September 2005 **LIDS Publication # 2665**

Research supported in part by:

National Science Foundation Award
ACI-0121182 and the Air Force
Aerospace Research Award FA9550-04
-1-0351

Learning the Dynamics of Deformable Objects and Recursive Boundary Estimation Using Curve Evolution Techniques

Walter Sun

Learning the Dynamics of Deformable Objects and Recursive Boundary Estimation Using Curve Evolution Techniques

by

Walter Sun

B.S. Electrical Engineering,
Georgia Institute of Technology, 1995

S.M. Electrical Engineering and Computer Science,
Massachusetts Institute of Technology, 1997

E.E. Electrical Engineering and Computer Science,
Massachusetts Institute of Technology, 2004

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

September 2005

Signature of Author: ______________________________

Dept. of Electrical Engineering and Computer Science
July 18, 2005

Certified by: ______________________________

Alan S. Willsky
Edwin S. Webster Professor of Electrical Engineering
Thesis Supervisor

Accepted by: ______________________________

Arthur C. Smith
Professor of EECS
Chair, Committee for Graduate Students

Learning the Dynamics of Deformable Objects and Recursive Boundary Estimation Using Curve Evolution Techniques

by Walter Sun

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering
and Computer Science at the Massachusetts Institute of Technology

Abstract

The primary objective of this thesis is to develop robust algorithms for the incorporation of statistical information in the problem of estimating object boundaries in image data. We propose two primary algorithms, one which jointly estimates the underlying field and boundary in a static image and another which performs image segmentation across a temporal sequence. Some motivating applications come from the earth sciences and medical imaging. In particular, we examine the problems of oceanic front and sea surface temperature estimation in oceanography, soil boundary and moisture estimation in hydrology, and left ventricle boundary estimation across a cardiac cycle in medical imaging.

To accomplish joint estimation in a static image, we introduce a variational technique that incorporates the spatial statistics of the underlying field to segment the boundary and estimate the field on either side of the boundary. For image segmentation across a sequence of frames, we propose a method for learning the dynamics of a deformable boundary that uses these learned dynamics to recursively estimate the boundary in each frame over time. In the recursive estimation algorithm, we extend the traditional particle filtering approach by applying sample-based methods to a complex shape space. We find a low-dimensional representation for this shape-shape to make the learning of the dynamics tractable and then incorporate curve evolution into the state estimates to recursively estimate the boundaries.

Experimental results are obtained on cardiac magnetic resonance images, sea surface temperature data, and soil moisture maps. Although we focus on these application areas, the underlying mathematical principles posed in the thesis are general enough that they can be applied to other applications as well. We analyze the algorithms on data of differing quality, with both high and low SNR data and also full and sparse observations.

Thesis Supervisor: Alan S. Willsky
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

Throughout the duration of my Ph.D., I attained much insight from my interactions with many individuals who kindly gave of their time. In this section, I gladly acknowledge those that have made my Ph.D. experience a good one.

First and foremost, I want to thank my advisor Professor Alan Willsky for all of the supervision, advice, and recommendations he has provided me. I am grateful for the wealth of knowledge he has shared with me over the years, both in terms of research and life in academics. I also want to thank Mujdat Cetin for being a co-mentor, also providing me with advice and recommendations as well as encouragement and support throughout the entire PhD. His availability as an initial sounding board for ideas and questions was invaluable to me. Both Alan and Mujdat taught me good technical writing during the revision phase of papers, journals, and this thesis. Comparing the final version of documents I wrote near the end of my thesis compared with early drafts showed me the magnitude of their influence. In addition, I appreciate their patience during their thorough analyses of my written drafts.

I want to thank my thesis committee, which included Alan, Mujdat, and Polina Golland. Polina provided another perspective regarding my research. This allowed me to think about and investigate different questions and helped me to consider better ways to explain and present my work. I thank her for the thorough reading of my thesis and the useful recommendations for its improvement.

Besides the individuals affiliated with MIT, I had the opportunity to collaborate with several individuals outside of the MIT community, including W. Carlisle Thacker of the National Oceanic and Atmospheric Administration (NOAA), T. Mike Chin and Dmitris Menemenlis of the Jet Propulsion Laboratory (JPL), Ray Chan and Fred Holmvang of the Cardiovascular MR-CT program at Massachusetts General Hospital (MGH), and Vivek Reddy of the Electrophysiology and Pacemaker Laboratory at MGH. I thank them for their willingness to work with me.

I am also grateful to Andrew Kim and Dewey Tucker, both of whom came to MIT with me from Georgia Tech way back in the mesozoic era when dinosaurs and the NCSA Mosaic web browser ruled the earth. Both have provided encouragement over the years. Without that, I perhaps may have never returned to finish my PhD.

In addition, one of the great aspects of being part of a research group is the opportunity to interact academically and socially amongst peers. I want to thank Ayres Fan for being available to discuss issues ranging from different ways to discretize the Laplacian operator to the numerous reasons why the ACC is the best basketball conference in America (e.g. winning the ACC/Big Ten challenge six years in a row). I also want to thank Erik Sudderth, who in addition to being someone I could play tennis

with, was kind enough to share his knowledge with me through discussing his research. Similarly, I want to thank Alex Ihler for sharing his knowledge with me and letting me re-use some of his code for my work on learning the dynamics which had its origins from his Master's thesis. Thanks go out to Dmitry Malioutov for being available to help me understand aspects of his research and for reading parts of my thesis. I also want to thank John Fisher, Andy Tsai, Junmo Kim, Kush Varshney, and Ayres for the interesting discussions in our weekly grouplet. I also want to thank Ron Dror, Lei Chen, Jason Johnson, Jason Williams, Pat Kreidl, and Emily Fox for their friendship as labmates.

One of the benefits of doing research in the EECS department is the opportunity to work with others in different fields. My funding source came from a cross-department NSF project which included individuals in the earth sciences. I thank Dennis McLaughlin, Paola Rizzoli, Carl Wunsch, Sai Ravela, Xiaoyun Zang, Markus Jochum, and Yuhua Zhou for their assistance and discussions. Also, from the bi-weekly seminars that spawned from this project, I had the opportunity to meet Greg Lawson who provided me with great insight in application areas that I was not as familiar with, but ones in which he knew very well. Greg provided excellent feedback on a geoscience paper I worked on. In addition, it was nice to hear of his personal experiences towards finishing his PhD. I also want to thank Mike Perrott for being a good friend and sounding board during my return to school. It was good to hear things from the perspective of a professor with regards to the graduate school process.

I want to thank my family for making this possible. I thank my brother, Ray, and sister, Nancy, for pushing me to be my best at all times. I also thank my parents for instilling in me the value of a PhD from a young age. I always wondered how hard it was for my father to put together that large bounded thesis on the bookshelf and also how long it took for my mother to type it up for him on a typewriter (makes me appreciate computers even more given my frequent revisions).

Finally and most importantly, I want to dedicate this thesis to my wife Evie. I thank Evie for her constant encouragement and her support for my returning to school. I am also grateful for her moral support throughout my graduate experience. I would not have made it through without her.

Contents

List of Figures

List of Tables

Notational Conventions

Symbol	Definition
General Notation	
$\lvert \cdot \rvert$	absolute value
$\lvert\lvert \cdot \rvert\rvert$	2-norm, unless indicated otherwise
∇	the gradient operator
$\nabla\cdot$	the divergence operator
$\iint_R \cdot \, dA$	area integral over the region R
$\oint_{\vec{C}} \cdot \, ds$	line integral along the close curve $\vec{C}$
$\vec{\alpha}_i$	the ith component of the vector α
$A \setminus B$	the set A minus the set B
$\vec{C}_t^*$	*optimal* estimate of $\vec{C}_t$
$\mathcal{N}(\mu, \sigma^2)$	represents a Gaussian random variable with mean μ and variance σ^2
$p(f)$	probability for the random vector f
$p(f\|g)$	probability of f given g
$p(f, g)$	joint probability for the random vector f and g
$\mathbb{R}$	the field of real numbers
$\mathbb{R}^n$	the space of n-dimensional vectors
Markov Random Fields	
Z	normalization constant for the Gibb's distribution
$N(\mathbf{x})$	denotes the set of neighbors of pixel $\mathbf{x}$
Curve Evolution	
$\vec{C}$	denotes a family of smooth curves
$\vec{C}_t$	curve evolution equation or the flow for curve $\vec{C}$
$\frac{\partial \vec{C}}{\partial t}$	same as $\vec{C}_t$
ds	infinitesimal arclength
κ	principle curvature of a curve
$\vec{\mathcal{N}}$	outward unit normal of a curve
Level Set Methods	
Ψ	level set function
Ψ_t	equation of motion for the level set function Ψ

Common Acronyms

BLUE	best linear unbiased estimate
BP	belief propagation
CDF	cumulative distribution function
CLM	Community Land Model
ED	end diastole
EKG	electrocardiogram
ES	end systole
GOES	Geostationary Operational Environmental Satellite
IID	independent, identically distributed
LLSE	linear least-squares error
LV	left ventricle
MAP	maximum *a posteriori*
MC	Monte Carlo
MMS	modified Mumford-Shah
MMSE	minimum mean-squares error
MR	magnetic resonance
MRF	Markov random field
NBP	non-parametric belief propagation
NSD	normalized symmetric difference
PTI	piecewise time-invariant
PCA	principal components analysis
PDF	probability density function
SDF	signed distance function
SNR	signal-to-noise ratio
SST	sea surface temperature
TI	time-invariant

Chapter 1

Introduction

THE primary objective of this thesis is to develop robust algorithms for the incorporation of statistical information in the problem of estimating object boundaries in image data. We propose two primary algorithms, one which jointly estimates the underlying field and boundary in a static image and another which performs image segmentation across a temporal sequence. To accomplish the former, we introduce a variational technique that incorporates the spatial statistics of the underlying field to segment the boundary and estimate the field on either side of the boundary. For the latter, we propose a method for learning the dynamics of a deformable boundary that uses these learned dynamics to recursively estimate the boundary in each frame over time. Experimental results are obtained on cardiac magnetic resonance images, sea surface temperature data, and soil moisture maps. This thesis shows how the proposed methods are robust to low quality data or sparse measurements. In Section 1.1, we introduce the problem of image segmentation first performed jointly with static field estimation and then computed across a sequence of frames. In Section 1.2, we describe the main contributions of this work. We complete the chapter by outlining the organization of the thesis in Section 1.3.

■ 1.1 Image Segmentation Problem

Image segmentation involves locating feature boundaries in an image. This problem has been approached using a variety of different techniques, including adaptive binary thresholding methods [133, 152], local edge detection [21, 126] coupled with a post-processing step of grouping edges (based on Gestalt proximity laws [14]) to form a continuous boundary [181, 202], and approaches based on minimum description length criteria [93,110,112]. Alternatively, much work on segmentation is based on the so-called active contours approach [22, 30, 94, 108, 114, 142, 143]. Originated by Kass *et al.* [94], active contours are curves that evolve towards the boundary of an object according to some force function. When the segmentation is formulated as the optimization of a variational problem, this force is determined by the first variation of an associated functional whose minimization yields the boundary estimate.

Image segmentation has applications in many different scientific disciplines, including computer vision, oceanography, hydrology, medical image analysis, and im-

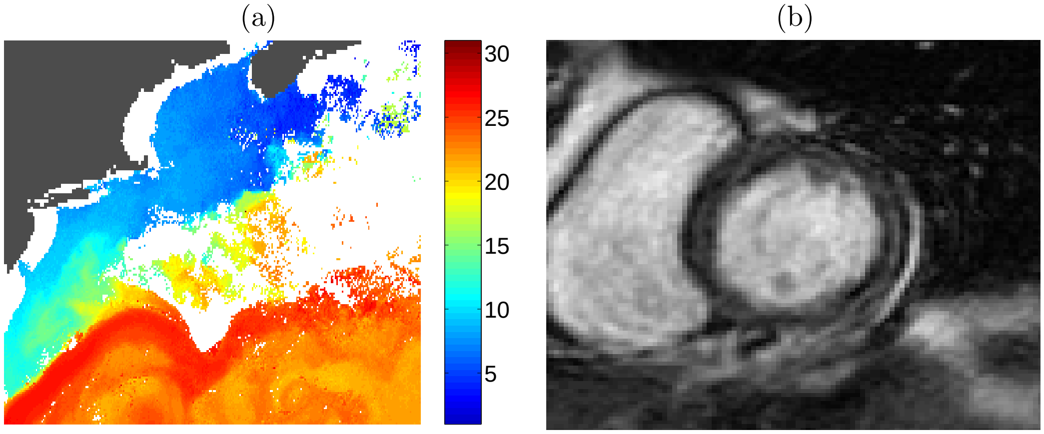


Figure 1.1. (a) Sample SST image obtained from a Geostationary Operational Environmental Satellite (GOES) [132]. The dark region in the upper left is land, while white areas are unobserved. The color bar indicates temperature in degrees Celsius. (b) Single frame of cardiac cycle (courtesy of the Cardiovascular MR-CT Program at Massachusetts General Hospital).

age and video coding. For video sequences, semantic object[1] tracking, which requires segmentation, is used for some methods of compression [75, 76, 199, 200] (e.g. it is a coding option in the MPEG-4 standard). In oceanography, localizing oceanic fronts [24,63,121,146,147,150,191] by segmentation provides value for ship navigators, ecologists, and oil rig drill teams. For illustration, Figure 1.1(a) shows a sea surface temperature map of the North Atlantic Ocean with the Gulf Stream's north wall, an example of an oceanic front, separating warm (red) waters from cooler (blue) waters. The localization of the boundary is made difficult because of missing observations and the diffusion of warm and cool waters in the eastern portion of the North Atlantic. In medical imaging, automatic segmentation of internal organs such as the cardiac left ventricle [25, 65, 71, 141, 156, 190, 201] or subcortical structures of the brain such as the hippocampus, thalamus, globus pallidus, and putamen [4, 79, 95, 153], provides assistance in medical diagnosis and surgical planning. Without prior information, the segmentation problem for medical imaging is difficult because the region of interest has non-constant intensity and other high intensity objects appear nearby such as the right ventricle in left ventricle (LV) segmentation (see Figure 1.1(b), an example of a cardiac magnetic resonance (MR) image).

■ 1.1.1 Joint Field and Boundary Estimation

Besides segmentation, there is an interest in estimating the underlying field, generally assumed smooth, from a given image. The problem of obtaining a smooth field estimate has been addressed by approaches such as kriging [41], gradient smoothing [174], and smoothing splines [184]. In some instances, there is a desire to locate the boundaries of

---

[1] Semantic objects are objects in a video that correspond to some real world object.

a region as well as estimate the field within each region. Stephenson *et al.* [168] pose a variant of kriging which simultaneously determines region boundaries and piecewise smooth field estimates by kriging within each region. Mumford and Shah [130,131] also pose a mathematical model that solves this joint problem using a variational formulation. In particular, they find the set of discontinuities and the underlying field which optimize an associated functional.

Tsai *et al.* [180], restricting the discontinuities to a family of closed curves, develop a curve evolution-based method for minimizing the Mumford-Shah functional. In that work, the curve evolution equation necessary to evolve the curve is explicitly computed. In all of these joint estimation methods, the underlying field models are spatially invariant. For instance, the Mumford-Shah approach smoothes the underlying field by enforcing a gradient penalty on the field, implicitly assuming a constant spatial mean. Such an assumption limits the class of problems that can be estimated accurately using this approach. For instance, the sea surface temperature field exhibits a spatially varying mean (one can see from Figure 1.1(a) that the temperatures generally become warmer at points farther south in the North Atlantic). Therefore, something besides a constant mean model is needed to estimate such fields accurately.

■ 1.1.2 Segmentation of Dynamically Evolving Objects

Temporally adjacent frames in a video sequence often exhibit strong correlation. Thus, incorporating information from neighboring frames can improve the segmentation process. For example, consider the problem of segmenting the LV boundary across a cardiac cycle. If we have a dynamic model for the LV as well as an estimate of the LV in adjacent frames, this information can be incorporated with the current observation to obtain a reasonable estimate of the boundary. For illustration, Figure 1.2 shows a low SNR cardiac image with estimates of the LV obtained in two different ways. Using a static segmentation (yellow curve), a suboptimal solution is found. However, by incorporating information from adjacent frames using an approach we discuss in Chapter 4, we obtain a better estimate (red curve).

Past work has incorporated prior knowledge of an object's shape to aid the segmentation process in a noisy image or one in which regions of an image are missing [35,36,101,113,178]. This prior knowledge can come from a variety of sources such as historical data or a physical model of the object. Alternatively, as illustrated, information from temporally neighboring frames can also aid the segmentation problem. In the application of LV boundary estimation, early work using temporal information for segmentation rely on taking the previous frame's estimate of the boundary and using it as a prediction for the current frame [26,86]. More recently, the predictions have been improved by incorporating knowledge of the dynamic evolution of an object [156,201]. In particular, given the previous frame's LV boundary estimate, one can dynamically evolve the boundary from a model to obtain the prediction for the boundary in the current frame. However, these papers which include evolution models for LV boundary estimation have assumed that the dynamics are linear. Such an assumption may limit

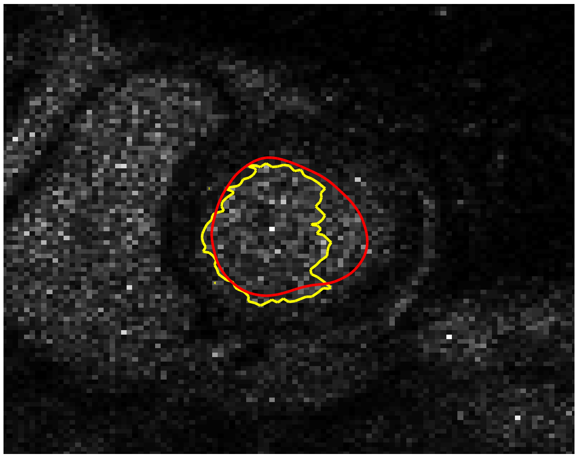


Figure 1.2. Segmentation of the left ventricle using only the image data (yellow) as compared to a segmentation that incorporates information from neighboring frames (red).

the effectiveness of the prediction, making a more complex dynamic model more desirable. In addition, the dynamic models are either assumed known [201] or determined based on the mean shape at only two points (end systole and end diastole) [156]. A more general approach for learning these dynamics is needed to accurately model deformable objects such as the LV.

■ 1.2 Contributions

■ 1.2.1 Incorporation of Spatially Varying Statistics

The first major contribution of the thesis is the incorporation of spatially varying statistics in the joint estimation of a boundary and the underlying field. This method provides improved field and boundary estimates for applications such as SST estimation in oceanography in which the underlying field exhibits systematic spatial variability. Furthermore, this formulation does not require a specific field prior. Rather, the framework is general enough that for a given application the appropriate field prior can be used based, for instance, on the statistics of the underlying field.

