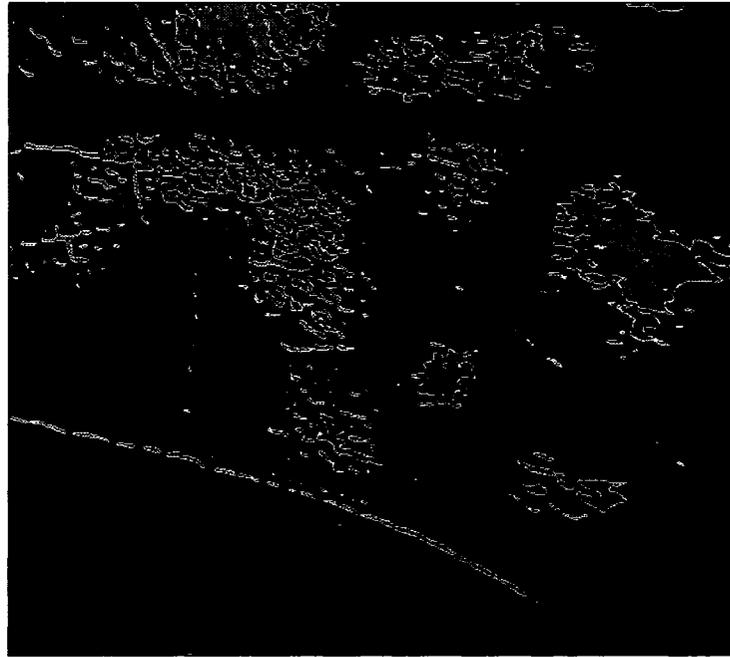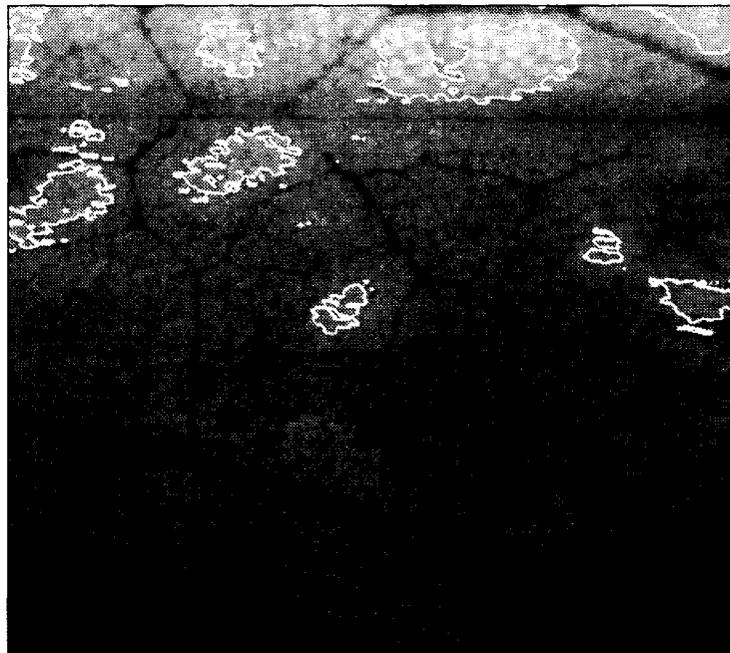
Image 150