In addition to proposing a functional to enforce these statistics, another major contribution is the derivation of the curve evolution equation necessary to evolve the curve to the final solution. In particular, to iterate towards the solution, we apply coordinate descent, alternating between intermediate field estimates and gradient descent steps for the curve. To accomplish the latter, we take the first variation of the functional we propose (with respect to the boundary) to determine in what direction we should evolve the curve to most rapidly decrease the functional.

■ 1.2.2 Principled Recursive Boundary Estimation using a Low-Dimensional Representation

The second major contribution of the thesis is formulating a principled framework for recursively segmenting a deformable object through the use of a low-dimensional representation. Representing the evolution of deformable objects is difficult because the object itself is of high dimensionality. Using an implicit parametric state representation based on principal components analysis (PCA) [113,180], we capture the salient parts of the object's evolution in a low-dimensional state. Since this information is not sufficient to estimate the boundary of the object, we incorporate an active contours approach with the predictions based on this state representation to produce an estimate for the deformable boundary.

We pose the estimation problem in three different ways, with the specific choice of method dependent on the data available and the structure of the problem. First, we formulate the problem as one of causal filtering. Using this approach, posterior density estimates of the state are obtained based on current and past observations. Next, given the fact that non-causal processing is possible for some types of problems, we pose the problem as one of fixed-interval smoothing on a simple chain. The incorporation of the entire set of observations can lead to improvements in the posterior estimates at each frame. Finally, exploiting the fact that some applications, such as LV boundary estimation, have quasi-periodic structure, we model the system as a graphical model having a single cycle. This interpretation allows for a stronger coupling between estimates at the beginning and the end of a sequence. We apply a sample-based technique known as non-parametric belief propagation (NBP) [171] in the fixed-interval smoothing and loopy graph problems.

■ 1.2.3 Learning the Dynamics of a Deformable Object

The third major contribution of the thesis is learning the dynamics of a deformable object using an information-theoretic approach. This pure statistical approach allows one to determine the dynamics even without a physical model or prior information about the deformable object. Different types of dynamic models should be considered depending on the availability of data and the type of problem to be solved. For instance, in the LV boundary estimation problem, the limited amount of training data requires that the dynamic model be different in each phase of the cardiac cycle rather than different at every instance of time. In addition, when using NBP, both the forward and backward dynamics need to be learned.

■ 1.3 Thesis Summary

The thesis is organized as follows: an introduction, an overview of background information, three chapters on the technical aspects of the thesis, and a final chapter providing a summary and discussion of future work. In this section, we detail the contents of each chapter.

Chapter 2, Background

This chapter presents background material relevant to the development and understanding of the thesis. It begins by explaining existing field estimation algorithms used for interpolation on noisy data and observations having unobserved regions. Then, active contour methods for image segmentation are discussed. Existing active contour algorithms based on the so-called level set method are explained in detail. Next, the idea of joint field estimation and image segmentation is analyzed. The chapter then discusses the concept of shape priors for use in segmentation. Subsequently, methods for recursive estimation over time are discussed. In particular, sample-based methods such as particle filters are explained. Finally, belief propagation and the more general non-parametric belief propagation algorithm are discussed.

Chapter 3, Joint Boundary and Field Estimation in Fields with Spatially-Varying Statistics

This chapter describes both the application of a recently developed variational technique called the Mumford-Shah functional to problems in earth science and the proposal of a new method to jointly estimate the field and boundary of an image that allows for more general prior means and covariances on the field statistics.

The algorithmic framework developed is an extension of existing work that involve energy minimizations that jointly localize a boundary as well as estimate the surrounding field. We apply this technique to remote sensing data containing regions of missing observations and demonstrate how using this additional knowledge helps to locate the discontinuity in the data as well as provide a reasonable field estimate despite the sparsity and noisiness of measurements.

Chapter 4, Recursive Segmentation of Deformable Objects through Learning the Dynamics

In this chapter, we consider the concept of recursively estimating a boundary across a time sequence applied to LV boundary estimation. Using an implicit parametric representation for the boundary [113, 180], we determine a low-dimensional state for each LV boundary. This state attempts to capture the portion of the boundary which undergoes dynamic evolution.

Next, using a training set, we learn a non-parametric conditional density of the current state conditioned upon the previous state. We incorporate this density into a Bayesian framework to create a method for recursively estimating the state, which in turn allows us to obtain estimates of the LV boundary. The recursive estimation of posterior densities is approximated using sample-based methods such as the particle filter [3, 50]. We justify why our particular choice of a time-varying dynamic is reasonable for the application of left ventricle segmentation by discussing our dataset and comparing the results with an alternate choice of dynamic model.

Two different approaches for the filtering problem are examined, one using sample-based methods alone and another that uses the simplifying assumption of a Gaussian posterior density. We compare the two approaches and explain when it is appropriate to use a Gaussian assumption. In addition, the observations used for this chapter and the next include both low and high SNR data.

Chapter 5, Smoothing in Markov Chains and Single Cycle Graphs for Deformable Objects

This chapter generalizes the discussion of the previous one by introducing the concept of smoothing for the estimation of deformable objects. In smoothing, future observations are incorporated with the present and past data from the filtering problem to produce an improved estimate.

Here, we apply two different approaches to solving the problem. First, we apply a forward-backward algorithm on a sample-based approach which merely re-weights the filtered samples based on the future observations. Second, we use non-parametric belief propagation (NBP) to solve the problem. For the latter, we model the cardiac cycle as a simple Markov chain with non-parametric probability densities. We apply the message-passing algorithm of NBP to approximate the posterior densities. Curve evolution is then applied to these densities to segment the image.

We alternatively reformulate the segmentation problem to enforce the periodicity of the left ventricle. This involves a graphical model having a single cycle loop by adding an edge between the end nodes of the Markov chain. We approximate the solution using loopy NBP. Quantitative results of all methods are compared.

Chapter 6, Contributions and Suggestions for Future Research

The thesis concludes by summarizing the major contributions of the thesis. In addition, there is a discussion of future research based on the work presented in the thesis.

Chapter 2

Background

THE primary focus of this background chapter is to provide details for some specific concepts which are relevant to the development of the material in the thesis. In Section 2.1, we provide an overview of some existing field estimation algorithms. Next, we discuss existing active contour-based image segmentation methods in Section 2.2. In Section 2.3, we consider the case of jointly solving the boundary and field estimation problem. We then provide background on shape priors in Section 2.4. In Section 2.5, we discuss the use of sample-based methods to approximate probability densities, and in Section 2.6, we introduce the concept of belief propagation (BP) as a tool for estimation. Finally, in Section 2.7, we discuss a generalization to BP called non-parametric belief propagation (NBP).

■ 2.1 Field Estimation Algorithms

Many applications involve estimating an underlying field from noisy and/or incomplete observations. For example, remotely-sensed measurements from satellites may exhibit missing observations due to cloud cover, and the radiance measurements obtained are also noisy observations of the true field of interest [17,132]. Similarly, raw data acquired from magnetic resonance (MR) imaging can contain different levels of noise [164].

Methods for estimating a field from such data often involve a smoothness assumption on the field. In this section, we introduce Markov random fields (MRFs), commonly used in the computer vision community in estimation problems, as a prior for a field to be estimated. Then, we discuss algorithms for field estimation, including smoothing splines, which uses specific MRF models as a prior, and kriging. Both of these techniques are common in the earth science community.

■ 2.1.1 Bayesian Models

The field estimation problem involves the use of a Bayesian model which consists of a prior model $p(\mathbf{f})$, where $\mathbf{f}$ is a discrete random field, and a likelihood model $p(\mathbf{y}|\mathbf{f})$, where the collection of variables comprising $\mathbf{y}$ are noisy observations of the field [172]. The combination of these two models allows one to obtain a posterior model $p(\mathbf{f}|\mathbf{y})$

using Bayes' Rule. In particular,

$$p(\mathbf{f}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{y})}. \tag{2.1}$$

Once the posterior model $p(\mathbf{f}|\mathbf{y})$ is obtained, we are often interested in finding a single *optimal* field $\mathbf{f}^*$. The best choice depends on the metric chosen for cost, which is determined based on the relative importance for different types of errors in the estimate. Three common estimates are the MAP estimate, the minimum mean-squared error (MMSE) estimate, and the linear least-squares error (LLSE) estimate [193] (Appendix A provides a brief discussion of these estimators while a more detailed analysis may be found in [91]). For example, the MAP estimate of $\mathbf{f}$ can be written as[1]

$$\mathbf{f}^* = arg\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{y}) = arg\min_{\mathbf{f}}[-\log p(\mathbf{f}|\mathbf{y})] = arg\min_{\mathbf{f}}[-\log p(\mathbf{y}|\mathbf{f}) - \log p(\mathbf{f})]. \tag{2.2}$$

■ 2.1.2 Markov Random Fields

In field estimation problems, we often desire some form of spatial smoothness. This can be accomplished through the choice of a prior model which describes the correlation between neighboring points. One such model often used is the Markov random field. A Markov random field is a probability distribution defined over a field where the probability density at any point $\mathbf{x}$ in the field $\mathbf{f}$ depends only on a small set of neighboring points. In our work, we deal only with the discrete version of Markov random fields. Thus, continuous-space fields encountered are converted to discrete-space ones through sampling. Mathematically, we write

$$p(\mathbf{f}(\mathbf{x})|\mathbf{f}) = p(\mathbf{f}(\mathbf{x})|\mathbf{f}_{N(\mathbf{x})}), \tag{2.3}$$

where $\mathbf{f}_{N(\mathbf{x})}$ represents field values for neighbors of $\mathbf{x} \equiv (x_1, x_2)$, a two-dimensional position vector. Alternatively, we can represent an MRF using an undirected graphical model where the local structure can be determined from the graph. This is possible because the conditional independence of the variables is represented in the graphical model through graph separation. For example, consider the graph shown in Figure 2.1. We can conclude that, conditioned on $\mathbf{f}(X_3)$ and $\mathbf{f}(X_4)$, random variables $\mathbf{f}(X_1)$ and $\mathbf{f}(X_2)$ are independent of $\mathbf{f}(X_5)$ and $\mathbf{f}(X_6)$.

The Hammersley-Clifford Theorem [11, 15] relates MRFs to a Gibbs distribution. Mathematically, the field prior can be written (with Z a normalizing constant) as

$$p(\mathbf{f}) \equiv \frac{1}{Z} exp(-\frac{E(\mathbf{f})}{T}), \tag{2.4}$$

a Gibbs distribution with T some parameter (in thermodynamics, where the Gibbs distribution has its origins, T represents the temperature) and $E(\mathbf{f})$ represents some

---

[1] We write only the formula for the MAP estimate because this specific estimate is referenced later in the chapter.

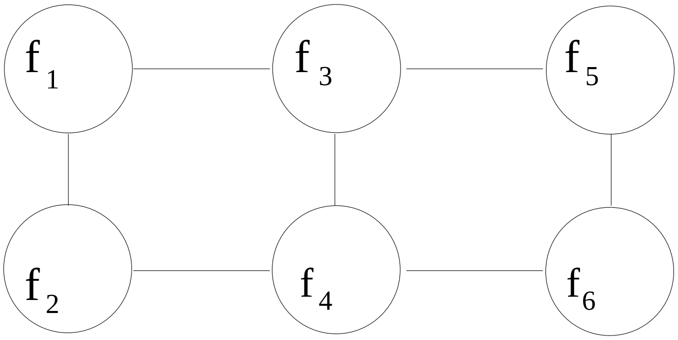


Figure 2.1. Sample undirected graph where each node represents a random variable $\mathbf{f}(i)$.

energy function that can be further decomposed into a sum of local energies

$$E(\mathbf{f}) = \sum_{\mathbf{x} \in \mathcal{N}} E_{N(\mathbf{x})}(\mathbf{f}), \tag{2.5}$$

with each $E_{N(\mathbf{x})}(\mathbf{f})$ depending only on the neighboring points $N(\mathbf{x})$ and $\mathcal{N}$ is some subset of points. This representation shows the Markovian nature of the problem.

In the next section, we discuss how MRFs can be used as a prior in the field estimation problem. In particular, we discuss the method of smoothing splines.

■ 2.1.3 Smoothing Splines

Smoothing splines [184] create smoothness in the interpolation process by minimizing derivatives of the underlying field. The main goal is to obtain a field estimate which trades off, depending on the choice of tunable parameters, field smoothness with fidelity to the potentially noisy observations. Mathematically, the functional to be minimized is

$$E_m(f) = \alpha \int_U (y(\mathbf{x}) - f(\mathbf{x}))^2 d\mathbf{x} + \int_\Omega \sum_{i=0}^{m} (\frac{\partial^m f}{\partial x_1^i \partial x_2^{m-i}})^2 d\mathbf{x}, \tag{2.6}$$

where Ω is the overall two-dimensional spatial region of interest, U is the subset of Ω where data are available, f is the unknown field to be estimated, y is the observed data, and α is a tunable parameter. The discrete form of Equation (2.6) is[2]

$$E_m(\mathbf{f}) = \alpha \sum_{\mathbf{x} \in U} (\mathbf{y}(\mathbf{x}) - \mathbf{f}(\mathbf{x}))^2 + \mathbf{f}^T L_m^T L_m \mathbf{f}, \tag{2.7}$$

where L_m is a matrix version of the differential operator. We can relate this to the posterior model of Equation (2.1) in the following manner. If the noise in the measurement model is independent, identically-distributed (IID) and Gaussian, the likelihood

---

[2] Note that $\mathbf{f}$ and $\mathbf{y}$ are the discrete counterparts to f and y, respectively.

function may be written as

$$p(\mathbf{y}|\mathbf{f}) \propto \prod_{\mathbf{x}\in U} exp(-\beta(\mathbf{y}(\mathbf{x}) - \mathbf{f}(\mathbf{x}))^2), \tag{2.8}$$

where β is proportional to the inverse of the variance. Next, we assume that the prior model is an MRF written as a Gibbs distribution (see Equation (2.4)). Exploiting the local nature of the energies (Equation (2.5)), the MAP estimate shown in Equation (2.2) can be written as

$$\mathbf{f}^* = arg \min_{\mathbf{f}} \sum_{\mathbf{x}} (\mathbf{y}(\mathbf{x}) - \mathbf{f}(\mathbf{x}))^2 + \frac{1}{T} \sum_{\mathbf{x}\in\mathcal{N}} E_{N(\mathbf{x})}(\mathbf{f}). \tag{2.9}$$

From this equation, one can observe that finding the MAP estimate of $\mathbf{f}$ is equivalent to finding the $\mathbf{f}$ which minimizes Equation (2.7). This relation holds because the differential operator acts locally on the field. In particular, as shown for the $m = 1$ case below (see Equation (2.12)), the second term of Equation (2.7), $\mathbf{f}^T L_m^T L_m \mathbf{f}$, can be written as a sum of local energies $E_{N(\mathbf{x})}(\mathbf{f})$.

In Chapter 3, we compare our proposed method of field estimation to smoothing splines of the first and second order. Because we provide results using those two methods, we focus the following discussion on these particular two choices of MRF priors.

First-Order Smoothing Splines

First-order smoothing splines are also referred to as gradient smoothing [174]. Gradient smoothing creates a field estimate which is made smooth through a gradient penalty term. Mathematically, this involves a minimization of

$$E_{m=1}(f) = \alpha \int_U (y(\mathbf{x}) - f(\mathbf{x}))^2 d\mathbf{x} + \int_\Omega |\nabla f(\mathbf{x})|^2 d\mathbf{x}, \tag{2.10}$$

where the variables are as previously defined and ∇ is the gradient operator. For convenience, we solve the discrete version of the estimation problem which involves minimizing

$$E_{m=1}(\mathbf{f}) = \alpha \sum_{\mathbf{x}\in U} (\mathbf{y}(\mathbf{x}) - \mathbf{f}(\mathbf{x}))^2 + \mathbf{f}^T L_1^T L_1 \mathbf{f}, \tag{2.11}$$

with $L_1^T L_1 = L_{x_1}^T L_{x_1} + L_{x_2}^T L_{x_2}$, where L_{x_1} is a matrix of first differences in the x_1 direction and L_{x_2} is a matrix of first differences in the x_2 direction. If we write the Gibbs energy as $E(\mathbf{f}) = \mathbf{f}^T L_1^T L_1 \mathbf{f}$, it is easy to show that this can be decomposed into a sum of local energies, where at each point,

$$E_{N(x_1,x_2)}(\mathbf{f}(x_1,x_2)) \equiv (4\mathbf{f}(x_1,x_2) - [\mathbf{f}(x_1-1,x_2) + \mathbf{f}(x_1+1,x_2) + \mathbf{f}(x_1,x_2-1) + \mathbf{f}(x_1,x_2+1)])^2. \tag{2.12}$$

From an estimation-theoretic standpoint, the preference for smooth fields can be viewed as a specific prior on the first and second order statistics of the field. For instance, the gradient terms impose a particular covariance (the inverse of $L_1^T L_1$) and a spatially-constant mean. Gradient smoothing is also equivalent to finding the MAP estimate using the so-called thin membrane prior model [12, 37] with IID and Gaussian noise.

Second-Order Smoothing Splines

Second-order smoothing splines [184] create more smoothness than gradient smoothing by minimizing the second order order derivatives of the field. Mathematically, we minimize

$$E_{m=2}(f) = \alpha \int_U (y(\mathbf{x}) - f(\mathbf{x}))^2 d\mathbf{x} + \int_\Omega [(\frac{\partial^2 f}{\partial x_1^2})^2 + (\frac{\partial^2 f}{\partial x_2^2})^2 + 2(\frac{\partial^2 f}{\partial x_1 \partial x_2})^2] d\mathbf{x}. \qquad (2.13)$$

In discrete form, we have

$$E_{m=2}(\mathbf{f}) = \alpha \sum_{\mathbf{x}\in U} (\mathbf{y}(\mathbf{x}) - \mathbf{f}(\mathbf{x}))^2 + \mathbf{f}^T L_2^T L_2 \mathbf{f}, \qquad (2.14)$$

where L_2 is a matrix determined by second differences. Second-order smoothing splines are equivalent to finding the MAP estimate using a thin plate prior model.

■ 2.1.4 Kriging

Kriging is a method of field estimation commonly used to deal with missing measurements in earth science applications, such as in oceanographic and hydrological studies [6, 9, 119, 175]. Also referred to as optimal analysis [54, 62], kriging attempts to provide a smooth estimate of a field given some observations by obtaining an estimate which minimizes the mean-squared estimation error. As with smoothing splines, the estimate at each point is directly dependent only on a local neighborhood of points, the size being a tunable parameter. Appendix B briefly describes the process of kriging [41], mainly to provide terminology to a reader unfamiliar with this specific technique.

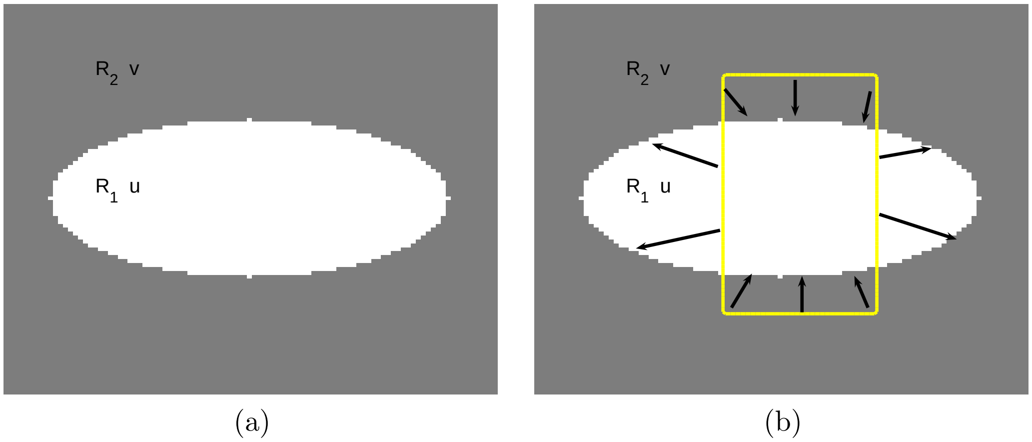
■ 2.2 Active Contour Methods for Image Segmentation

In this section, we discuss the topic of active contour methods for image segmentation. First, we discuss the concept of active contours, or curve evolution. Then, we examine how these methods are applied to image segmentation problems.

■ 2.2.1 Active Contours

Active contours [22, 30, 94, 108, 114, 142, 143] are curves which evolve to minimize an associated objective functional that incorporates constraints from available data (e.g. imagery). The goal of this technique is to obtain a segmentation (i.e. locate a boundary) in the spatial field or image of interest. Mathematically, this amounts to determining

$$\hat{C} = \underset{\vec{C}}{argmin}[E(\vec{C})], \qquad (2.15)$$



(a) (b)

Figure 2.2. (a) Given an image with two regions R_1 with mean u and R_2 with mean v, we want a curve which can separate them. (b) An initialization of the active contour evolution process shows the curve $\vec{C}$ and arrows indicating its evolution towards the desired final solution.

where $\hat{C}$ represents the segmentation and E is the energy functional to be minimized. If we introduce an iteration time parameter t, we may evolve our curve according to a differential equation of the form

$$\frac{\partial \vec{C}}{\partial t} = -\mathcal{F}(\vec{C}), \tag{2.16}$$

where $\mathcal{F}(\vec{C})$ is a force functional. Choosing $\mathcal{F}(\vec{C})$ as the first variation of $E(\vec{C})$ allows the curve to move in the direction of steepest descent. The curve is evolved until steady-state is reached (i.e. $\mathcal{F}(\vec{C}) = 0$).

The evolution of such a contour can be performed in a variety of ways. One standard approach involves discretizing the curve into a set of marker points that are moved from iteration to iteration [32, 94]. At each iteration, a curve can be constructed by some method of interpolation. Kass [94] parameterizes an initial curve and moves individual marker points to evolve the curve. The evolving curves in such approaches are commonly referred to as "snakes." This technique has several drawbacks [158, 177]. First, nothing in the equations of motion prevent the discretized points from moving together, leaving an unevenly discretized curve. Second, stable evolution requires a very small time-step. Third, marker points cannot easily handle changes in topology of the evolving curve. Some modifications can help to remedy these problems, such as more frequent resampling of the points on the curve to redistribute them evenly. However, the added complexity leads many, including us, to consider alternate methods. One such technique, which we use for our curve evolution, involves level set methods. Level set methods, developed by Osher and Sethian [137, 159], provide a robust way of evolving a contour because they handle topological changes seamlessly and do not require labels or marker points. Instead of evolving discretized points on the curve, level set methods evolve a surface whose zero level set represents the curve.

Level Sets

Level sets provide a robust way of evolving the contour towards the solution through the process of representing an n-dimensional boundary in an $(n+1)$-dimensional space as an $(n+1)$-dimensional surface in an $(n+2)$-dimensional space. Devised by Osher and Sethian [137,159], this Eulerian technique does not have the same problems inherent in Lagrangian methods[3], as it handles topology changes seamlessly and does not require labels or marker points. For visualization purposes, we explain the level set method for the case of a one-dimensional curve in a two-dimensional space. The technique generalizes to arbitrary n-dimensional boundaries in an $(n+1)$-dimensional space. We begin with an initial (one-dimensional) curve $\vec{C}$. This curve is an initial guess for the desired location of the contour. Then, we create a two-dimensional surface (embedded in $\Re^3$) whose zero level set is the curve. Since an infinite number of surfaces satisfy this condition, one often makes the choice that the initial level set is the signed distance function, where each point (x_1, x_2) in the original two-dimensional space maps to the point (x_1, x_2, x_3) on the surface, where $|x_3|$ is the Euclidean distance to the curve, with $x_3 > 0$ on one side of the boundary, $x_3 < 0$ on the other, and $x_3 = 0$ on the boundary itself. When the boundary is a closed curve, we choose as a matter of convention $x_3 < 0$ inside the boundary. The signed distance function has the special characteristic that it satisfies $|\nabla\Psi| = 1$ (where Ψ is the surface), known as the Eikonal equation [169].

Figure 2.3 shows an example of an elliptical boundary evolving into two objects, a circle and a rectangle. The evolution is performed using level sets, and is shown in (a-c). Shown in (d-f) is the shape formed at the zero level set of the surface. Note how the topology change from (e) to (f) occurs naturally with level sets in the evolution from frame (b) to frame (c).

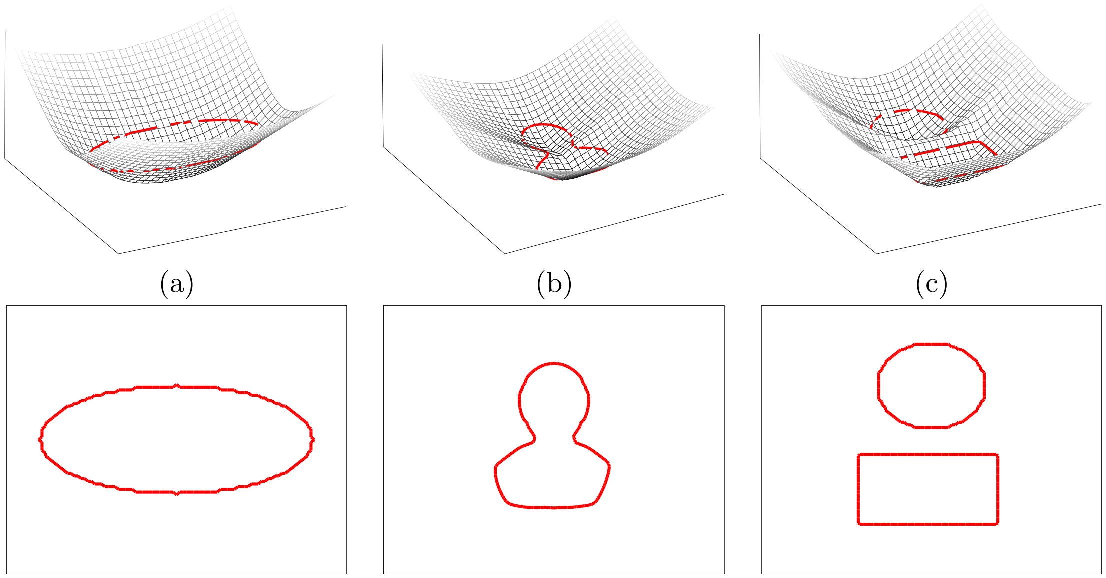
2.2.2 Gradient-Based Methods

Different methods exist for active contour segmentation, with most of the early work focused on edge-based models [22,23,94,99]. Kass *et al.* [94] are generally given credit as the first to use active contours for segmentation. Caselles *et al.* [22] and Malladi *et al.* [122,123] introduce the use of level sets to solve the segmentation problem, creating a stopping function that reaches a minimum at points of maximal gradient. In further work, Caselles *et al.* [23] relates energy-based active contours to the computation of geodesics, minimal distance curves for a given metric. They derive a new flow equation and prove the existence, uniqueness, and stability of the solution using the theory of viscosity [31,39]. Meanwhile, Kichenassamy *et al.* [99] introduce a similar stopping term in their energy functional which encourages curves to settle in a so-called *potential well* occuring in high gradient areas. They further extend the gradient-based segmentations to three dimensional problems.

All of these edge-based methods rely primarily on local information; in particular,

---

[3] Eulerian methods rely on a fixed coordinate grid, whereas Lagrangian methods involve a parameterized set of points that represent the front or boundary of interest, which move over time.



(a) (b) (c)

Figure 2.3. Visualization of the level set method. Frames (a-c) show the two-dimensional surface evolving in three-dimensional space, with the zero level set drawn. Frames (d-f) show the corresponding evolution as seen in the original two-dimensional space for the one-dimensional boundary.

a strong image gradient. In some instances, region statistics provide useful information for segmentation. This is the subject of our next section.

■ 2.2.3 Region-Based Methods

Region-based information is particularly useful when no sharp image discontinuity exists at the object's boundary. In such problems, gradient methods may fail. For example, the boundary is not well-defined by gradients in images where there is a smooth transition from one region to the other. Alternatively, the data may be so noisy that it is difficult to determine which points of strong gradient are indeed part of the boundary. In both instances, region-based methods may be more appropriate than edge-based approaches for segmentation of such images.

While early active contour methods focused on gradient terms to locate boundaries, Ronfard [151], Mumford and Shah [130], and Zhu and Yuille [203] assert that edge-based terms alone are not always sufficient in finding the boundary. As a result, they separately propose region-based techniques which consider the statistics of the entire region on either side of the curve to aid in the boundary localization problem. More recently, region-based methods have been extended by Chan and Vese [27, 28], Tsai *et al.* [177, 179], Paragios and Deriche [142, 143], Kim *et al.* [103], and Chakraborty *et al.* [25], with Kim's work involving higher order statistics of the regions via non-parametric density estimates. The following subsections provide more detail on some of the different region-based image segmentation techniques introduced above. In particular, we discuss separation of means and variances [177], the Chan-Vese functional [29], and the Mumford-Shah functional [130].

Building upon the development of edge and region-based methods, Chakraborty *et al.* [25] introduce a method which combines these two techniques (along with a shape prior, which we will discuss in Section 2.4) to produce an improved algoritm for segmentation. Citing that each method alone has drawbacks, they describe a probabilistic framework which depends on a shape term, an image gradient term, and a region homogeneity term. Similarly, Paragios [141] uses gradient and region techniques jointly for segmentation.

Separation of Means

We consider a technique which uses region statistics for image segmentation. In particular, we seek the boundary which maximizes the difference in mean intensities of the two regions separated by the boundary. Proposed by Tsai [177], the objective can be written as

$$E_{mean}(\vec{C}) = -\frac{1}{2}(u-v)^2 + \gamma \oint_{\vec{C}} ds, \tag{2.17}$$

where u and v are the mean intensity values on either side of the boundary separated by $\vec{C}$, and the second term is a regularization term which encourages a smooth curve. Without this curve length term, the minimized objective can result in a jagged curve or boundaries almost everywhere, something not meaningful or useful in an image segmentation. This objective functional works well when the two regions have clearly different mean values.

Chan-Vese Functional

Chan and Vese [27] use a similar functional involving the mean statistics on either side of the curve. Using the assumption that an image of interest is formed by two regions of approximately piecewise-constant intensities, they minimize

$$E_{Chan/Vese}(\vec{C}) = (\int_{R_1} |I(x_1,x_2) - u|^2 dA + \int_{R_2} |I(x_1,x_2) - v|^2 dA) + \gamma \oint_{\vec{C}} ds, \tag{2.18}$$

where $I(x_1, x_2)$ is the intensity of the image at (x_1, x_2), and u and v are the mean values of the observations in R_1 and R_2, the two regions separated and defined by the boundary. Again, the other term of the functional is a regularization term which enforces smoothness through a curve length penalty.

Separation of Variances

When the mean values of two regions are the same, but the variances are different, consider an objective of the form

$$E_{var}(\vec{C}) = -\frac{1}{2}(\sigma_u^2 - \sigma_v^2)^2 + \lambda \oint_{\vec{C}} ds, \tag{2.19}$$

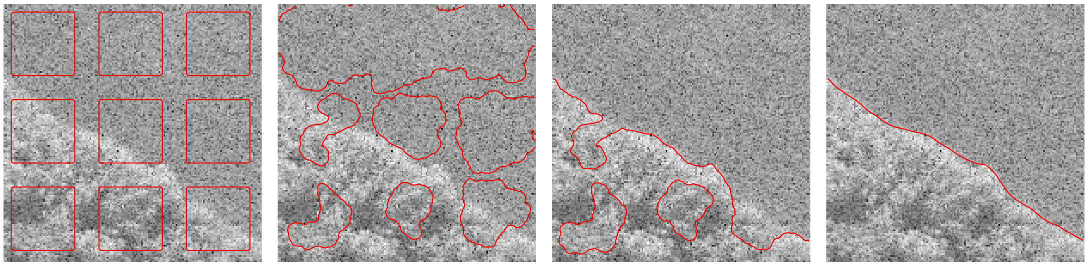


Figure 2.4. The tree line shown in the SAR image is captured by using gradient flow to separate two regions having the same mean but different variances. The four images show the evolution of the curve, from initialization (left) to the final segmentation (right). Taken from [177].

where σ_u^2 and σ_v^2 are the variances of the two regions, respectively. Separation of means and the Chan-Vese functional described above will not separate the regions in this case because the means are the same. However, using this energy functional, we can separate these regions having the same mean by different variances. Figure 2.4 shows four steps of the contour evolution achieved using $E_{var}(\vec{C})$ as the objective to segment trees from a grassy region in a noisy synthetic aperture radar (SAR) image, where the two regions have similar mean intensities.

■ 2.3 Joint Segmentation and Field Estimation Algorithms

In Section 2.1 we discussed methods of field estimation, while in Section 2.2 we examined techniques of estimating boundaries. In some instances, we are interested in solving both problems simultaneously. For example, several problems require localizing a boundary (e.g. an oceanic front, the boundary between different soils, or the left ventricle) while obtaining a smooth field estimate on either side of such a boundary (e.g. sea surface temperature, soil moisture, or blood intensity in a magnetic resonance image). In this section, we examine methods which jointly segment a boundary and estimate a field.

■ 2.3.1 Thin Plate and Thin Membrane with Line Processes

In Section 2.1, Markov random fields are used as a prior for the estimation of smooth fields. In certain applications, however, discontinuities in the field may be expected. To address this issue, Geman and Geman [66] discuss the addition of a so-called *line process* to allow for discontinuities in the field.

Blake and Zisserman [12] refer to prior models which allow for line processes as weak plate and weak membrane models (as opposed to thin plate and thin membrane for fields which do not have a discontinuity). They examine the sensitivity of the estimation to different types of discontinuities and provide an algorithm for numerically solving a discretized version of the joint estimation problem. They formulate an energy functional to be minimized across field values and locations of discontinuities. The discrete energy functional to be minimized over f and l (written in 1-D) is

$$E(\mathbf{f},l) = \alpha \sum_x (\mathbf{f}(x) - \mathbf{y}(x))^2 + \sum_x [(1 - l(x))(\mathbf{f}(x) - \mathbf{f}(x-1))^2 + \gamma l(x)], \tag{2.20}$$

where the line process function l is a binary function which takes a value of 1 when there is a line or discontinuity between x and $x-1$, and 0 when there is not. In the second term, $(\mathbf{f}(x) - \mathbf{f}(x-1))^2$ represents a gradient penalty on the field (other smoothness penalties can be used). This formulation, known as a weak membrane model, is also equivalent to the functional proposed by Mumford and Shah [130,131]. We now discuss an active contour approach that minimizes this functional to find the optimal field and boundary.

2.3.2 Active Contour Implementation of the Mumford-Shah Functional

In early applications of Mumford-Shah, or equivalently the weak membrane model, finding the solution involved the arduous process of searching for the optimal set of arbitrary discontinuities in the data. Tsai *et al.* [180], restricting the set of discontinuities to a family of closed curves, propose a curve evolution-based technique that minimizes the Mumford-Shah functional to produce a boundary and field estimate. Their method results in a smooth closed curve which separates two discontinuous regions and also provides smooth estimates of the field within each region.

The Mumford-Shah functional (as re-formulated by Tsai *et al.*) is a three-term energy functional written in continuous space as

$$E_{MS}(f,\vec{C}) \equiv \alpha \int_U (y(\mathbf{x}) - f(\mathbf{x}))^2 d\mathbf{x} + \beta \int_{\Omega \backslash \vec{C}} |\nabla f(\mathbf{x})|^2 d\mathbf{x} + \gamma \oint_{\vec{C}} ds, \tag{2.21}$$

where the variables in common with Equation (2.6) are defined in the same way, $\vec{C}$ is the unknown boundary to be estimated, and α, β, and γ are tunable parameters. The first term in the functional, the likelihood term in the Bayesian formulation shown in Equation (2.1), is also known as the data fidelity term. This term penalizes deviations of the estimated field from the observed data. The second term penalizes the gradient of the field everywhere except across the closed curve, thus enforcing smooth fields on either side of the boundary while not penalizing a potentially sharp discontinuity across the boundary. Finally, the third term is the regularization term. This regularization captures the physical characteristic of producing smooth, non-jagged estimates of the boundary. The minimization of the Mumford-Shah functional using active contours as described in Section 2.2.1 produces a piecewise smooth field estimate.

2.3.3 Kriging with Multiple Regions

Similar to Geman and Geman's extension of thin plate and thin membrane models to fields with discontinuities, Stephenson *et al.* [168] have proposed a variant of kriging that handles discontinuous spatial fields. Citing the poor performance of kriging when the field is non-stationary and the need for some method to handle fields with discontinuities, they use the Bayesian Partition Model of Denison *et al.* [48] to define a set of disjoint regions obtained using a Voronoi tessellation. Each region is assumed to have stationary statistics, but discontinuities are allowed across the region boundaries. The approach involves first an initialization of the number of different regions in the field with their corresponding approximate centers. The technique iteratively refines the locations of the centers and the corresponding estimates of the field in each region until convergence. One limitation of this method is that regions are constrained to be polygons. In most instances, the shape of field discontinuities is not so constrained. However, this method allows kriging to be performed in fields where discontinuities are known to exist.

■ 2.4 Shape Priors

The statistical theory of shape has been described by the works of Kendall [96] and Small [166] regarding the use of marker points to represent a shape. Shape priors have been used to aid in the segmentation of an object [25,36,101,113,180]. Cootes *et al.* [36] first describe the use of principal components analysis (PCA) to compute the average shape and the primary modes of variability of a training set assuming a Gaussian model for each of the feature points. Their marker points were located manually. They use this representation as a means to deform training shapes to match the segmentation of new data. In later work by the same group, Davies *et al.* [44, 45] propose the use of a minimum description length criterion to automatically locate the landmarks.