Figure 9.1: $D(\cdot, \cdot, 150)$ with Segmentation Contour Overlayed

Image 275

Figure 9.2: $D(\cdot, \cdot, 275)$ with Segmentation Contour Overlayed

Image 400

Figure 9.3: $D(\cdot, \cdot, 400)$ with Segmentation Contour Overlayed

Image 150

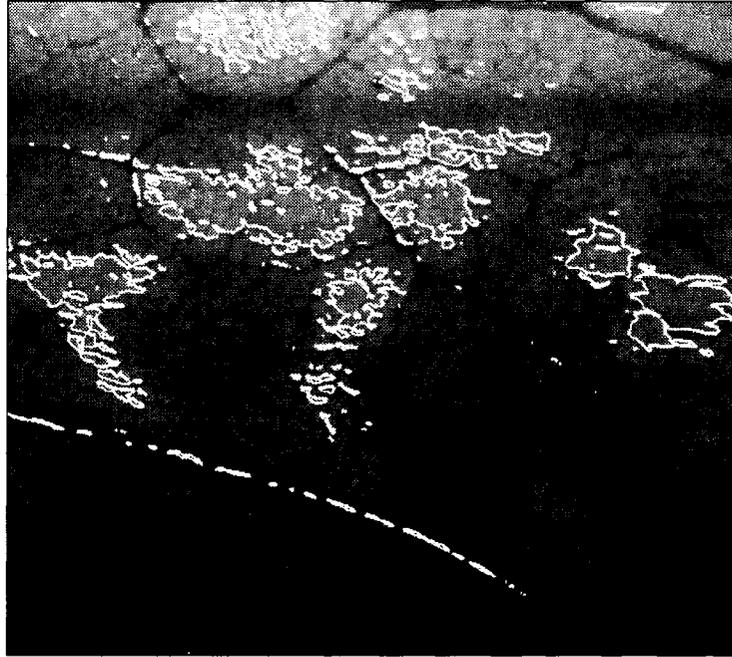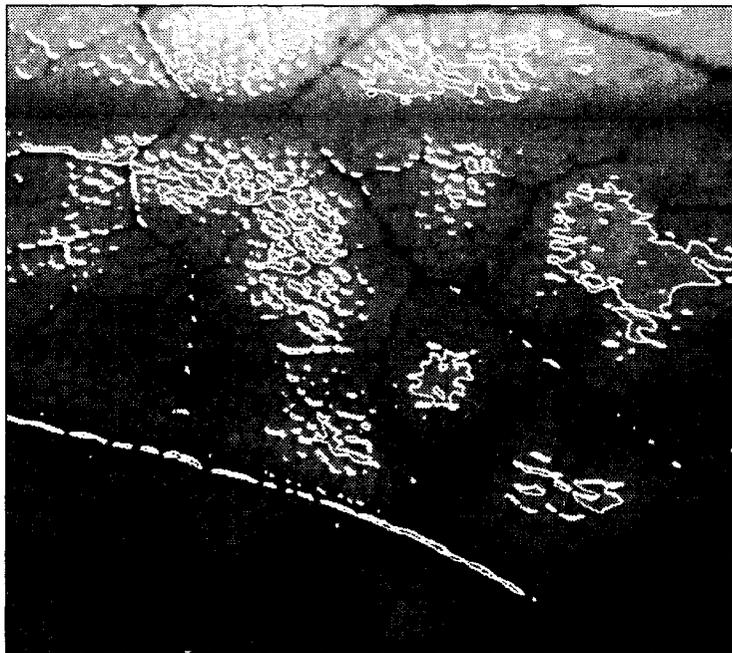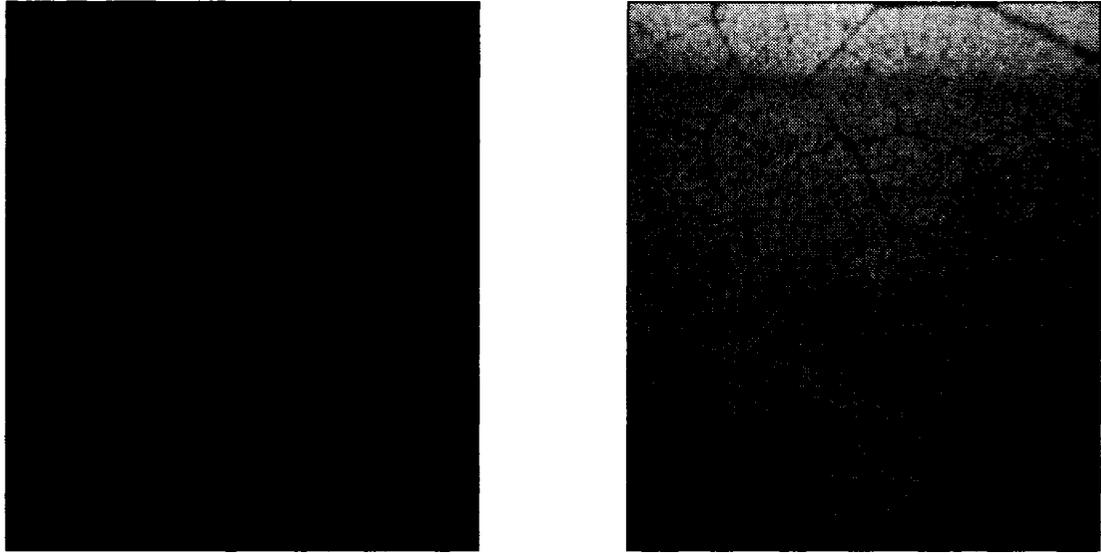Figure 9.4: $I(\cdot, \cdot, 150)$ with Segmentation Contour Overlayed

Image 275

Figure 9.5: $I(\cdot, \cdot, 275)$ with Segmentation Contour Overlayed

Image 400

Figure 9.6: $I(\cdot, \cdot, 400)$ with Segmentation Contour Overlayed

Image 2

Figure 9.7: $D(\cdot,\cdot,2)$ and $I(\cdot,\cdot,2)$ with Segmentation Contour Overlayed