Staib and Duncan [167] create statistical priors on Fourier coefficients to represent the shapes. Their work builds on the early work of Scott [154] who models shapes with trigonometric functions. Alternatively, Tagare [173] proposes an approach for segmentation where image boundaries are matched to two-dimensional shape templates. This work extends Yuille's [198] early work for using deformable templates to extract features from faces. These analyses work well whenever well-defined landmarks can be used for the marker points. Stated another way, good correspondence is useful if not necessary when using a marker point approach.

As described in Section 2.2.1, Osher and Sethian [137] introduced level sets that could be used to represent shapes without the need for landmark or marker points. Using this method, several authors [30, 40, 113, 178] have analyzed and represented shapes in the segmentation problem. Chen *et al.* [30] create an energy functional which contains a gradient-based term and a shape prior term. The former incorporates the measured data, while the latter encourages the segmentation to be similar to existing shapes from the training database. Their shape prior seeks curves which are closely aligned to training shapes, allowing for scaling, translation, and rotation. Cremers *et al.* [40] incorporate shape information jointly with the Mumford-Shah functional, which we described in Section 2.3.2. Tsai *et al.* [178], leveraging on work by Leventon *et al.* [113], minimize an energy functional within the space of shapes obtained using PCA on level sets.

Other recent work in shape analysis include Fletcher *et al.*'s [57] principal geodesic analysis of shapes, an analogy of PCA in Euclidean spaces for shape parameters residing in Lie groups [52]. The use of Lie groups applies naturally for their medial representation of geometric objects. Klassen *et al.* [105] represent shapes using direction functions and determine the geodesic path connecting any two shapes. They use this path to compute the distance between the two shapes, thus providing a metric for shapes. Kim [101] estimates shape densities from a training set and incorporates shape prior densities into image segmentation. In particular, he proposes non-parametric density estimation techniques on the space of signed distance functions, a non-linear manifold. He shows the robustness of the algorithm by producing a shape density on a multi-modal set of training shapes (a handwritten digit data set). He then computes the MAP estimate of the estimated densities by curve evolution.

■ 2.5 Sample-Based Methods for Recursive Inference

Complex probability densities often cannot be expressed by a simple set of parameters. As a result, non-parametric kernel or Parzen [144] densities are often used to represent the distributions. When there exists a dynamic system and a recursive equation relating the evolution of a state having non-parametric density, sample-based methods, such as particle filters [3, 50, 51, 106], are often used to recursively approximate the density. These techniques represent the density through a discrete set of weighted samples drawn from that density (which in turn approximates the density through a kernel density estimate). In this section, we describe particle filters and the related method of particle smoothing.

■ 2.5.1 Particle Filtering

Particle filtering [3, 50] is a sequential Monte Carlo (MC) approach where a set of particles is used to approximate a probability distribution. Known also as the condensation algorithm [82], bootstrap filtering [68], and interacting particle approximations [42], particle filtering is particularly useful when the distributions are non-Gaussian.[4]

Suppose we have a state X_t which we wish to estimate. However, we cannot observe it directly. Instead, we have measurements y_t at each time that represent a noisy observation of the true state X_t. If we assume that X_t is a Markov process, then a diagram of this hidden Markov model is shown in Figure 2.5. Now, the objective is to determine the posterior distribution $p(X_t|y_{1:t})$, where $y_{1:t} \equiv [y_1, y_2, \ldots, y_t]$, the set of all observations from the initial time to time t. Applying Bayes' Rule, we can write

$$p(X_t|y_{1:t}) = p(X_t|y_t, y_{1:t-1}) = \frac{p(X_t, y_t|y_{1:t-1})}{p(y_t)} \propto$$

---

[4] For Gaussian problems, techniques such as Kalman filtering [92] allow us to solve the problem exactly without need for approximation.

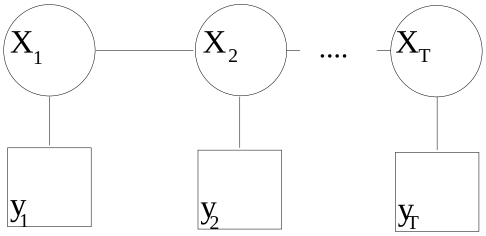


Figure 2.5. Graphical model representation of a T node Markov chain, with curve $\vec{C}_t$ and observation nodes y_t explicitly shown.

$$p(y_t|X_t, y_{1:t-1}) \int_{X_{t-1}} p(X_t|X_{t-1}, y_{1:t-1}) p(X_{t-1}|y_{1:t-1}) dX_{t-1}, \tag{2.22}$$

which simplifies to

$$p(X_t|y_{1:t}) \propto p(y_t|X_t) \int_{X_{t-1}} p(X_t|X_{t-1}) p(X_{t-1}|y_{1:t-1}) dX_{t-1} \tag{2.23}$$

if we assume that X is a Markov process and that observations y_t are independent from each other conditioned on their associated state X_t. From this equation, we can observe the recursive nature of our problem. We initialize the problem with a set of M samples $x_1^{(i)}$ having weights $w_1^{(i)}$ to represent the initial posterior $p(X_1|y_1)$. In particular, the continuous distribution at any time t is approximated by

$$p(X_t|y_{1:t}) \approx \sum_{i=1}^{M} w_t^{(i)} k(x - x_t^{(i)}), \tag{2.24}$$

where $k(\cdot)$ usually represents a Gaussian kernel but could be any other kernel.

Starting at time $t-1$, we have samples $x_{t-1}^{(i)}$ having weights $w_{t-1}^{(i)}$ to approximate $p(X_{t-1}|y_{1:t-1})$. The points propagate through the dynamics to sample points $x_t^{(i)}$ having weights $w_{t|t-1}^{(i)}$. These points are then weighted by the likelihood, namely

$$w_t^{(i)} = \frac{1}{Z} p(y_t|x_t^{(i)}) w_{t|t-1}^{(i)}, \tag{2.25}$$

where Z is a normalizing constant which ensures that the weights sum to 1.

Importance Sampling

One of the key concepts of particle filters is the idea of importance sampling. If a given distribution $p(x)$ cannot be sampled, we resort to sampling from a proposal density $\pi(x)$ and apply the weight of $\frac{p(x)}{\pi(x)}$ to the samples. The weighting for each sample $x_t^{(i)}$

is necessary to adjust for the fact that we sample from $\pi(x)$ rather than $p(x)$, the distribution of interest. Mathematically,

$$w_t^{(i)} \propto \frac{p(x_t^{(i)})}{\pi(x_t^{(i)})}, \tag{2.26}$$

where in order to maintain a valid probability distribution we normalize the weights such that

$$\sum_i w_t^{(i)} = 1. \tag{2.27}$$

Implicitly, $p(x)$ needs to be a density which can be evaluated, but such a requirement is much more likely than the likelihood that we are able to sample $p(x)$.

Resampling

Another feature of particle filters is resampling. After a few iterations in the recursive filtering process, most of the particle weight may become concentrated among a small set of samples, leading to a degenerate distribution. The process of resampling can mitigate this problem.

Re-sampling involves generating a new set of samples $x_t^{(j)}$ from the approximate representation $p(x) \approx \sum_{i=1}^{M} w_t^{(i)} k(x; x_t^{(i)})$, where $k(x; x^0)$ represents some kernel function centered at x^0. Re-sampling can be accomplished in many different ways. A simple approach is to create the cumulative distribution function (CDF) associated with the approximation for $p(x)$. Then, to obtain each new sample, draw a random number from the uniform distribution of real numbers between 0 and 1 and determine the associated sample by the sample value in the CDF which corresponds to this random number.

■ 2.5.2 Particle Smoothing

Sometimes the estimation process can be performed offline, meaning that an estimate of X_t can be based not only on past and present observations y_t, but also future ones. This process is called smoothing. Work has been done to generalize the process of particle filtering to incorporate future observations [58, 67]. This method is known as particle smoothing.

For this particular analysis, we look at fixed-interval smoothing[5]. Fixed-interval smoothing [192] involves estimating the state X_t for all t in a given interval $[0, T]$ conditioned on observations from the entire interval $y_{1:T} \equiv [y_1, y_2, \dots, y_{t-1}, y_t, y_{t+1}, \dots, y_T]$. Fong *et al.* [58] propose an approach which post-processes the particle filter as described in Section 2.5.1 by re-weighting the existing sample points based on future observations.

---

[5] The other two types of smoothing which are commonly considered are fixed-point smoothing, which estimates $p(X_t|y_{1:\tau})$ with fixed $\tau > t$, and fixed-lag smoothing, which estimates $p(X_t|y_{1:t+\tau})$ for fixed $\tau > 0$ [192].

The posterior based on all observations $p(X_t|y_{1:T})$ can be written as

$$p(X_t|y_{1:T}) = \int p(X_{t+1}|y_{1:T}) \frac{p(X_t|y_{1:t})p(X_{t+1}|X_t)}{p(X_{t+1}|y_{1:t})} dX_{t+1}. \tag{2.28}$$

In particular, suppose the particle filtering process yields at time t a set of points $x_t^{(i)}$ with associated weights $w_t^{(i)}$. Once the filtering process reaches the endpoint at time T, the smoothing process involves recursively reweighting the samples going backwards. So, assume that we have samples $x_{t+1}^{(j)}$ with updated weights $w_{t+1|T}^{(j)}$, then we can reweight the samples of $x_t^{(i)}$ according to Equation (2.28) by

$$w_{t|T}^{(i)} \propto \sum_j w_{t+1|T}^{(j)} w_t^{(i)} p(x_{t+1}^{(j)}|x_t^{(i)}). \tag{2.29}$$

Once all of the weights are computed, we normalize them so that they sum to 1.

Because this process does not resample the points, we need to have a dense enough set of samples to represent the density at each time. Otherwise, the reweighting process will not accurately capture changes in the distribution. One alternate method to solving the smoothing problem uses belief propagation. We discuss this concept next.

■ 2.6 Belief Propagation

Belief propagation (BP) [145] is an inference algorithm for graphical models [87,88,111] that computes the posterior $p(X_t|y_{1:T})$ through a series of message passing operations. The field of graphical models deals with inference on distributions defined on graphs. Such distributions satisfy conditional independence relations with respect to the graph. A graph contains a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$, each of which connects a distinct[6] pair of nodes. A random variable X_t is associated with each node $t \in \mathcal{V}$, while the edges between the nodes represent a relationship between the random variables.

Using the same framework for state estimation as that described in Section 2.5, we can create a graphical model representation for this Markov process (the graph is as shown in Figure 2.5). In such an undirected graph, $p(X_t|X_1, X_2, \cdots, X_{t-1}) = p(X_t|X_{t-1})$ and $p(X_t|X_{t+1}, \cdots, X_{T-1}, X_T) = p(X_t|X_{t+1})$. To provide a concrete example, X_1 and X_3 are independent conditioned on X_2. Also, each y_t is independent of the other observations y_τ (for $\tau \neq t$) conditioned on X_t. As in the smoothing problem, the aim is to obtain the posterior; that is, we estimate the unobserved states X_t through all observations $y_{1:T}$. When the state variables are jointly Gaussian, a fixed-interval Kalman smoother solves the problem exactly. Belief propagation can solve a generalization of this problem for any tree-structured graph (i.e. no cycles). For any graph

---

[6] All graphs we consider will have no self-loops (a node having an edge to itself) and no multiple edges between two different nodes.

without cycles, the prior $p(X_t)$ can be written as

$$p(X) = \prod_{(a,b)\in\mathcal{E}} \frac{p(X_a, X_b)}{p(X_a)p(X_b)} \prod_{c\in\mathcal{V}} p(X_c), \tag{2.30}$$

where (a, b) are edges connecting nodes a and b, and X represents all nodes in the system (such a factorization usually does not exist in a graph with cycles). If we define pair-wise potential functions $\psi_{a,b}(X_a, X_b) = \frac{p(X_a,X_b)}{p(X_a)p(X_b)}$ and self-potentials for prior models $\psi_c(X_c) = p(X_c)$, Equation (2.30) can be rewritten as

$$p(X) = \prod_{(a,b)\in\mathcal{E}} \psi_{a,b}(X_a, X_b) \prod_{c\in\mathcal{V}} \psi_c(X_c), \tag{2.31}$$

a product of self-potentials defined at each node and pair-wise potentials defined between pairs of nodes.

Rewriting the posterior distribution $p(X_t|y_{1:T})$ using Bayes' Rule, we have

$$p(X_t|y_{1:T}) \propto p(X_t)p(y_t|X_t) \prod_{a\in N(t)} p(Y_{a\backslash t}|X_t), \tag{2.32}$$

where $N(t) \equiv \{a|(a,t) \in \mathcal{E}\}$ denotes the neighborhood of a node t and $Y_{a\backslash t}$ represents the set of all observation nodes in the tree with root at node a except for nodes in the subtree having a root at node t. Similarly, we can apply Bayes' Rule and a marginalization identity on $p(Y_{a\backslash t}|X_t)$ to obtain [170]

$$p(Y_{a\backslash t}|X_t) = \frac{p(X_t|Y_{a\backslash t})p(Y_{a\backslash t})}{p(X_t)} \propto \int_{X_a} \frac{p(X_t, X_a|Y_{a\backslash t})}{p(X_t)} dX_a \propto \int_{X_a} \frac{p(X_t, X_a)p(Y_{a\backslash t}|X_a)}{p(X_t)} dX_a$$

$$= \int_{X_a} \frac{p(X_t, X_a)}{p(X_a)p(X_t)} p(X_a)p(y_a|X_a) \prod_{b\in N(a)\backslash t} p(Y_{b\backslash a}|X_a) dX_a. \tag{2.33}$$

Now, combining Equations (2.32) and (2.33), we obtain

$$p(X_t|y_{1:T}) = \alpha p(y_t|X_t) \prod_{a\in N(t)} m_{at}(X_t), \tag{2.34}$$

where α is a constant parameter and

$$m_{at}(X_t) = \int_{X_a} \psi_{t,a}(X_t, X_a)\psi_a(X_a, y_a) \prod_{b\in N(a)\backslash t} m_{ba}(X_a) dX_a, \tag{2.35}$$

with $\psi_{a,b}(X_a, X_b) \equiv \frac{p(X_a,X_b)}{p(X_a)p(X_b)}$ being the pair-wise potential between nodes a and b as defined before and $\psi_a(X_a, y_a) \equiv p(y_a|X_a)$ being a self-potential which includes the observation y_a. So, given some algorithm to compute the *messages* m_{at} in Equation (2.35), the posteriors can be determined from Equation (2.34). The process of iteratively computing the messages shown in Equation (2.35) is known as belief propagation. This is accomplished by first initializing each message $m^0_{ab}(X_b)$ to some arbitrary value and then iteratively applying the message update equation

$$m^n_{ab}(X_b) = \int_{X_a} \psi_{b,a}(X_b, X_a) p(y_a|X_a) \prod_{c \in N(a) \backslash b} m^{n-1}_{ca}(X_a) dX_a \tag{2.36}$$

until convergence. Convergence, in graphs without loops, is guaranteed after M steps, where M is the *diameter* of the graph [170]. The diameter of a graph is defined to be the longest of the shortest paths between any pair of nodes in the graph. In graphs with loops, convergence is not guaranteed. In addition, when convergence is achieved, the results, in general, differ from the true probabilities.

■ 2.7 Non-parametric Belief Propagation

The derivations presented so far hold for arbitrary potentials. For discrete state problems, the transmitted messages are a vector of numbers. Furthermore, for continuous variables having Gaussian densities, the represention of messages is exact because only the means and covariances need to be transmitted. Most other continuous densities, however, cannot be transmitted easily. In particular, the computation of Equation (2.36) is often not tractable [84].

In Section 2.5.1, we discuss how particle filters can be used to approximately solve a recursive estimation problem for Markov chains. However, such an approach only holds for a simple chain. Sudderth *et al.* [171] combine techniques of BP and particle filtering to create a technique called non-parametric belief propagation (NBP). NBP is an inference algorithm for graphical models [87, 88, 111] which is tractable for continuous non-Gaussian densities. Previously, inference on such problems was performed by first discretizing the potentials. However, when the dimensionality is very large, the discretization and subsequent inference is computationally intractable. NBP addresses this issue by representing the messages in the message-passing process as non-parametric kernel densities.

The key novelties of NBP are two-fold. First, the message-update algorithms of BP are adapted so that non-parametric representations of the messages are created using kernel density estimates [165]. Second, the message products, a product of Gaussian mixtures, are calculated using an efficient local Gibbs sampling [66] procedure. This process involves importance sampling as described in particle filtering.

Chapter 3

Joint Boundary and Field Estimation in Fields with Spatially-Varying Statistics

IN this chapter, we propose a joint field and boundary estimation method which allows for spatially-varying statistics. We focus the algorithm on earth science applications because existing interpolation methods applied to this area do not handle fields containing discontinuities. In particular, techniques (e.g. Mumford-Shah [130,131,180]) which jointly estimate the field and the location of a discontinuity within the field have not been applied to such problems. We first propose the use of Mumford-Shah to more accurately estimate sea surface temperature and soil moisture given noisy and/or partial observations. In addition, we propose a modified formulation which allows for spatially-varying means and a general covariance. This new method has a number of desirable features, especially when the spatial statistics vary in some systematic way, as in sea surface temperature measurements where it is known that a longitudinal temperature gradient exists (i.e. temperatures are warmer near the equator than near the poles).

In Section 3.1, we motivate the need for estimation in the areas of oceanography and hydrology. In Section 3.2, we describe how we apply the method of Mumford-Shah to sea surface temperature and soil moisture measurements. In Section 3.3, we propose a generalization of Mumford-Shah which we call the modified Mumford-Shah (MMS) functional. In Section 3.4, we provide an extensive set of results comparing MMS and Mumford-Shah to other interpolation techniques. We summarize the work explained in this chapter in Section 3.5.

■ 3.1 Description of Motivating Problems

One application area for our work is oceanography. In this research area, ocean circulation [60] and, more specifically, oceanic fronts [24,63,121,146,147,150,191] have been analyzed and modeled extensively. An oceanic front, such as the Gulf Stream's north wall which separates regions of warm and cold waters in the North Atlantic, provides a strong current which results in a temperature discontinuity perpendicular to the front.

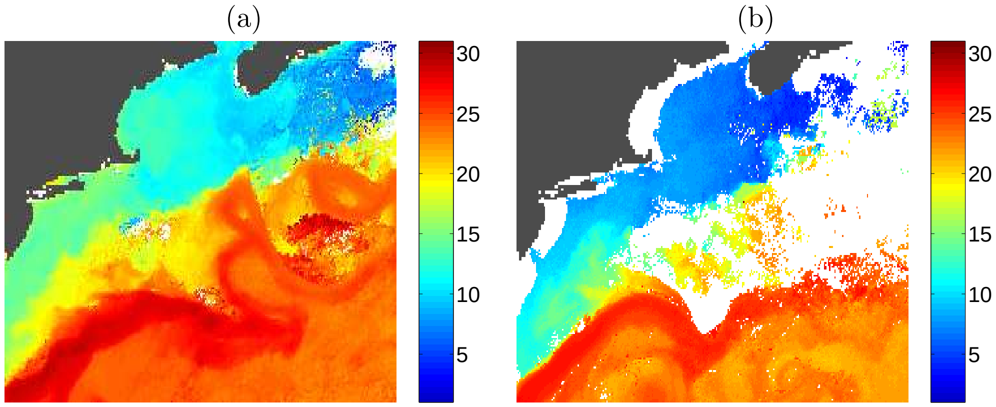


Figure 3.1. Sample SST images obtained from a Geostationary Operational Environmental Satellite (GOES). The dark region in the upper left on both images is land mass, while white regions are areas of missing measurements. Colorbar on right indicates temperature in degrees Celsius. (a) Image with 2% data missing due to cloud cover. (b) Image on another day with 34% data missing.

Determining the location of these ocean surface features not only helps the field estimation process, but is of practical significance in itself in marine-based transportation, ecological studies, and oil exploration. Estimating the field on both sides of these fronts is useful to gain a better understanding of events such as coral reef bleaching [69] and global climate changes [19]. Sea surface temperature (SST) is often used to observe fronts and ocean circulation. SST can be measured by ship- or aircraft-based hydrographic surveys, or by satellite-based infrared (or microwave) sensors. While the former only offer sparse and irregular sampling patterns concentrated mostly along commercial ship lanes, the satellite measurements, which have maximum spatial resolution of approximately five kilometers, are significantly attenuated by atmospheric water vapor and droplets, resulting in large data voids. These regions of missing observations are often common over oceanic regions covered by clouds over periods as long as weeks, thus making the localization of oceanic fronts a non-trivial problem [191]. For illustration, Figure 3.1(a) shows an SST map for a region adjacent to the Atlantic seaboard showing the Gulf Stream's north wall when 2% of the data points in the ocean are unobserved. Figure 3.1(b) shows another image with more cloud cover, resulting in 34% data missing. Given such data, the problem is to estimate the location of the north wall as well as the entire underlying temperature field, interpolating over regions of missing observations.

Another application area for our techniques is hydrology. Work has been done in this area to investigate methods of estimating soil moisture fields [124], a problem in which measurements may be incomplete (e.g. due to a sparse set of ground sensors). Variations in soil moisture can significantly impact the climate and land use [5, 56, 61, 109, 194], hence soil moisture estimation is of interest. Different soil textures tend to have different levels of water saturation [20, 55, 129]. As a result, moisture fields tend to exhibit clear discontinuities at the boundary between soil types.

■ 3.2 Application of Mumford-Shah to Earth Science Problems

Detecting oceanic fronts (boundaries between two distinct water masses) in SST maps and the boundaries between different soil types in soil moisture measurements are examples of remote sensing problems involve estimating a field which contains discontinuities. Traditional methods of interpolation and smoothing are often used to produce estimates of these fields. Examples of existing techniques include kriging and smoothing splines, as described in Section 2.1, as well as optimal interpolation [16, 43, 54, 62, 89, 125, 176] and smoothing by local regression [34]. Because these interpolation techniques do not account for the discontinuities, they blur across the boundaries when estimating the field [41, 168]. To address this problem, we propose a solution based on an approach that jointly locates an unknown discontinuity and produces a smooth field estimate on either side of the discontinuity.

In particular, we apply the Mumford-Shah model [130, 131] to applications within oceanography and hydrology. Introduced in Section 2.3.2 in its continuous form and repeated here for convenience, the Mumford-Shah functional,

$$E_{MS}(f,\vec{C}) \equiv \alpha \int_U (y(\mathbf{x}) - f(\mathbf{x}))^2 d\mathbf{x} + \beta \int_{\Omega\backslash\vec{C}} |\nabla f(\mathbf{x})|^2 d\mathbf{x} + \gamma \oint_{\vec{C}} ds, \tag{3.1}$$

can be written in discrete form as

$$E_{MS}(\begin{bmatrix} \mathbf{f}_{R_1} \\ \mathbf{f}_{R_2} \end{bmatrix}, \vec{C}) = \alpha[||H_{R_1}\mathbf{f}_{R_1} - \mathbf{y}_{R_1}||_2^2 + ||H_{R_2}\mathbf{f}_{R_2} - \mathbf{y}_{R_2}||_2^2] + \beta[\mathbf{f}_{R_1}^T L^T L \mathbf{f}_{R_1} + \mathbf{f}_{R_2}^T L^T L \mathbf{f}_{R_2}]$$
$$+\gamma l(\vec{C}) \tag{3.2}$$

where the matrix L is the discrete representation of the gradient operator, R_1 and R_2 are the regions on either side of the boundary, $\mathbf{y}_{R_1}$ and $\mathbf{y}_{R_2}$ are SST or soil moisture observations lexicographically unwrapped and written in vector form, $\mathbf{f}_{R_1}$ and $\mathbf{f}_{R_2}$ are similarly unwrapped vectors representing the underlying field estimates of temperature or moisture, and H_{R_i} associates $\mathbf{f}_{R_i}$ with the associated observation $\mathbf{y}_{R_i}$ (i.e. if the m^{th} observation in $\mathbf{y}$ corresponds to a measurement of the n^{th} element of $\mathbf{f}_{R_i}$, then $H_{R_i}(m,n) = 1$, and it is only these entries of H_{R_i} that are nonzero). Finally, $l(\vec{C})$ represents the length of curve $\vec{C}$, the Gulf Stream's north wall or the boundary between soil types.

In this framework, we formulate the problem of joint boundary localization and field estimation as an optimization problem. The $||H\mathbf{f} - \mathbf{y}||_2^2$ terms in the functional impose data fidelity in the field estimate. The gradient terms ($\mathbf{f}^T L^T L \mathbf{f}$) indicate a specific preference for field smoothness on either side of the boundary, while the curve length term encourages a smooth boundary.

3.3 Alternative Model for Spatial Regularization

In some applications, the model assumptions implied by Mumford-Shah are inaccurate. In such instances, a more accurate, yet still simple, prior model may be available. For instance, sea surface temperatures have spatially-varying means. More specifically, we expect temperatures north (south) of the Gulf Stream's north wall to decrease (increase) as a function of the distance from this discontinuity. Furthermore, an information matrix different from the $L^T L$ of Equation (3.2) may provide a better characterization of the second order statistics within the regions. In this section, we propose a generalization of the Mumford-Shah functional which incorporates general first and second order statistics of the field on either side of the discontinuity.

3.3.1 The Modified Mumford-Shah (MMS) Functional

We first introduce the MMS functional in continuous form:

$$\begin{aligned} E_{MMS}(f,\vec{C}) = \alpha \int_{U \cap R_1} (f(\mathbf{x}) - y(\mathbf{x}))^2 d\mathbf{x} + \alpha \int_{U \cap R_2} (f(\mathbf{x}) - y(\mathbf{x}))^2 d\mathbf{x} \\ + \beta \int_{R_1}\int_{R_1} \mathcal{D}[f(\mathbf{x}) - \mu_{R_1}(\mathbf{x})] b_{R_1}(\mathbf{x},\mathbf{z}) \mathcal{D}[f(\mathbf{z}) - \mu_{R_1}(\mathbf{z})] d\mathbf{x} d\mathbf{z} \\ + \beta \int_{R_2}\int_{R_2} \mathcal{D}[f(\mathbf{x}) - \mu_{R_2}(\mathbf{x})] b_{R_2}(\mathbf{x},\mathbf{z}) \mathcal{D}[f(\mathbf{z}) - \mu_{R_2}(\mathbf{z})] d\mathbf{x}\, d\mathbf{z} + \gamma \oint_{\vec{C}} ds. \end{aligned} \tag{3.3}$$

In this equation, U is the region where data are available, $\mathbf{x} \equiv (x_1, x_2)$ and z each represent two-dimensional position vectors, s denotes the arc length, f is the unknown field to be estimated, $b_{R_i}(\mathbf{x},\mathbf{z})$ provides a relationship between two points $\mathbf{x}$ and $\mathbf{z}$ in R_i (thus permitting a non-local dependence that Mumford-Shah does not allow[1]), $\vec{C}$ is the unknown boundary to be estimated, g is the observed data, α, β, and γ are positive constants, $\mathcal{D}$ is an arbitrary linear differential operator, and $\mu_{R_1}(\mathbf{x})$ and $\mu_{R_2}(\mathbf{x})$ are spatially-varying mean functions in the respective regions. In particular, we model the means as functions of the distance from the boundary. So, $\mu_{R_i}(\mathbf{x}) = \phi_i(d(\mathbf{x},\vec{C}))$, where $d(\mathbf{x},\vec{C})$ is the distance of point $\mathbf{x}$ to $\vec{C}$, and ϕ_i is some function. The data fidelity and curve length terms of the MMS functional are the same as those in Mumford-Shah. However, we replace the second term in Equation (3.1) with terms which incorporate the prior means μ_{R_1} and μ_{R_2} and an arbitrary linear differential operator $\mathcal{D}$ on the field f. Note that this reduces to the Mumford-Shah functional in the special case where the means are zero, the $\mathcal{D}$ operator is the gradient operator, and b is the Dirac delta function. Various possibilities for the priors exist (e.g. historical data can be used to estimate the mean and covariance) [33, 64, 78, 149, 186]. In our examples, we make specific choices for these parameters to illustrate the characteristics and advantages of this approach. We now describe how Equation (3.3) is minimized numerically using coordinate descent.

---

[1] Note that b must be a positive semi-definite function.

■ 3.3.2 Numerical Solution

Ideally, we wish to solve for f and $\vec{C}$ simultaneously. However, given that the gradient of $E_{MMS}(f, \vec{C})$ is not easily attainable, we approach the solution to this problem using the technique of coordinate descent. This method is an iterative approach in which at each step, all but one of the varying parameters is held fixed while the remaining parameter is updated in a manner which decreases the functional. Within each iteration, each of the parameters is updated exactly once.

Using this technique, we divide each iteration into two main steps. First, having $\vec{C}$ fixed, we compute the value of f which minimizes $E_{MMS}(f, \vec{C})$ for the particular value of $\vec{C}$.[2] Then, with f fixed, we evolve $\vec{C}$ in the direction of the first variation of $E_{MMS}(f, \vec{C})$ with respect to $\vec{C}$. For this curve evolution step, we ideally want both b and μ to vary with the curve, since both can in general be functions of $\vec{C}$. However, for computational simplicity, we hold b and μ fixed as we evolve $\vec{C}$ and then recompute b and μ to accurately reflect the new position of $\vec{C}$ before the next iteration. The iterative process is repeated until we converge to an equilibrium. The method of coordinate descent does not guarantee convergence to the global minimum, but given a reasonable choice of initialization, our algorithm appears to converge to a sufficiently good solution as seen from the experiments in Section 3.4.

In the following, we derive the closed form solution of f given a fixed $\vec{C}$ (correspondingly, R_1 and R_2 are fixed) in a discrete space. First, we need to write the discrete space representation of the MMS functional. We can write the third term in Equation (3.3) as $(D(\mathbf{f}_{R_1} - \mathbf{m}_{R_1}))^T B_{R_1}(D(\mathbf{f}_{R_1} - \mathbf{m}_{R_1}))$, where D is a matrix approximation of the $\mathcal{D}$ operator, B_{R_1} is a matrix which represents the function b_{R_1}, and $\mathbf{f}_{R_1}$ and $\mathbf{m}_{R_1}$ are vectors representing the discretized version of f_{R_1} and μ_{R_1}, respectively. Simplifying this term, we have $(\mathbf{f}_{R_1} - \mathbf{m}_{R_1})^T D^T B_{R_1} D(\mathbf{f}_{R_1} - \mathbf{m}_{R_1})$. Similarly, the fourth term in Equation (3.3) involving R_2 becomes $(\mathbf{f}_{R_2} - \mathbf{m}_{R_2})^T D^T B_{R_2} D(\mathbf{f}_{R_2} - \mathbf{m}_{R_2})$. We can think of the term $D^T B_{R_i} D$ as an inverse covariance matrix ($K_{R_i}^{-1} \equiv D^T B_{R_i} D$); so, the discretized version of MMS can be written, with $\mathbf{f} = \begin{bmatrix} \mathbf{f}_{R_1} \\ \mathbf{f}_{R_2} \end{bmatrix}$, as

$$E(\begin{bmatrix} \mathbf{f}_{R_1} \\ \mathbf{f}_{R_2} \end{bmatrix}, \vec{C}) = \alpha[||H_{R_1}\mathbf{f}_{R_1} - \mathbf{y}_{R_1}||_2^2 + ||H_{R_2}\mathbf{f}_{R_2} - \mathbf{y}_{R_2}||_2^2] +$$

$$\beta[(\mathbf{f}_{R_1} - \mathbf{m}_{R_1})^T K_{R_1}^{-1}(\mathbf{f}_{R_1} - \mathbf{m}_{R_1}) + (\mathbf{f}_{R_2} - \mathbf{m}_{R_2})^T K_{R_2}^{-1}(\mathbf{f}_{R_2} - \mathbf{m}_{R_2})] + \gamma l(\vec{C}). \quad (3.4)$$

where as in Equation (3.2) $\mathbf{y}_{R_1}$ and $\mathbf{y}_{R_2}$ are observations lexicographically unwrapped and written in vector form, $\mathbf{f}_{R_1}$ and $\mathbf{f}_{R_2}$ are similarly unwrapped vectors, and H_{R_i} associates $\mathbf{f}_{R_i}$ with the associated observation $\mathbf{y}_{R_i}$ (i.e. if the m^{th} observation in $\mathbf{y}$ corresponds to a measurement of the n^{th} element of $\mathbf{f}_{R_i}$, then $H_{R_i}(m, n) = 1$, and it is only these entries of H_{R_i} that are nonzero). Finally, $l(\vec{C})$ represents the length of curve $\vec{C}$.

---

[2] One could take a gradient step in the direction which decreases the functional rather than find the minimum and still attain a solution [177], but we actually solve for the minimizing f in each step.

By differentiating $\mathbf{f}_{\mathbf{R_1}}{}^*$ which minimizes Equation (3.4) is

$$\mathbf{f}^*_{R_1} = K_{R_1} H^T_{R_1} (H_{R_1} K_{R_1} H^T_{R_1} + \frac{\beta}{\alpha} I)^{-1} (\mathbf{y}_{R_1} - H_{R_1} \mathbf{m}_{R_1}) + \mathbf{m}_{R_1}, \tag{3.5}$$

where I is the identity matrix whose dimension is equal to the number of observations in the region. The result for $\mathbf{f}_{R_2}$ is similar to Equation (3.5) with R_2 replacing every instance of R_1.

After solving for and updating $\mathbf{f}$ for a given curve $\vec{C}$, the next step is to evolve $\vec{C}$ with $\mathbf{f}$ fixed. This is accomplished by taking the first variation of $E(\mathbf{f}, \vec{C})$ with respect to $\vec{C}$. The result provides us the direction of evolution of $\vec{C}$ for which $E(\mathbf{f}, \vec{C})$ decreases most rapidly. For the sake of brevity, we state the main result, namely the active contour expression here, and refer readers to Appendix C for the derivation of the result. The outcome of our derivation is

$$\begin{aligned} \frac{\partial \vec{C}}{\partial t} &= \alpha[([H_{R_2} \mathbf{f}_{R_2}] - \mathbf{y}_{R_2})^2 - ([H_{R_1} \mathbf{f}_{R_1}] - \mathbf{y}_{R_1})^2]\vec{N} \\ &+2\beta([diag[D(\mathbf{f}_{R_2} - \mathbf{m}_{R_2})] B_{R_2} (D(\mathbf{f}_{R_2} - \mathbf{m}_{R_2}))] \\ &-[diag[D(\mathbf{f}_{R_1} - \mathbf{m}_{R_1})] B_{R_1} (D(\mathbf{f}_{R_1} - \mathbf{m}_{R_1}))])\vec{N} - \gamma\kappa\vec{N}, \end{aligned} \tag{3.6}$$

where t represents iteration time during the evolution of the active contour, κ is the curvature of $\vec{C}$, $diag[\cdot]$ is a diagonal matrix whose diagonal elements are given by the elements of the vector inside the square brackets, and $\vec{N}$ is the outward unit normal with respect to R_1. The numerical evolution is performed iteratively, alternating between updates of $\mathbf{f}$ and $\vec{C}$, until convergence.

■ 3.4 Experimental Results

We perform experiments on SST data and soil moisture maps. We demonstrate the advantage of using our simultaneous boundary detection and field estimation method by comparing results from Mumford-Shah and MMS to existing techniques of kriging, gradient smoothing, and smoothing splines. The methods of kriging, gradient smoothing, and smoothing splines fundamentally have no provision to handle the presence of discontinuities. The illustrations from our results show how the failure to handle discontinuities leads to blurring across the boundaries. Before presenting the results, we note that the tunable parameters α, β, and γ in Equations (3.1) and (3.3) have been subjectively chosen to provide reasonable results. Similarly, the choices of parameters for the methods we compare against, namely kriging, gradient smoothing, and smoothing splines, have been made with an attempt to provide as good results as possible.