9.1.2 Sensitivity to Parameters

All of the segmentations in Figures 9.1-9.6 were computed using the same set of parameter values. The segmentation is not, however, very sensitive to the choice of the parameter values α and β. The important quantity is the ratio of α to β, and example images are presented in Figure 9.8 computed using several different values for α and β. Note that the 6 examples represent ratios of α to β of $1/5, 1/3, 1/2, 1, 2, 3, 5$. As can be seen in the figure, the segmentation is similar for each set of parameters such that $\alpha/\beta \leq 1$. For the cases where $\alpha >$ β, the regularization dominates the segmentation and the correct regions are not segmented. Thus the algorithm is sensitive only to which factor is dominating the segmentation, and, as long as the segmentation is the dominant term (i.e. $\beta > \alpha$), the algorithm is robust to the actual value of the ratio α/β.
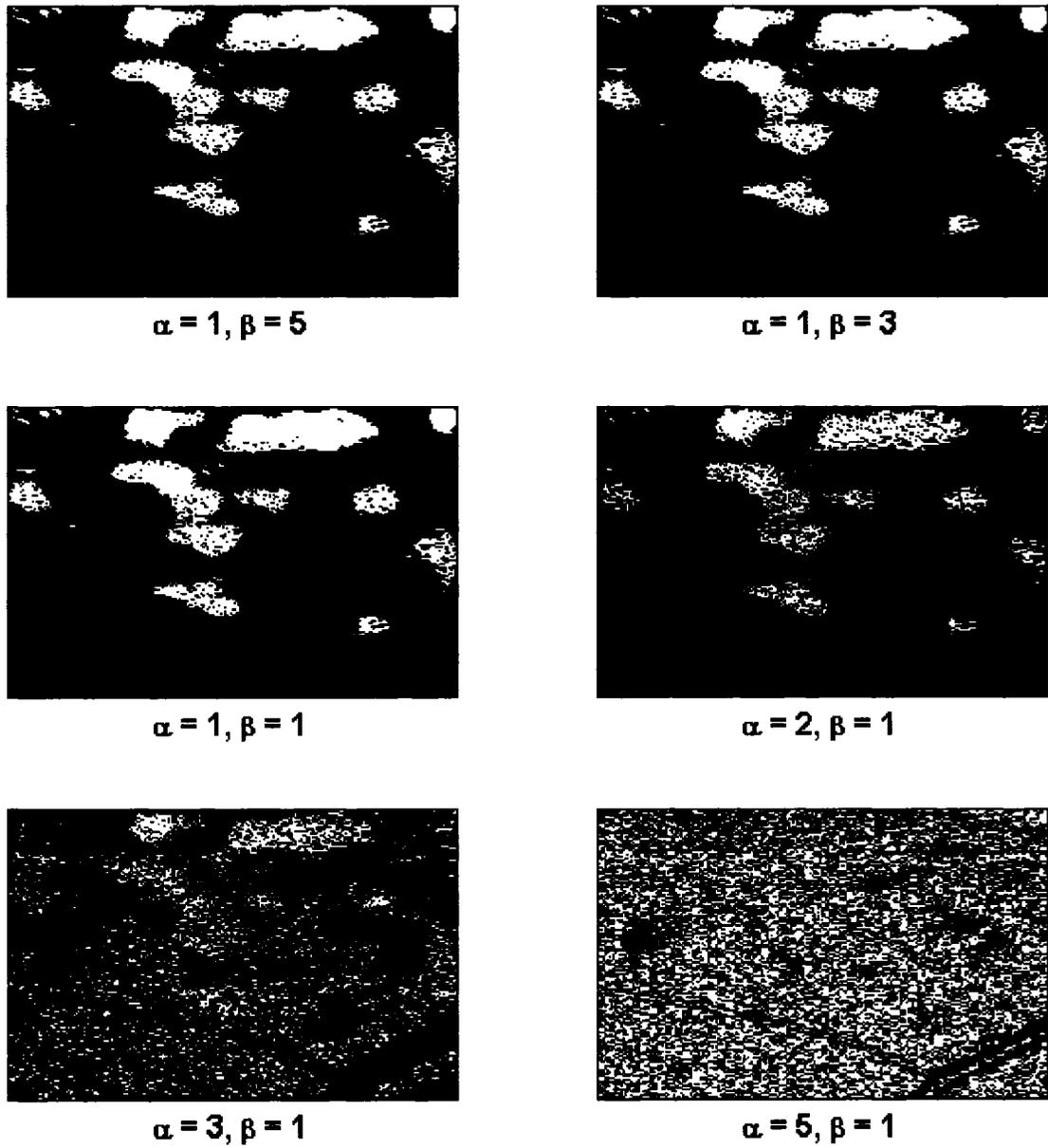
Figure 9.8: Segmentation of Image 300 with Differing Values of α and β