■ 3.4.1 Simulated Soil Moisure Data

We examine soil moisture measurements generated from a simulation based on the Community Land Model (CLM), a model designed at the National Center for Atmospheric

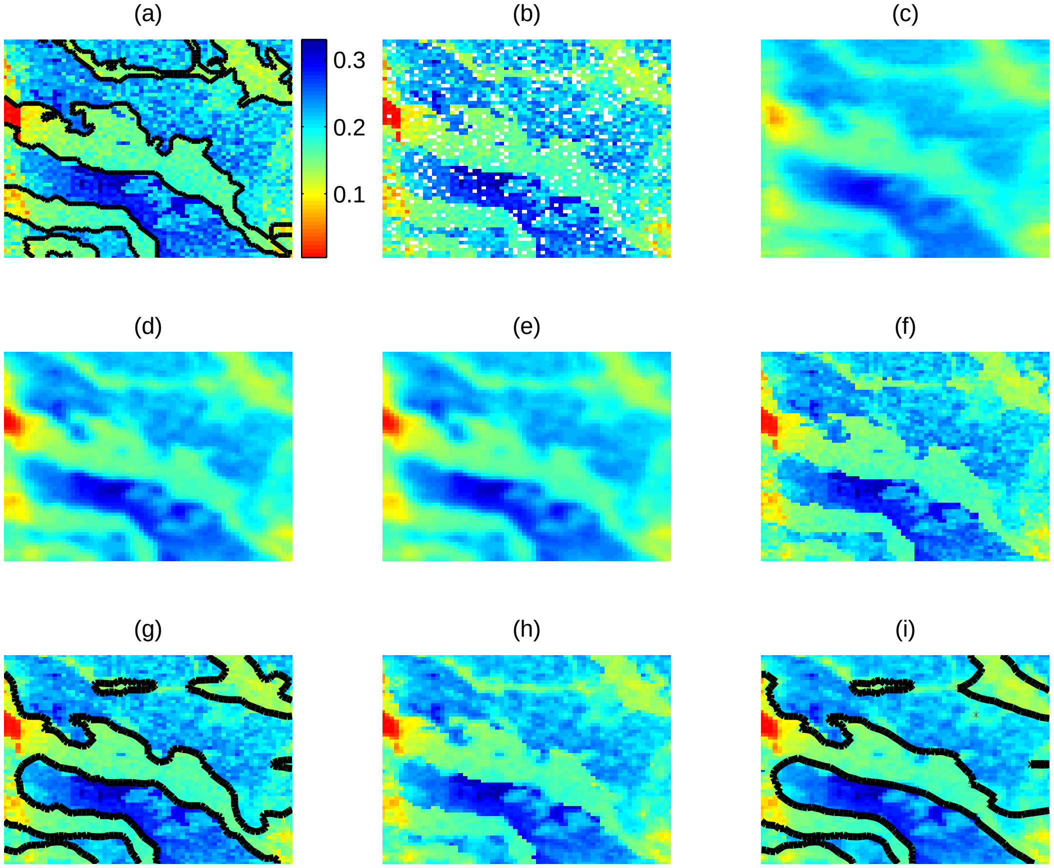


Figure 3.2. Estimation of soil moisture in an example where 90% of the data is observed. Images show fractional volume of water in the soil. (a) The original soil moisture map with colorbar (colorbar valid for all images) and true boundary overlaid. (b) The observed data. (c) Estimate using kriging. (d) Estimate using gradient smoothing. (e) Estimate using smoothing splines with second-order derivatives. (f) Field estimate using Mumford-Shah. (g) Field and boundary estimates using Mumford-Shah. (h) Field estimate using MMS. (i) Field and boundary estimates using MMS.

Research (NCAR) [13]. The CLM produces variable soil moisture in a spatial domain having different soil types. The particular simulation we consider assumes a uniform spatial distribution of rainfall. After time elapses from a period of rainfall, the surface soil moisture of the sand region tends to be much drier than any of the other soil types present. As a result, we pose the problem of segmenting the boundary (at a given time after a recent rainfall) between two regions, the first containing sand and the second containing other soil types, as well as estimating the soil moisture in each of these regions. Figure 3.2(a) shows an example of a soil moisture map with the boundary (given from the simulation) separating the sand region from the other soil types.

For the MMS method, we compute the mean moisture for sand and that for the other region from simulated data and use these as values for μ_{R_1} and μ_{R_2} in Equation (3.3), respectively. The results are sensitive to the choice of μ, so the simulated data and the test data need to have similar mean moistures. For this application, the means are chosen to be spatially non-varying because at this scale, no systematic variability can be discerned from the mean field. We choose $\mathcal{D}$ to be the second derivative operator and b to be the delta function. The choice of a second derivative operator for $\mathcal{D}$ means we enforce an even greater amount of smoothness than that obtained in Mumford-Shah, which uses a first derivative operator.

To provide perspective for how well the estimation methods which explicitly consider boundaries perform, we compare the estimation of the field using Mumford-Shah and MMS with a few standard methods of interpolation and smoothing. In particular, we consider kriging [41, 46, 107][3], gradient smoothing [174], and a second order smoothing spline [184]. In essence, gradient smoothing, as defined in Equation (2.10), involves applying Mumford-Shah without any notion of a boundary. Similarly, the second order smoothing spline is shown in Equation (2.13). Proposed by Wahba and Wendelberger [184], this method is identical to MMS in its choice of $\mathcal{D}$ and b, except that it does not account for the possible presence of boundaries.

For the field estimates, the particular simulation we use is at a much finer spatial scale than the field information we want to capture. Hence, the simulation essentially provides a *noisy version* of the coarser scale soil moisture field we are interested in estimating. Thus, to evaluate field estimation accuracy, we compare the resulting estimates with a coarser scale version of the simulations obtained using gradient smoothing within each of the regions (assuming that the boundary is known).

We consider three examples having different coverages of observed data. For the first example, we consider the case where 90% of the region is observed (Figure 3.2(b)). Figure 3.2(c) shows the interpolation resulting from kriging, Figure 3.2(d) illustrates the estimate from gradient smoothing, and Figure 3.2(e) shows the result from second order smoothing splines. In these three cases, some of the measurement noise, apparent from Figure 3.2(a), is removed, but the boundaries are slightly blurred.

Figure 3.2(f) shows the field estimate using Mumford-Shah while Figure 3.2(g) shows the same estimate with the boundary overlaid (for reference). Similarly, Figure 3.2(h) shows the field estimate using MMS while Figure 3.2(i) shows the estimate with the boundary overlaid. A binary threshold was used to determine the initial boundary for the active contour evolution. For reference, Figure 3.3 shows the initialization determined by the binary thresholding. Quantitatively, Table 3.1 shows the standard error per pixel for the variational techniques as well as the traditional methods. From these results, it can be seen that the field estimation errors from the variational techniques we propose are smaller than those from the traditional methods. To compare the accuracy of the boundary localization, we first need to find a way to generate a boundary from the

---

[3] The kriging results were obtained using ordinary kriging with nugget effect using a 5×5 window of neighbors for prediction. Other methods of kriging with different windows sizes can be considered. See Gratton and Lafleur's MATLAB Kriging Toolbox [72] for different options.

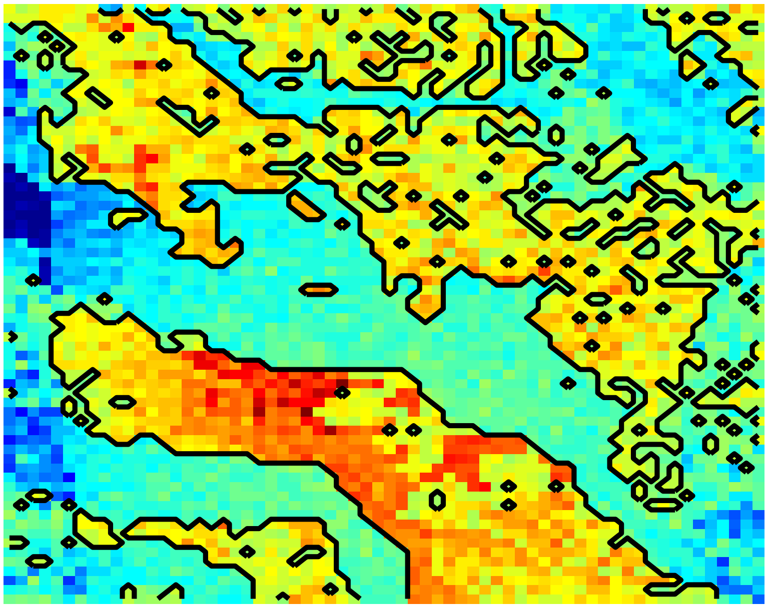


Figure 3.3. Initialization for curve evolution based on binary thresholding at a moisture level of 20% water fraction.

	Kriging	Gradient Smoothing	Smoothing Splines	Mumford-Shah Functional	MMS Functional
90% observed	0.0299	0.0231	0.0240	0.0183	0.0186
81% observed	0.0328	0.0269	0.0274	0.0224	0.0256

Table 3.1. Standard error per pixel between the field estimates and a smoothed version of the simluated observation. Note that the variational methods we propose have smaller errors.

field estimates of the traditional approaches. A direct approach involves using adaptive binary thresholding, where for each instance, we find the threshold which most closely matches the ground truth according to the normalized symmetric difference (NSD). NSD is defined as

$$NSD(R_1, R_2) \equiv \frac{A(R_1 \cap \bar{R}_2) + A(\bar{R}_1 \cap R_2)}{A(R_1 \cup R_2)}, \tag{3.7}$$

where $\bar{R}_i$ represents the complement of region R_i and $A(R_i)$ is the area of region R_i. NSD is 0 for a perfect match and 1 for a complete mismatch. Alternatively, we also compute the dice coefficient [49], which is defined as

$$Dice(R_1, R_2) \equiv \frac{2A(R_1 \cap R_2)}{A(R_1) + A(R_2)}, \tag{3.8}$$

where $R_i \cap R_j$ represents the intersection of regions R_i and R_j and $A(R_i)$ is the area of region R_i. The dice measure evaluates to 1 when two regions are a perfect match and 0 when the regions are a complete mismatch.

After boundary localization through such manual adaptive processing, we quantify the closeness of the resulting boundary to the truth by computing the NSD and dice coefficient. Note that for the traditional methods, the thresholding process is in practice unrealizable since we do not necessarily have ground truth. Thus, the boundaries from the traditional approaches are actually the best case results that can come from these methods. For the boundaries shown in Figure 3.4, the NSD for the boundary determined using kriging is 0.3469, that for gradient smoothing is 0.4158, and that for second-order smoothing splines is 0.4105. In contrast, the NSD for Mumford-Shah is 0.1502 while that for MMS is 0.1789. Hence, our proposed methods do better in finding boundaries than the traditional approaches. Note that for the soil moisture examples, we do not expect MMS to necessarily outperform Mumford-Shah because the fields are assumed to have spatially constant means (unlike the SST example described in the next section).

These results illustrate that Mumford-Shah and MMS can locate the boundary and produce field estimates that maintain a sharp transition at the region boundaries. In addition, we note that through the use of level set methods [137, 159] for our curve evolution, we are able to detect boundaries which are not simple closed curves, as shown in Figures 3.2(f) and (g).

For the second example, we consider the situation where a large rectangular region is unobserved, as shown in Figure 3.5(b) (compare with the full observation in Figure 3.5(a)). This may simulate a scenario where blocks of measurements are locally obtained by regional stations, but because of a gap in the spatial coverage of the stations, certain areas are unobserved. Again, Figures 3.5(c), (d), and (e) show interpolations using kriging, gradient smoothing, and second order smoothing splines, respectively. In all three images, we observe a smooth estimate in the missing rectangular region as well as a smoothing of the observed data. Furthermore, as expected, the discontinuity at the soil boundary is not well-defined in the missing region for any of these three field estimates.

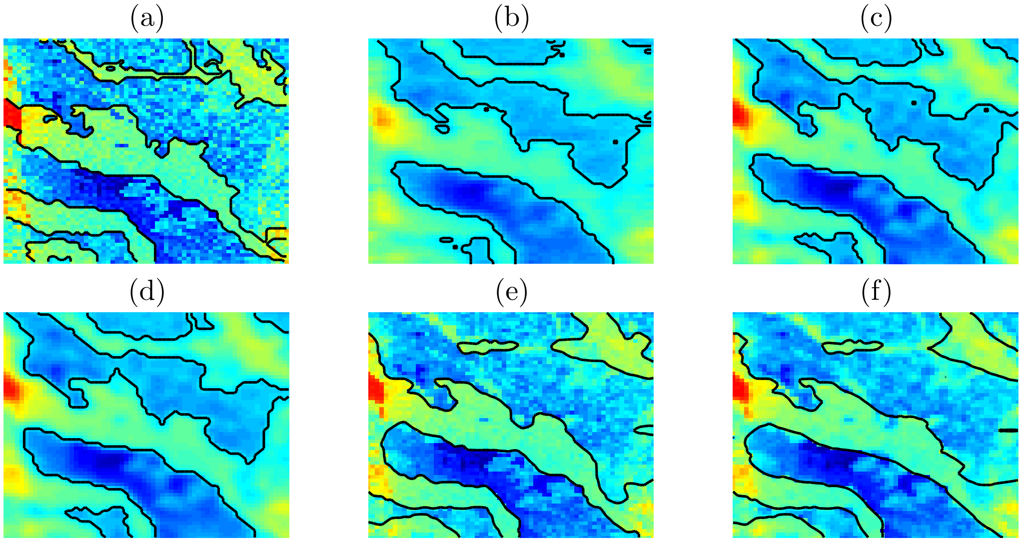


Figure 3.4. Estimates of the boundary between soil types given the observation shown in Figure 3.2. (a) The original soil moisture map with true boundary shown. (b) Estimate using kriging. (c) Estimate using gradient smoothing. (d) Estimate using smoothing splines. (e) Estimate using Mumford-Shah. (f) Estimate using MMS.

Figure 3.5(f) shows the field estimate using Mumford-Shah with the boundary localization overlaid in Figure 3.5(g) while Figures 3.5(h) and (i) shows the same using the MMS functional. As the initialization for the curve evolution of Mumford-Shah and MMS, we again use a threshold of the moisture data where there are observations. Across the unobserved block, we linearly interpolate the thresholded boundary. The results from Mumford-Shah and MMS preserve the boundary between the two soil types, which lead to more accurate field estimates. In particular, Figures 3.5(f) and (h) do not exhibit blurring across the soil boundaries which is the case in Figures 3.5(c), (d), and (e).

Again, we evaluate the boundaries using binary thresholding. For the boundaries shown in Figure 3.6, the NSD for the boundary determined using kriging is 0.4327, that for gradient smoothing is 0.3519, and 0.3569 for second-order smoothing splines. In contrast, the NSD for Mumford-Shah is 0.2909 while that for MMS is 0.3469. Table 3.2 summarizes the results of the boundary estimates for these two examples.

In our third example, we simulate the situation where we have a sparse set of sensors to collect measurements. In particular, suppose our observations are 11% of the spatial field of interest, as shown in Figure 3.7(b). Figure 3.7(c) shows the result using kriging, Figure 3.7(d) shows the field estimate using gradient smoothing, while Figure 3.7(e) illustrates the result using second order smoothing splines. As one would expect, each of these three methods smoothes the region between the sparse set of observations.

Now, given the measurements alone, it would be difficult to find a reasonable initial-

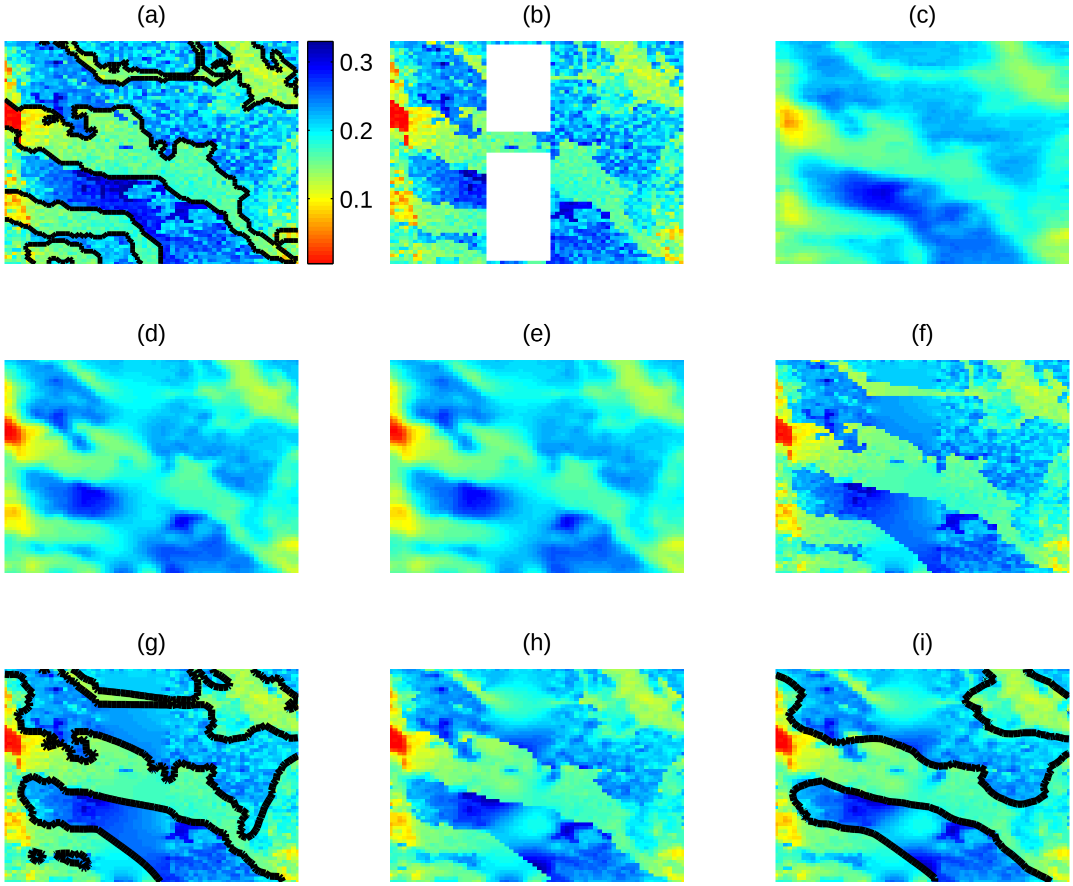


Figure 3.5. Estimation of soil moisture. Images show fractional volume of water in the soil. (a) The original soil moisture map with colorbar (valid for all images) and true boundary overlaid. (b) The observed data, a moisture map with a rectangular region unobserved (representing 81% observed data). (c) Estimate using kriging. (d) Estimate using gradient smoothing. (e) Estimate using smoothing splines with second-order derivatives. (f) Field estimate using Mumford-Shah. (g) Field and boundary estimates using Mumford-Shah. (h) Field estimate using MMS. (i) Field and boundary estimates using MMS.

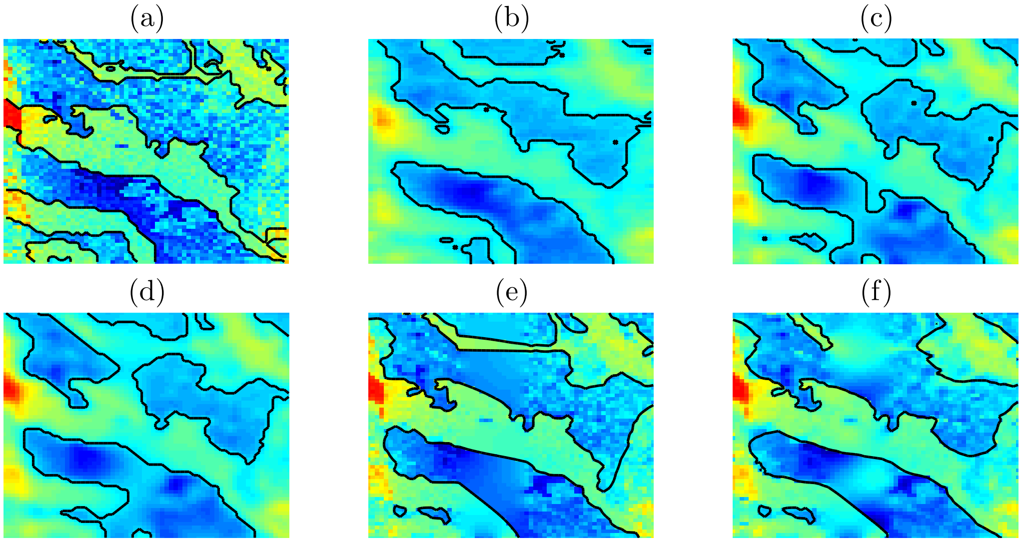


Figure 3.6. Estimates of the boundary between soil types given the observation shown in Figure 3.5. (a) The original soil moisture map with true boundary shown. (b) Estimate using kriging. (c) Estimate using gradient smoothing. (d) Estimate using smoothing splines. (e) Estimate using Mumford-Shah. (f) Estimate using MMS.

	% obs	Kriging	Gradient Smoothing	Smoothing Splines	Mumford-Shah Functional	MMS Functional
NSD	90%	0.3469	0.4158	0.4105	0.1502	0.1789
	81%	0.4327	0.3519	0.3569	0.2909	0.3469
Dice	90%	0.6987	0.7438	0.7403	0.8544	0.8254
	81%	0.6549	0.7005	0.7045	0.7682	0.7157

Table 3.2. Normalized symmetric difference and dice coefficient between true boundary and estimated boundary. For the 90% observed example, we see that Mumford-Shah and MMS do much better than the traditional methods while for 81% observed, these two methods do slightly better.

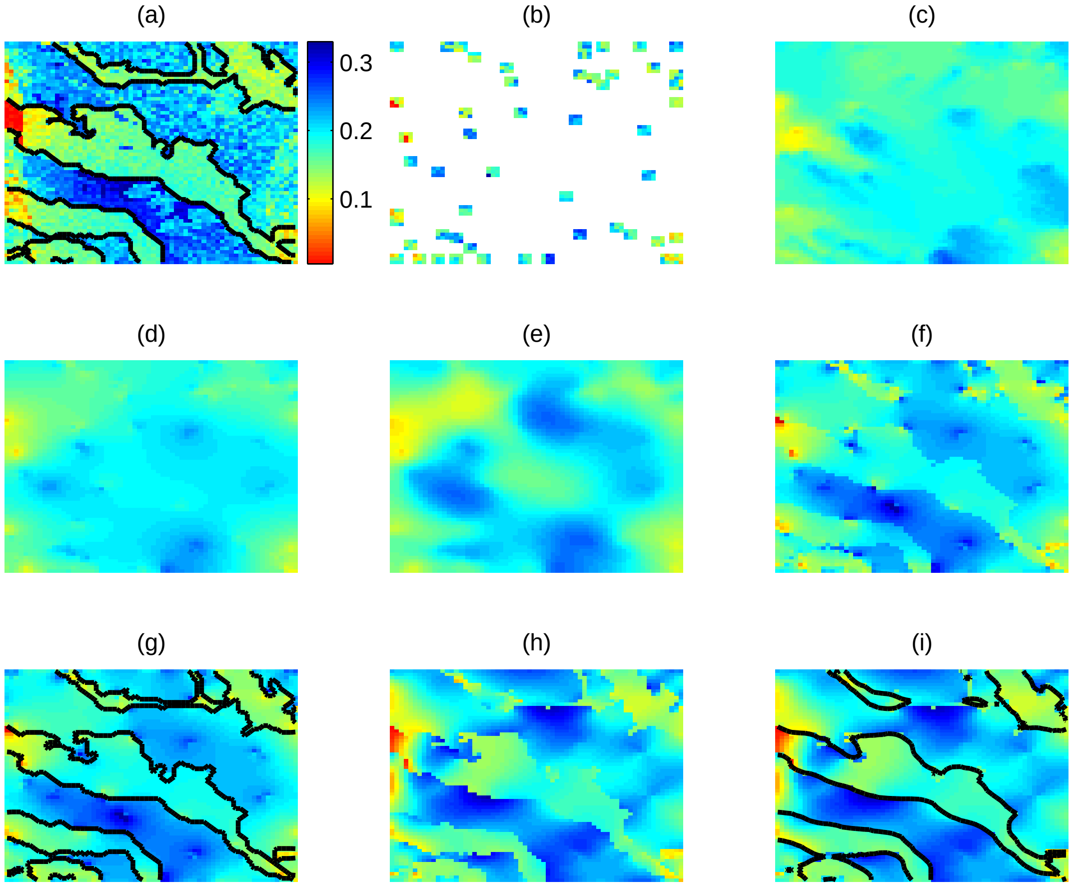


Figure 3.7. Estimation of soil moisture where only a sparse set of measurements are made. Images show fractional volume of water in the soil. (a) The original soil moisture map with colorbar (valid for all images in this figure) and true boundary overlaid. (b) Observed data representing 11% of the entire spatial field. (c) Estimate using kriging. (d) Estimate using gradient smoothing. (e) Estimate using smoothing splines with second-order derivatives. (f) Field estimate using Mumford-Shah. (g) Field and boundary estimates using Mumford-Shah. (h) Field estimate using MMS. (i) Field and boundary estimates using MMS.

ization for the curve evolution methods (Mumford-Shah and MMS). So suppose that we have some knowledge of the boundary, either from a segmentation of the same spatial domain at a previous time in the current observation sequence or from an ensemble of historical data. In our experiments, we take the segmentation from a different simulation of this region and use this as the initialization. With this as a starting point, we show that we can provide a reasonable estimate for the boundary location as well as a field estimate of the spatial domain that appears more accurate than the conventional interpolation methods shown in Figures 3.7(c-e).

Figure 3.7(f) shows the result using Mumford-Shah (the boundary is shown in Figure 3.7(g)), while Figures 3.7(h) and (i) illustrate the same using MMS. In Mumford-Shah and MMS, the field is smoothed within each region, but the discontinuity is preserved. Furthermore, finding the optimal field for each region separately provides a more accurate moisture estimate for both regions.

Given a sparse dataset such as this example, localization of the boundary would be difficult without a good initialization. However, the purpose of this example is to illustrate that, by incorporating any available prior information, our method uses this knowledge to provide an improved boundary estimate based on the measurements as well as a reasonable estimate of the field.

Before we conclude this discussion on soil moisture data, we discuss the computational complexity of each of the methods analyzed. The following computation times are obtained using a desktop personal computer having a Xeon $2.2GHz$ processor running MATLAB version 6.5 on a Linux OS. The computed times are based on the time required to produce a field estimate of the soil moisture maps as shown in Figure 3.5. The soil moisture field consists of a 64×64 grid of pixels. Kriging requires 8.56 seconds to produce the field estimate, while gradient smoothing takes 1.75 seconds. Iterating Mumford-Shah for 100 steps (an approximate time for convergence) requires 20.53 seconds.

Mumford-Shah uses first derivatives for interpolation in a manner similar to that in gradient smoothing. Although Mumford-Shah has greater computational complexity, it provides an improved field estimate over gradient smoothing. Furthermore, Mumford-Shah performs the additional task of localizing the boundary. To provide a similar comparison with second order smoothing splines, we choose the second derivative operator for $\mathcal{D}$ in MMS for our analysis. Second order smoothing splines requires 10.73 seconds, while the computation time for MMS for this choice of $\mathcal{D}$ is 32.98 seconds for 100 iterations. MMS results in somewhat higher computational complexity than second order smoothing splines, however in return it provides an improved field estimate as well as boundary localization.

■ 3.4.2 Sea Surface Temperature Data

In soil moisture estimation, because each soil type has a mean that does not vary systematically, Mumford-Shah and MMS yield similar results. However, in the oceanographic application we consider here, it is known that sea surface temperatures vary spatially in a systematic manner. In the following analysis of SST measurements, we show that in an application with spatially-varying fields, MMS provides better estimates than Mumford-Shah.

Synthetic SST Data

We initially analyze a synthetic data sequence designed to simulate a region in the North Atlantic. First, we define a temperature discontinuity characteristic of an oceanic front. From this, a spatially-varying mean temperature map is generated using the following assumptions: (a) south of the discontinuity, the mean temperature is $25 + 0.002d(\mathbf{x})$ degrees Celsius, where $d(\mathbf{x})$ is the distance in kilometers of the point $\mathbf{x}$ from the front, (b) north of the discontinuity, the mean temperature is $20-0.01d(\mathbf{x})$ degrees Celsius. We use a linear temperature model because we assume that we are operating in the linear portion of a particular temperature model $T = T_0 cos(sin(\theta))$, where θ is degrees latitude and T_0 is the equatorial temperature. This model is obtained as an approximation of surface temperature based on a 1-D energy balance model of climate [18, 155]. Next, we define a Gaussian covariance matrix for each region to enforce correlation between nearby points. In particular, we choose $K[\mathbf{x}, \mathbf{z}] = exp[-0.02d^2(\mathbf{x}, \mathbf{z})]$ (where $d(\mathbf{x}, \mathbf{z})$ is the distance in kilometers between $\mathbf{x}$ and $\mathbf{z}$) for points $\mathbf{x}$ and $\mathbf{z}$ on the same side of the boundary[4]. Using these assumptions, our synthetic temperature map is generated and shown in Figure 3.8(a). Finally, independent, identically distributed pixelwise white noise with zero mean and standard deviation of two is added to simulate measurement noise. Figure 3.8(b) shows such synthetically generated measurements.

First, we apply the MMS functional to the dataset shown in Figure 3.8(b). Again, we make the assumption that b in Equation (3.3) is the Dirac delta function. However, instead of choosing the differential operator $\mathcal{D}$, we assume we know the true covariance K (given b, this choice of K implicitly determines the choice of $\mathcal{D}$). The results of the field estimation are shown in Figure 3.8(c), with the estimated boundary overlaid. From this image, we observe that the MMS method localizes the boundary and substantially reduces the noise in the field (quantitatively, the reduction in standard error is 26%).

Satellite measurements have unobserved regions due to a combination of cloud cover and the limited range of the satellite's swath. Motivated by this fact, Figure 3.9(a) shows an example where a spatial region is not fully observed (the area with no measurements is shown in white). As with the soil moisture examples, we compare the joint boundary and field estimation techniques with three traditional methods of interpolation. In particular, we apply kriging (Figure 3.9(b)), gradient smoothing (Figure 3.9(c)), and second order smoothing splines (Figure 3.9(d)) to the observation shown in Figure 3.9(a). In comparison with the true field in Figure 3.8(a), we observe that the field estimates using kriging, gradient smoothing, and smoothing splines do not accurately preserve the boundary in the unobserved region.

---

[4] The multiplicative constant 0.02 was chosen so that there would be some non-local dependencies, but it had to be chosen carefully to ensure that K was a mathematically valid covariance [140]. While not an issue for 1-D problems, a slowly-decaying Gaussian in 2-D can result in an indefinite matrix.

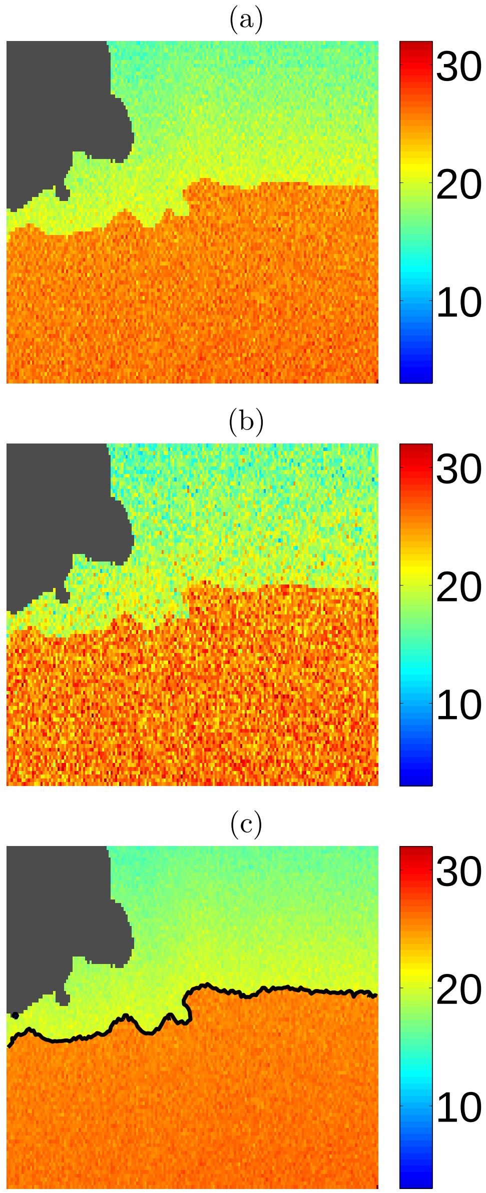


Figure 3.8. Boundary and field estimation of synthetic SST field (temperatures in degrees Celsius). (a) The original synthetic field to be estimated (dark mass at top left is land). (b) Simulated measurements containing additive IID noise. (c) Field estimate using MMS with boundary overlaid. Note that this estimate removes some of the IID measurement noise.

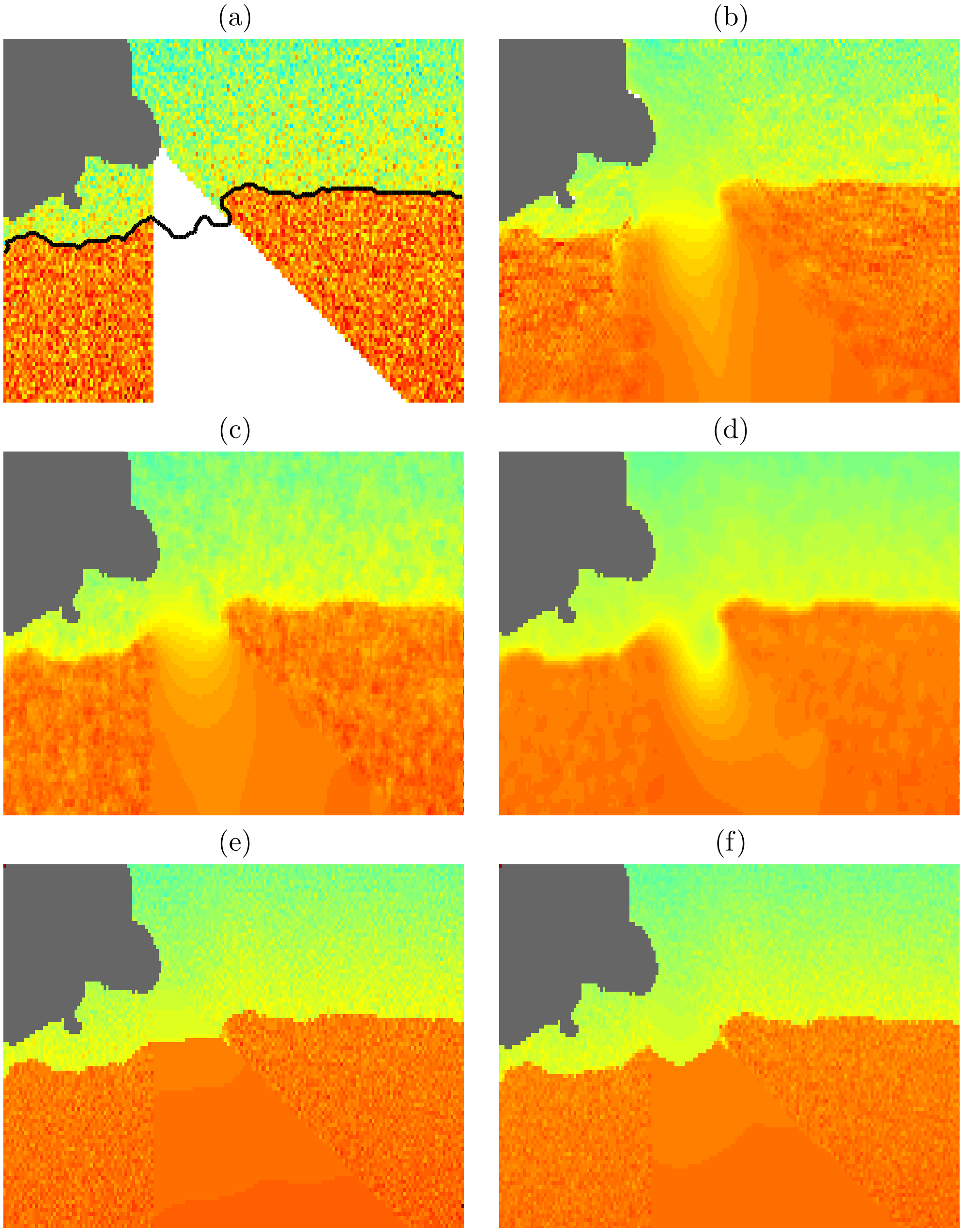


Figure 3.9. Boundary and field estimation of synthetic SST field. (a) The measurement data is as in Figure 3.8(b), except that a triangular region is unobserved. The true boundary is also shown. (b) Field estimate using kriging. (c) Field estimate using gradient smoothing. (d) Field estimate using second order smoothing splines. (e) Field and boundary estimate using the Mumford-Shah functional. Note that the boundary is a straight line across the region of missing observations. (f) Field and boundary estimate using the MMS functional.

	Kriging	Gradient Smoothing	Smoothing Splines	Mumford-Shah Functional	MMS Functional
Missing Triangle	2.41	1.42	1.71	1.36	1.15
Observed Region	1.64	1.05	0.99	0.45	0.55
Entire Image	1.84	1.14	1.19	0.76	0.73

Table 3.3. Standard error (in degrees Celsius) in the interpolated triangular region, the observed region, and the entire region of the synthetic temperature map shown in Figure 3.9.

Figure 3.9(e) shows the result using the Mumford-Shah functional from Equation (3.1), while Figure 3.9(f) shows the outcome using the MMS functional proposed in Equation (3.4). For the latter, $\mathbf{m}_{R_i}$ and $K_{R_i}^{-1}$ are the mean and covariance statistics, respectively, used in the simulation shown in Figure 3.8(a). Note that in the regions of missing observations, the estimate using MMS and Mumford-Shah both provide a smooth temperature map. However, the estimated curves differ in the unobserved region. With Mumford-Shah, the estimated curve is almost a straight line across the missing observations. This occurs because the primary force in this region is due to the curve length penalty. However, with MMS, the curve exhibits characteristics more similar to the true boundary seen in Figure 3.8(a). Minimizing the MMS functional yields this boundary because the prior means play a role in the minimization process, resulting in an upward concave curve where no measurements are present. Quantitatively, Mumford-Shah yields a standard error of 1.36 degrees per pixel in the triangular region, while MMS has a standard error of 1.15 degrees per pixel in the same region. Based on these results, we determine that the improved boundary estimate of MMS in the unobserved area results in a much better field estimate for that region. Table 3.3 shows the standard error exhibited by each method within the unobserved triangular region.

Figure 3.10 shows four steps of the curve evolution process from an initial boundary (Figure 3.10(a)), across two intermediate steps (first Figure 3.10(b) and then Figure 3.10(c)), and the final boundary estimate (Figure 3.10(d)). The boundary in each case is overlaid on the synthetic data.