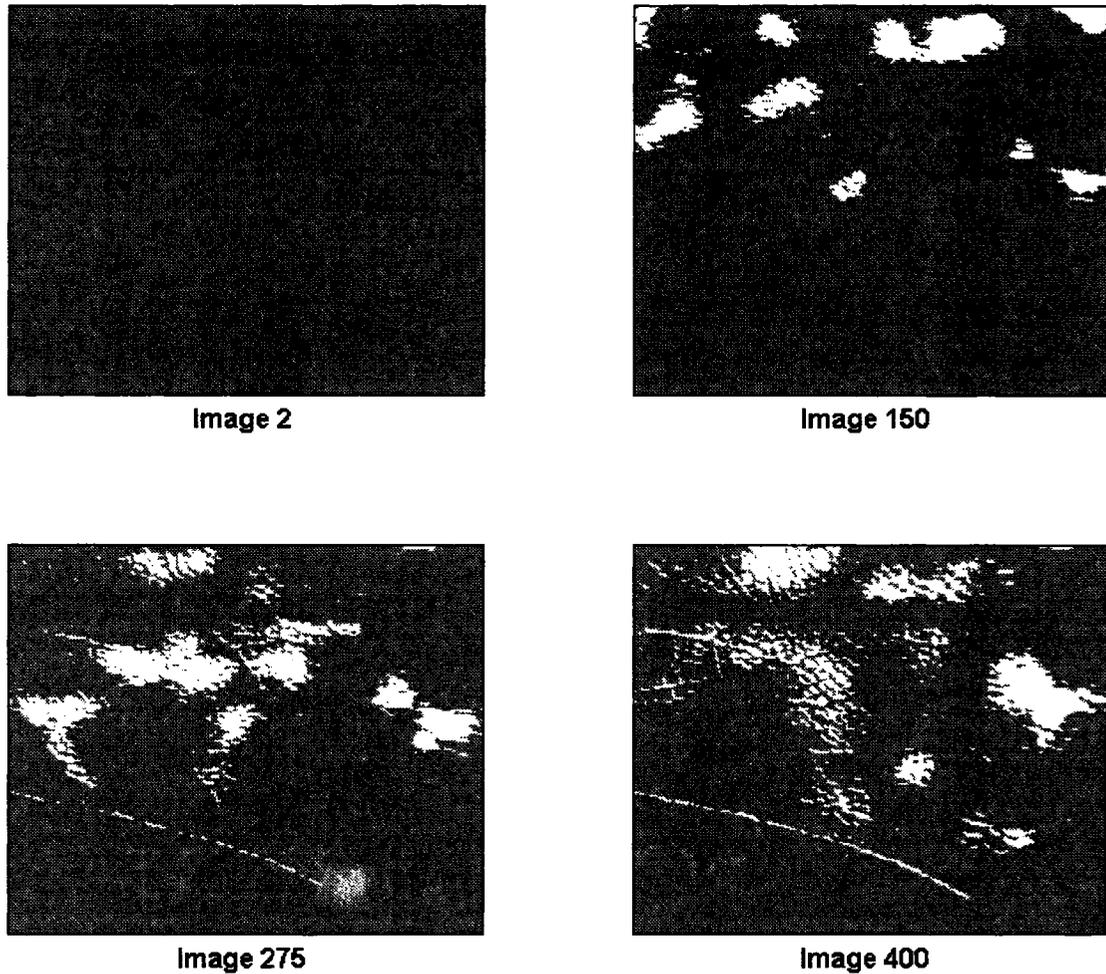Image 2   Image 150

Image 275   Image 400

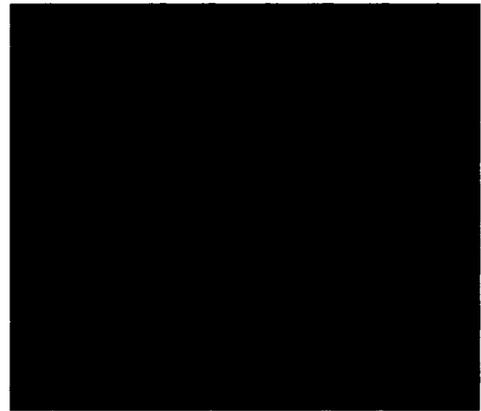Figure 9.9: Binary Segmented Images

9.1.3 Binary Segmentation and Pattern Analysis

In order to analyze the dynamics of the two-dimensional patterns seen in a leaf video, it is necessary to look at a binary form of the segmentation. Sample images can be seen in Figure 9.9. The white regions in the binary segmentation correspond to the areas of the leaf that are actively fluorescing, and the gray regions of each image correspond to areas of the leaf that are not fluorescing.

Given the binary segmentation, it is possible to use the methods of two-dimensional pattern
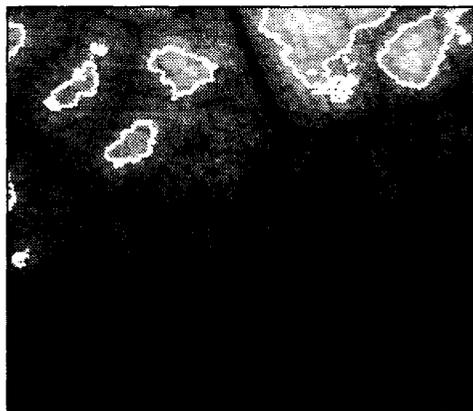
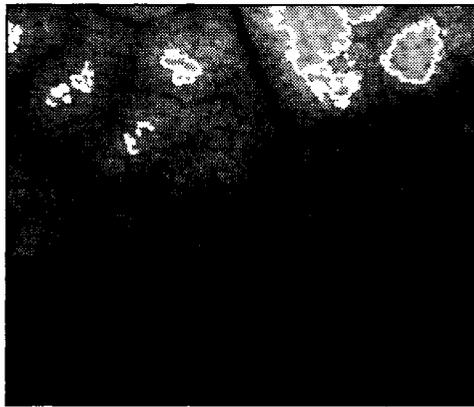Figure 9.10: Segmentation and Binary Results from a Second Leaf Video Sequence
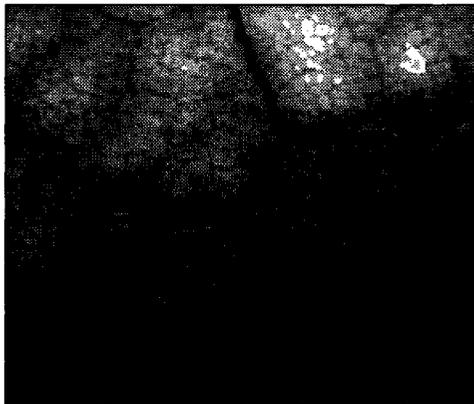
Image 150

Binary Segmentation

Image 200

Binary Segmentation

Figure 9.11: More Segmentation and Binary Results from Second Leaf Video Sequence
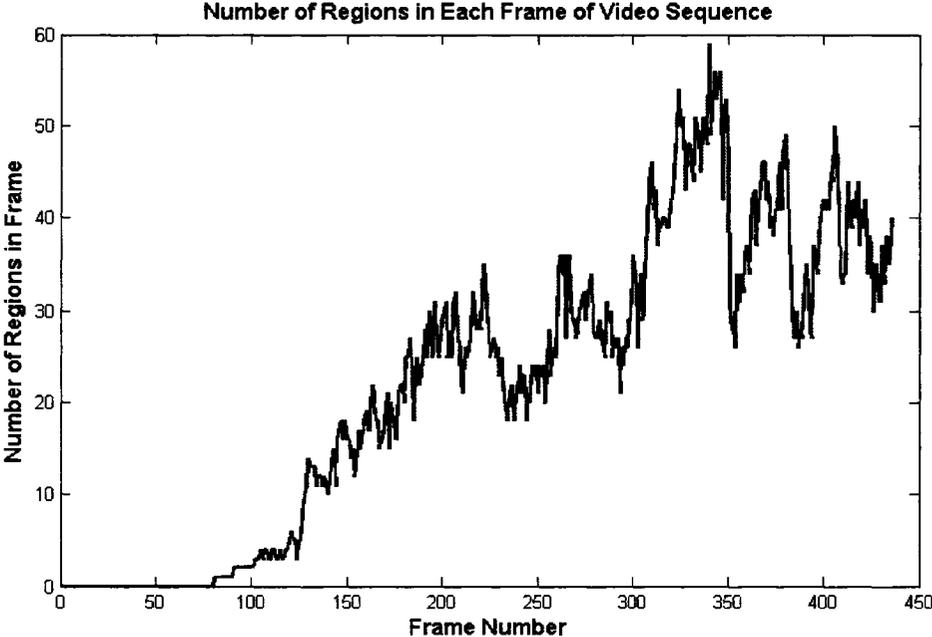
Number of Regions in Each Frame of Video Sequence

Figure 9.12: Number of Regions of Size 100 Pixels or Larger in Each Frame
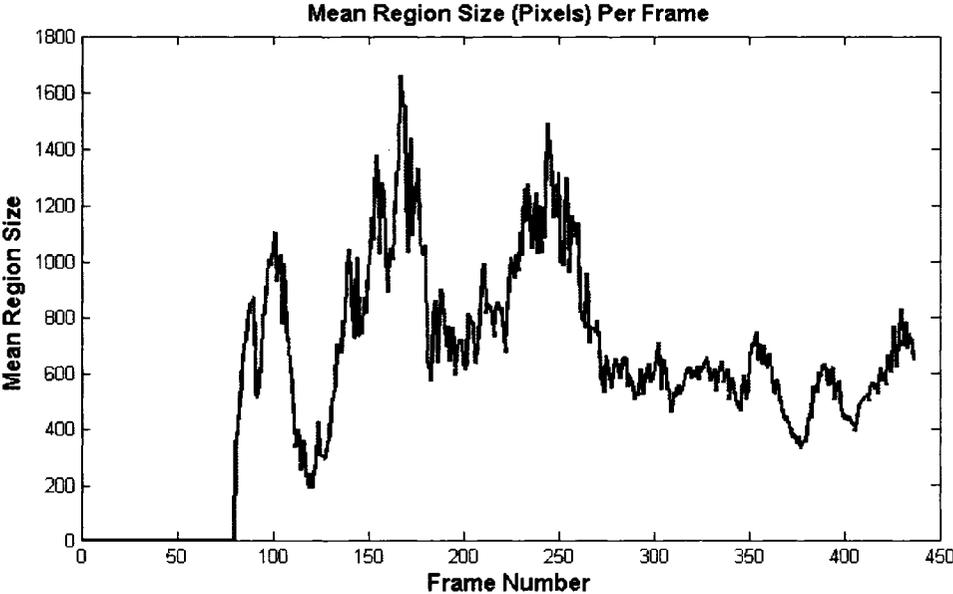
Mean Region Size (Pixels) Per Frame

Figure 9.13: Mean Region Size Per Frame

analysis to develop models based on statistical mechanics that can generate similar patterns. The first step is to reduce each two-dimensional pattern to a single measure, so the entire video sequence can be analyzed as a time series. There are several different measures that can be used, but the first is the number of fluorescing regions in each frame. A plot of the time series can be found in Figure 9.12. While the number of regions steadily increases over the first half of the experiment, it seems to stabilize after image 300. In the last 136 frames the number of regions oscillates but the oscillations decrease in amplitude. For this computation only regions containing 100 or more pixels are counted.

The size of each of these regions is also important, and a second useful measure is therefore the mean region size in each image. Figure 9.13 shows the time series associated with mean region size. Once the dynamics are initiated, the mean region size oscillates heavily, but after frame 275 the oscillations decrease substantially in amplitude. Thus the mean region size, as well as the number of regions, stabilizes towards the completion of the experiment.

9.2 Future Work

We can take several different directions in our future analysis. The first step in improving the segmentation algorithm is computing a more accurate background model. This will eliminate the incorrect segmentation along the veins. Improving the segmentation to yield more regular patterns, i.e. patterns with smoother boundaries, is another important step. More regular patterns are more likely to be modeled using known systems in mechanics. One way of producing patterns with smoother boundaries is by denoising the data using a Gaussian convolution. The bright regions in the non-linearly scaled data $D(\mathbf{x})$ would then have a more regular shape. As was noted in Chapter 3, this would remove all information about the vein boundaries but might yield patterns that can be modeled more readily.

A second improvement to the numerical segmentation algorithm could be accomplished by

incorporating the reinitialization directly into the segmentation functional as in [38]. This would remove the need for the reinitialization step. A drawback to such a method is that it requires an explicit numerical implementation, so the semi-implicit scheme outlined here would no longer be applicable.

As far as the binary patterns are concerned, a more sophisticated analysis of pattern dynamics can yield more information as to the possible physical system that produces the patterns seen in the leaf videos. For example, one can compute the two-dimensional Shannon entropy and an associated two-dimensional *excess entropy* [24] - using approximation methods for the Shannon entropy of the two-dimensional images [57] - in order to determine the scale of correlations among fluorescing pixels. This would yield further insight into the coupling between neighboring regions of fluorescence. Further analysis of stomatal patterns can be done using topological properties of the patterns, such as measuring higher-order homology [25]. One can also perform an archetypal analysis [65] to determine the characteristic patterns of the stomatal patches.

Chapter 10

Conclusion

Leaves have pores in their surfaces called stomata that open and close to regulate CO_2 uptake. Water is lost through open stomata via evaporation. It is believed that for a fixed amount of water loss a leaf will globally optimize CO_2 uptake, but it is not known how the stomata in a leaf regulate their apertures in order to accomplish this. It has been observed in over 200 plant species that neighboring stomata form patches across which apertures are uniformly regulated. These patches often result in less than optimal local CO_2 uptake, and it is not understood how such patches can produce a maximum in global CO_2 uptake. In order to visualize stomatal apertures, a video is taken of chlorophyll fluorescence in the leaf. A dye is injected into the leaf that fluoresces when reacting with CO_2. This leads to an inverse relationship between fluorescence and photosynthesis, which implies an inverse relationship between fluorescence and stomatal aperture.

In order to analyze the fluorescence patterns in a leaf video, it is first necessary to extract the regions in each frame that are actively fluorescing. A method for the segmentation of these regions is the primary focus of this work.

112

10.1 Preprocessing

Before the fluorescing regions can be extracted, the leaf video must first be preprocessed. There are three steps to the preprocessing: denoising, median-change normalization, and scaling. The denoising process removes the high-frequency instrumentation noise associated with the capture of digital signals. We use an L^1-fidelity variant the ROF scheme, which is a method that denoises locally within regions of relatively homogeneous intensity while preserving sharp edges. Preserving the edges that form the vein boundaries is necessary in order to determine the effect that veins have on the dynamics of the stomatal patches. The median-change normalization is designed to remove low-frequency noise effects caused by uncontrolled lighting conditions in the lab. These effects are not known *a priori*, but must be removed nonetheless. By assuming that the median intensity change is caused by global effects, these effects can be offset by subtracting them out of each frame in the video sequence. The last preprocessing step is a linear scaling of the video data.

10.2 Regular Surface Evolution

The video segmentation technique for extracting the fluorescing regions of the leaf video is a level set-based, variational formulation. It must be shown that such a method can yield the desired set of regions, and, in order to show this, we develop surface manifolds and regular surface evolutions. A regular surface evolution is a level set representation of a time-differentiable deformation of one surface into another, and it is shown that the gradient descent method for surface functionals yields a regular surface evolution. In particular, the regular surface evolutions corresponding to the steepest descent method for surface functionals of the

form

$$E(M) = \int_M g(\mathbf{x}, \Phi, \nabla\Phi) \, dS,$$

$$E(M) = \int_{\mathbb{R}^n} f(\mathbf{x}, \Phi, \nabla\Phi) \, d\mathbf{x}$$

are computed.

10.3 Background Model

The first step in developing the segmentation algorithm is to produce a background model, $B(\mathbf{x})$, for the video data, $I(\mathbf{x})$. Since the stomatal dynamics in a leaf experiment are initiated by the experimentalist, it is possible to take several images of the leaf prior to the start of the fluorescence. These frames are then averaged pixel-wise to produce an image of the leaf with no fluorescence. Thus, rather than segmenting based directly on the image intensities in the video sequence, we segment based on the absolute difference between the image intensities and the background model, $D(\mathbf{x}) = |I(\mathbf{x}) - B(x, y)|$. The resulting data is bright in regions corresponding to fluorescence and dark elsewhere. In order to further enhance the segmentation, the difference data is nonlinearly scaled to make the bright patches brighter and the dark patches darker.

10.4 Video Segmentation

The goal is to extract the bright regions of $D(\mathbf{x})$, and this can be done by computing an approximation, $u(\mathbf{x})$, to the data such that $u(\mathbf{x})$ is piecewise-constant taking on exactly two

values. This approximation is computed by minimizing the functional

$$E(u) = \int_\Omega (D(\mathbf{x}) - u(\mathbf{x}))^2 \, d\mathbf{x},$$

which is just the distance from D to u in $L^2(\Omega)$. This is a surface functional whose gradient descent is a regular surface evolution, so it can be minimized if it is reformulated in a level set framework. A regularization term is added to the functional to ensure the existence of minimizers.

10.5 Results

The segmentation algorithm produces contours that form the boundaries of the fluorescing regions in the video sequence. These contours can be viewed directly on the data in order to visualize the segmentation, but the more useful results are the binary images that are produced by setting the fluorescing regions to be white and all other regions to black (or gray). These images can then be analyzed using the methods of two-dimensional pattern analysis, such as computing the number of fluorescing regions and their sizes.

Possible future work includes more sophisticated methods of pattern analysis, including computing pattern measures such as two-dimensional mutual information, Shannon entropy, and excess entropy.

Bibliography

[1] Simon A. Levin (ed.), *Mathematics and Biology: The Interface*, Netherlands Society for Theoretical Biology. Theory and Mathematics in Biology and Medicine (TMBM - 99).

[2] Grégoire Allaire, François Jouve, and Anca-Maria Toader, *A Level-set Method for Shape Optimization*, C. R. Acad. Sci. Paris, Serie I **334** (2002), 1–6.

[3] Luigi Ambrosio, Nicola Fusco, and Diego Pallara, *Functions of Bounded Variation and Free Discontinuity Problems* (J. M. Ball, E. M. Friedlander, I. G. MacDonald, L. Nirenberg, R. Penrose, and J. T. Stuart, eds.), Oxford Mathematical Monographs, Clarendon Press, 2000.

[4] M. J. Black, G. Sapiro, D. Marimont, and D. Heeger, *Robust Anisotropic Diffusion*, IEEE Transactions on Image Processing **7** (1998), no. 3, 421–432. Special issue on Partial Differential Equations and Geometry Driven Diffusion in Image Processing and Analysis.

[5] Michael J. Black and Yaser Yacoob, *Tracking and Recognizing Rigid and Non-Rigid Facial Motions Using Local Parametric Models of Image Motion*, Proceedings of International Conference on Computer Vision, 1995, pp. 374–381.

[6] A. Blake, *Introduction to Active Contours and Visual Dynamics*, Web reference, 1999.

[7] Martin Burger, *A Level Set Method for Inverse Problems*, Inverse Problems **17** (2001), 1327–1356.

[8] _____, *A Framework for the Construction of Level Set Methods for Shape Optimization and Reconstruction*, Interfaces and Free Boundaries **5** (2003), 301–329.

[9] J. Canny, *A Computational Approach to Edge Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **8** (1986), no. 6.

[10] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro, *On Geodesic Active Contours*, International Journal of Computer Vision **22** (1997), no. 1, 61–79.

[11] Michele Ceccarelli, Valentina De Simone, and Almerico Murli, *Well-posed Anisotropic Diffusion for Image Denoising*, IEEE Proceedings on Vision, Image and Signal Processing **149** (2002), no. 4, 244–252.

116

[12] Tony F. Chan and Selim Esedoglu, *Aspects of Total Variation Regularized L^1 Function Approximation*, SIAM Journal on Applied Mathematics **65** (2005), no. 5, 1817–1837.

[13] Tony F. Chan and Jianhong Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, 2005.

[14] Tony F. Chan and Luminita A. Vese, *A Level Set Algorithm for Minimizing the Mumford-Shah Functional in Image Processing*, Proceedings of the First IEEE Workshop on Variational and Level Set Methods in Computer Vision, 2001, pp. 161–168.

[15] _____, *Active Contours Without Edges*, IEEE Transactions on Image Processing **10** (2001), no. 2, 266–277.

[16] L. O. Chua and L. Yang, *Cellular Neural Networks: Theory*, IEEE Transactions on Circuits and Systems **35** (1988), no. 10.

[17] Bernard Dacorogna, *Introduction to the Calculus of Variations*, Imperial College Press, 2004.

[18] G. Dal Maso, J. M. Morel, and S. Solimini, *A Variational Method in Image Segmentation: Existence and Approximation Results*, Acta Mathematica **168** (1992), 89–151.

[19] Ennio De Giorgi, *Free Discontinuity Problems in Calculus of Variations*, Frontiers in Pure and Applied Mathematics, A Collection of Papers Dedicated to J. L. Lions on the Occasion of his 60th Birthday (1991), 55–62.

[20] Manfredo Do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, 1976.

[21] R. E. Edwards, *Functional Analysis: Theory and Applications*, Holt, Rinehart, and Winston, 1963.

[22] B. Engquist and S. J. Osher, *Stable and Entropy-Satisfying Approximations for Transonic Flow Calculations*, Mathematics of Computation **134** (1980), no. 149, 45–75.

[23] Lawrence C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, vol. 19, American Mathematical Society, 2000.

[24] David P. Feldman and James P. Crutchfield, *Structural Information in Two-Dimensional Patterns: Entropy Convergence and Excess Entropy*, Physical review E, Statistical physics, plasmas, fluids, and related interdisciplinary topics **67** (2003), no. 5.

[25] Marcio Gameiro, William Kalies, and Konstantin Mischaikow, *Topological Characterization of Spatio-Temporal Chaos*, Physical Review E 70 (2004).

[26] I. M. Gelfand and S. V. Fomin, *Calculus of Variations* (Richard A. Silverman, ed.), Dover, 1963.

[27] B. Genty and S. Meyer, *Quantitative Mapping of Leaf Photosynthesis Using Chlorophyll Fluorescence Imaging*, Australian Journal of Plant Physiology **22** (1994), no. 2, 277–284.

[28] B. Goldlücke and M. Magnor, *Weighted Minimal Hypersurfaces and Their Applications in Computer Vision*, Proceedings of the European Conference on Computer Vision, 2004, pp. 366–378.

[29] Gilad Halevi and Daphna Weinshall, *Motion of Disturbances: Detection and Tracking of Multi-Body Non-Rigid Motion*, Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition, 1997, pp. 897.

[30] R. M. Haralick, *Edge and Region Analsysis for Digital Image Data*, Computer Graphics and Image Processing **12** (1980), 60–73.

[31] _____, *The Digital Edge*, Proceedings of the Conference on Pattern Recognition and Image Processing (1981), 285–294.

[32] Lars Hörmander, *The Analysis of Linear Partial Differtial Operators I*, Springer-Verlag, 1989.

[33] Anna K. Jerebko, Sheldon B. Teerlink, Marek Franaszek, and Ronald M. Summers, *Polyp Segmentation Method for CT Colonography Computer Aided Detection*, Proceedings of SPIE, 2003.

[34] Hailin Jin, Anthony J. Yezzi, and Stefano Soatto, *Region-based Segmentation on Evolving Surfaces with Application to 3D Reconstruction of Shape and Piecewise Constant Radiance*, Proceedings of the European Conference on Computer Vision, 2004.

[35] M. Kass, A.P. Witkin, and D. Terzopoulos, *Snakes: Active Contour Models*, International Journal of Computer Vision **1** (1988), no. 4, 321–331.

[36] Ron Kimmel, *Fast Edge Integration*, Geometric Level Set Methods, 2003, pp. 59–77.

[37] _____, *Numerical Geometry of Images: Theory, Algorithms, and Applications*, Springer, 2004.

[38] Chunming Li, Chenyang Xu, Changfeng Gui, and Martin D. Fox, *Level Set Evolution Without Reinitialization: A New Variational Approach*, IEEE International Conference on Computer Vision and Patter Recognition (CVPR) (2005), 430–436.

[39] T. Lü, P. Neittaanmäki, and X-C. Tai, *A Parallel Splitting Up Method and its Application to Navier-Stokes Equations*, Applied Mathematics Letters **4** (1991), 25–29.

[40] _____, *A Parallel Splitting Up Method for Partial Differential Equations and its Application to Navier-Stokes Equations*, RAIRO Mathematical Modeling and Numerical Analysis **26** (1992), 673–708.

[41] D. Marr and E. Hildreth, *Theory of Edge Detection*, Proceedings of the Royal Society of London **B207** (1980), 187–217.

[42] Jerrold E. Marsden and Anthony J. Tromba, *Vector Calculus*, Fourth Edition, W. H. Freeman and Company, 1996.

[43] Per Mattsson and Andreas Eriksson, *Segmentation of Carotid Arteries from 3D and 4D Ultrasound Images*, Master's Thesis.

[44] A. McIvor, *Background Subtraction Techniques*, Proceedings of Image and Vision Computing, 2000.

[45] K. W. Morton and D. F. Mayers, *Numerical Solutions of Partial Differential Equations*, Cambridge University Press, 1994.

[46] D. Mumford and J. Shah, *Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems*, Communications in Pure and Applied Mathematics **42** (1989), 577–685.

[47] James R. Munkres, *Topology: A First Course*, Prentice Hall, 1975.

[48] Arian Novruzi and Michel Pierre, *Structure of Shape Derivatives*, Journal of Evolution Equations **2** (2002), 365–382.

[49] Nikos Paragios and Rachid Deriche, *Coupled Geodesic Active Regions for Image Segmentation: A Level Set Approach*, Proceedings of the European Conference on Computer Vision, 2000.

[50] David Peak, Jevin West, Susanna Messinger, and Keith Mott, *Evidence for Complex, Collective Dynamics and Emergent, Distributed Computation in Plants*, Proceedings of the National Academy of Sciences **101** (2004), no. 4.

[51] P. Perona and J. Malik, *Scale Space and Edge Detection Using Anisotropic Diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990), no. 7, 629–639.

[52] R. Courant, K.O. Friedrichs and H. Lewy, *Über die Partiellen Differenzengleichungen der Mathematischen Physik*, Mathematische Annalen **100** (1928), no. 32.

[53] Randall J. LeVeque, *Numerical Methods for Conservation Laws*, Second Edition, Birkhäuser Verlag, 1992.

[54] Kenneth A. Ross, *Elementary Analysis: The Theory of Calculus*, Undergraduate Texts in Mathematics, Springer, 1980.

[55] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D **60** (1992), 259–268.

[56] Fadil Santosa, *A Level-Set Approach for Inverse Problems Involving Obstacles*, ESAIM: Control Optimisation and Calculus of Variations **1** (1996), 17–33.

[57] Thomas Schürmann and Peter Grassberger, *Entropy Estimation of Symbol Sequences*, CHAOS **6** (1996), 414–427.

[58] Stan Sclaroff and John Isidoro, *Active Blobs*, Proceedings of the International Conference on Computer Vision, 1998, pp. 1146–1153.

[59] James A. Sethian, *Level Set Methods and Fast Marching Methods*, Second Edition (P.G. Ciarlet, A. Iserles, R. V. Kohn, and M. H. Wright, eds.), Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 1999.

[60] K. Siebke and E. Weis, *'Assimilation images' of Leaves of Glechoma Hederacea: Analysis of Non-Synchronous Stomatal Related Oscillations*, Planta **196** (1995), 155–165.

[61] S. M. Smith, *A New Class of Corner Finder*, Proceedings of the 3rd British Machine Vision Conference (1992), 139–148.

[62] D. L. Snyder, A. M. Hammond, and R. L. White, *Image Recovery from Data Acquired with a Charged-Coupled-Device Camera*, Journal of the Optical Society of America A **10** (1993), 1014–1023.

[63] J.E. Solem and N.C. Overgaard, *A Geometric Formulation of Gradient Descent for Variational Problems with Moving Surfaces*, The 5th International Conference on Scale Space and PDE Methods in Computer Vision, Scale Space 2005, Hofgeismar, Germany, 2005.

[64] _____, *A Gradient Descent Procedure for Variational Dynamic Surface Problems with Constraints*, IEEE Proceedings of the Workshop on Variational, Geometric and Level Set Methods in Computer Vision, 2005, pp. 332–343.

[65] Emily Stone, *Exploring Archetypal Dynamics of Pattern Formation in Cellular Flames*, Physica D **161** (2002), 163–186.

[66] B. Sumengen, B. S. Manjunath, and C. Kenney, *Image Segmentation Using Multi-Region Stability and Edge Strength*, IEEE International Conference on Image Processing (ICIP), 2003, pp. 429–432.

[67] Earl W. Swokowski, *Calculus*, Fifth Edition, PWS Publishing Company, 1983.

[68] Xue-Cheng Tai and Tony F. Chan, *A Survey on Multiple Level Set Methods with Applications for Identifying Piecewise Constant Functions*, International Journal of Numerical Analysis and Modeling **1** (2004), no. 1, 25–47.

[69] J. A. Thorpe, *Elementary Topics in Differential Geometry* (F. W. Gehring and P. R. Halmos, eds.), Springer-Verlag, 1979.

[70] Joachim Weickert and Gerald Kühne, *Fast Methods for Implicit Active Contour Models*, Geometric Level Set Methods, 2003, pp. 43–57.

[71] Joachim Weickert, Bart M. ter Haar Romeny, and Max A. Viergever, *Efficient and Reliable Schemes for Nonlinear Diffusion Filtering*, IEEE Transactions on Image Processing **7** (1998).

[72] Richard L. Wheeden and Antoni Zygmund, *Measure and Integral*, Marcel Dekker, Inc., 1977.

[73] Qing Wu and Yizhou Yu, *Two-Level Image Segmentation Based on Region and Edge Integration*, Proceedings of the VIIth Digital Image Computing Conference: Techniques and Applications, 2003.

[74] G. Xu, E. Segawa, and S. Tsuji, *A Robust Active Contour Model with Insensitive Parameters*, International Conference on Computer Vision, 1993, pp. 562–566.

[75] H. J. Yun and M. H. Kim, *Multiphase Segmentation of the Coronary Artery in X-Ray Angiograms*, Visualization, imaging, and image processing, 2004.

[76] H. Zhao, T. Chan, B. Merriman, and S. Osher, *A Variational Level Set Approach to Multiphase Motion*, Journal of Computational Physics **127** (1996), 179–195.