To determine the robustness of the results, we provide some sensitivity analysis on the initializations and the choice of parameters. Using the simulated measurements shown in Figure 3.8(b), we provide some analysis of how the results change with a corresponding change in parameter or initialization. Figure 3.11(a) shows the result previously shown in Figure 3.8(c). For this example, the parameters used are $\alpha = 0.7$, $\beta = 0.8$, and $\gamma = 8$. Regarding sensitivity to initializations, we randomly choose 40 smooth initial curves which extend from the left edge of the image to the right edge (one example of which is the initialization shown in Figure 3.10(a)) which lie within the two bands shown in Figure 3.11(b). For each of these initializations, the algorithm converges to an estimate that is visibly indistinguishable from the one shown

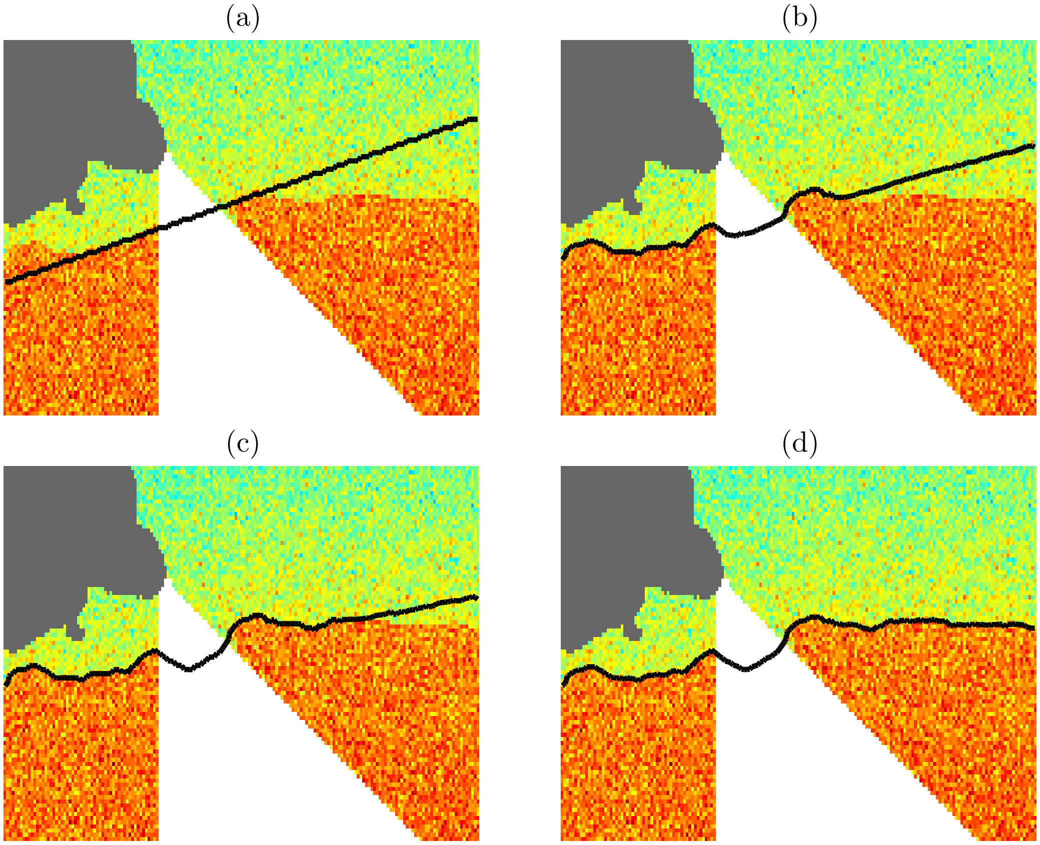


Figure 3.10. Demonstration of curve evolution using the MMS functional. (a) The initial boundary, which is a straight line overlaid on the measurement data. (b) An intermediate step in the evolution. (c) A later intermediate step in the evolution. (d) Final boundary in the curve evolution process.

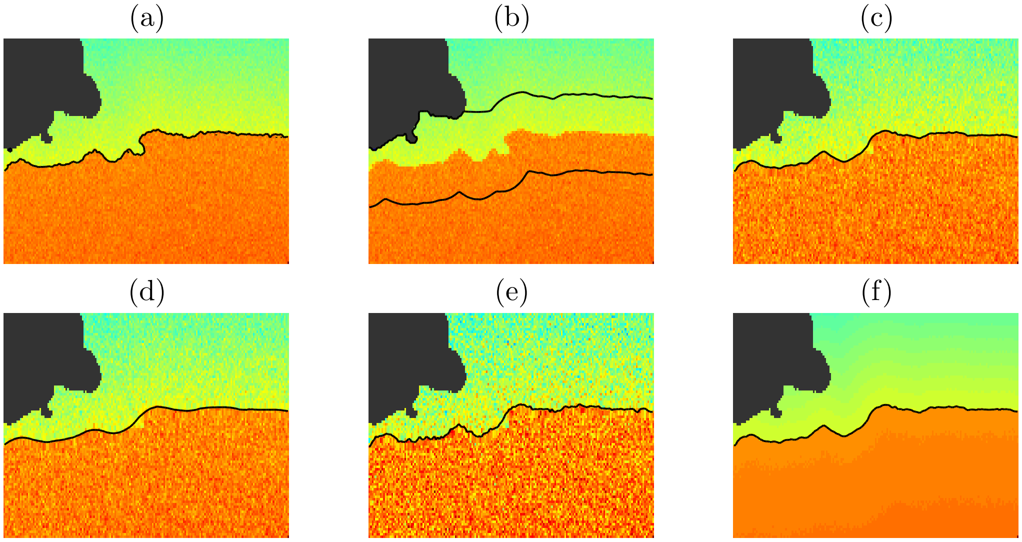


Figure 3.11. Sensitivity analysis for initializations and parameter choice. (a) The field and boundary estimate of Figure 3.8 using $\alpha = 0.7$, $\beta = 0.8$, and $\gamma = 8$. (b) Graph which shows band of possible initializations which result in the estimate in (a). (c) Result when the curve length parameter γ is set to 16. (d) Result when the curve length parameter γ is set to 40. (e) Result when the data fidelity parameter α is set to 7. (f) Result when the smoothness parameter β is set to 8.

in Figure 3.11(a). Such a result indicates that the method is reasonably robust to different initializations of the boundary. Next, we examine the effect of increasing γ while leaving the other parameters fixed. Increasing γ first to 16 (Figure 3.11(c)) and then to 40 (Figure 3.11(d)) results in a progressive smoothing of the boundary as expected. Resetting γ back to 8 and setting α to 7 results in Figure 3.11(e). Note that the result is noisier than Figure 3.11(a) because it adheres more strongly to the observations. Finally, having $\alpha = 0.7$ and $\gamma = 8$, we set β to 8, resulting in a smooth field estimate as shown in Figure 3.11(f).

Satellite SST Data

To demonstrate MMS on real data, we apply this functional to satellite SST data. In particular, we consider satellite images of the North Atlantic Ocean. Figure 3.12(a) is a satellite image which shows a SST map having very little missing data. In this image, the dark mass in the upper left is land, the white regions represent missing observations, the blue regions are regions of cool water, and the orange and red regions are regions of warmer water. A temperature discontinuity exists at the Gulf Stream's north wall, as seen by the sharp separation of cooler waters in the north from the warmer waters in the south in Figure 3.12(a). The satellite data is obtained from the National Oceanic and Atmospheric Administration (NOAA) Geostationary Operational Environment Satellite (GOES), courtesy of the Jet Propulsion Laboratory (JPL). For this set of data, the means were chosen to be linear functions of the distance from the curve, with the parameters determined empirically from examining historical sea surface temperature data of the North Atlantic. In particular, the mean temperature south of the Gulf Stream's north wall is assumed to be $24.17+0.01d(\mathbf{x})$ degrees Celsius, where $d(\mathbf{x})$ is the distance in kilometers of the point $\mathbf{x}$ from the front. Similarly, the mean temperature north of the discontinuity is modeled as $19.22-0.014d(\mathbf{x})$ degrees Celsius.

The covariance matrix was chosen to be the same as that used in the previous synthetic data example ($K[\mathbf{x},\mathbf{z}] = exp[-0.02d^2(\mathbf{x},\mathbf{z})]$). A Gaussian covariance is a common model for field covariance within geophysical applications [43]. Figure 3.12(a) shows a satellite SST image on a particular day where there is only a small percentage of missing observations due to cloud cover. By applying the MMS functional, the field estimate and boundary are shown in Figure 3.12(b). For this example, the previous day's boundary was used as an initial estimate for the curve. The knowledge that the north wall of the Gulf Stream does not vary much from day to day is incorporated here to provide us with a good initialization for the curve $\vec{C}$. Having a good initialization generally leads to faster convergence during curve evolution.

In the previous example, the amount of missing observations was small. In some cases, a large portion of the region of interest may be unobserved. In these cases, the prior knowledge of the means and covariance plays a larger role in both the interpolation and boundary estimation. To illustrate, we consider SST maps for two different days which have larger areas which are unobserved (see Figure 3.12(c) and Figure 3.12(e)). Whereas the image in Figure 3.12(a) had 87% of the spatial domain visible, the observation in Figure 3.12(c) has 66% data visible and the image in Figure 3.12(e) has 19% data visible. Figure 3.12(d) shows the boundary and field estimate based on observations from Figure 3.12(c), while Figure 3.12(f) shows the same given Figure 3.12(e). In both of these examples, we have used the previous day's segmentation for the initial curve. We note that initializations within several pixels of the final location will result in the same local minimum. As with the synthetic SST example, Figure 3.13 demonstrates the robustness of different initializations on the real data. Namely, we randomly choose 25 smooth initial curves which extend from the left edge of the image to the right edge within the two red bands shown in the image. Each of these initializations results in estimates visually indistinguishable from the one shown in Figure 3.12(f).

Since the Gulf Stream's north wall does not move very much from day to day, using the previous day's segmentation as the initialization helps to ensure that the final segmentation is likely the boundary we desire. Although we do not have access to the underlying truth in this experiment, we observe that the locations and shapes of the boundary estimates in all three examples are similar. Because the estimate based on 19% visible observations is similar to that based on 87%, we conclude that in the estimate shown in Figure 3.12(f), the MMS functional does well in locating the boundary despite the sparsity of observations. In these two cases, the prior mean, coupled with observations far from the curve in the eastern portion of the North Atlantic, allows us to provide a reasonable estimate of the sea surface temperature in the unobserved regions. The field estimates also appear reasonable when compared to the field estimate in Figure 3.12(b), an estimate made in the presence of most of the observations of that frame.

We now qualitatively examine the difference between the Mumford-Shah and MMS functionals on the GOES data. Figure 3.14 provides a visual comparison of the field estimate using each functional based on the measurements shown in Figure 3.14(a). Figure 3.14(b) shows the field estimate using Mumford-Shah. Note that unlike the temperature estimate using MMS (Figure 3.14(c)), the isotherms shown in Figure 3.14(b) are not parallel to the oceanic front north of the Gulf Stream's north wall. The estimate bears this characteristic because Mumford-Shah smoothes across the unobserved region north of the boundary. In this particular example, the warm observations (shown in green and yellow) are in the western portion of the region while the cooler observations (blue) are in the eastern part of the region. Consequently, the temperature smoothing results in a west to east gradient. If we examine Figure 3.12(a), we see that the region immediately north of the Gulf Stream's north wall is generally yellow in color. In Figure 3.14(b), the region immediately north of the north wall in the eastern portion of the image is blue, indicating that the temperature has been underestimated. On the other hand, MMS has the physically more consistent field estimate which exhibits a south to north cooling as shown in Figure 3.14(c), and is yellow in the region immediately north of the Gulf Stream's north wall. This is achieved because the MMS functional exploits the spatial statistics of the different regions.

■ 3.5 Conclusion

In this chapter, we have proposed methods of estimation and interpolation for earth science applications on a field having a discontinuity with unknown location and arbitrary shape. The two techniques we presented, Mumford-Shah and MMS, solve a joint problem of boundary localization and field estimation. These two methods preserve discontinuities in the field. However, the benefit of our proposed MMS technique is that it allows for the incorporation of a variety of first and second order statistics. We have shown that certain estimation problems (e.g. SST estimation) yield improved results when the spatial statistics of the underlying field can be included in the functional formulation. Using sea surface temperature and simulated soil moisture data as examples, we have demonstrated the advantages of the proposed methods over three existing techniques of interpolation and smoothing that do not handle discontinuities.

The Mumford-Shah and MMS methods can be applied to other earth science applications. In particular, we may want to locate the boundaries of the cusp-like structures of tropical instability waves near the equator from ocean color data [85] or locate rainstorms and estimate rainfall over a region using a combination of radar sensors [70].

The MMS functional can also be adapted to incorporate dynamic information about the boundary when this information is available, especially when we are estimating the field and boundary over a temporal sequence. In particular, we can replace the curve length penalty term of Equation (3.3) with a term that penalizes deviations of the current curve from a predicted curve. The next chapter examines the usefulness of dynamics in the segmentation problem.

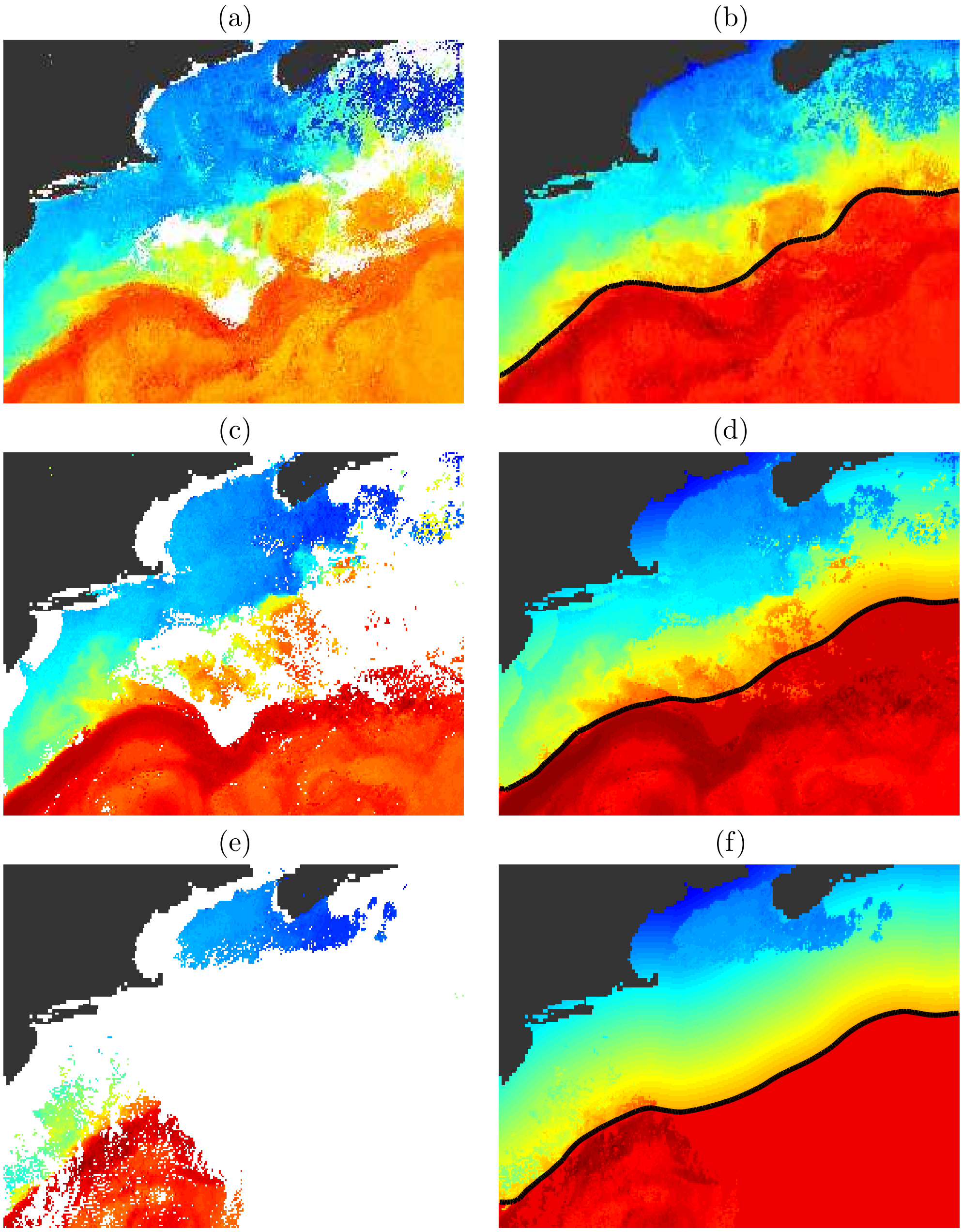


Figure 3.12. Satellite SST estimation using MMS on GOES data (different images over the course of a month) with missing observations (white regions) due to cloud cover. (a) A satellite SST measurement with relatively few missing observations. (b) Field and boundary estimate based on measurements in (a). (c) A different SST map with more missing regions in the measurements. (d) Field and boundary estimate based on the measurements in (c). (e) A third SST map with most of the measurement information missing. (f) Field and boundary estimate based on (e).

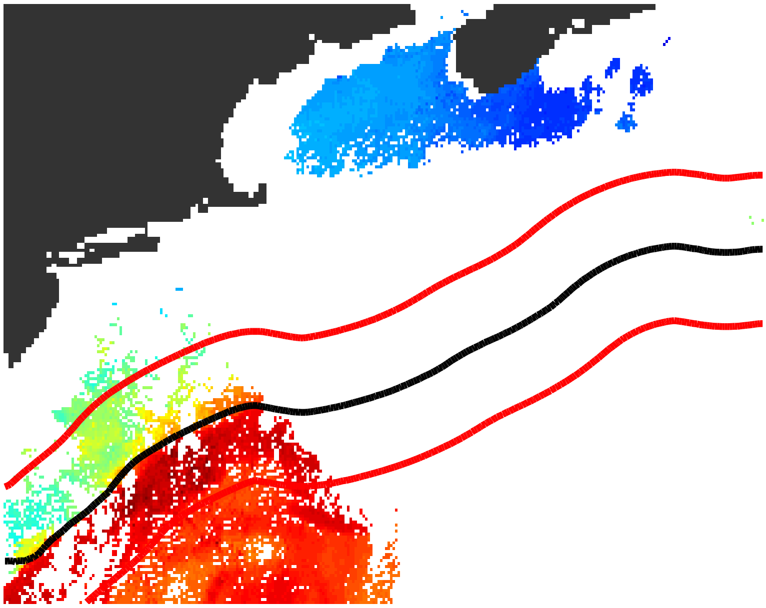


Figure 3.13. Sensitivity analysis of initializations. Any of the 25 random initializations contained within the red bands results in segmentations visually indistinguishable from the one shown in black.

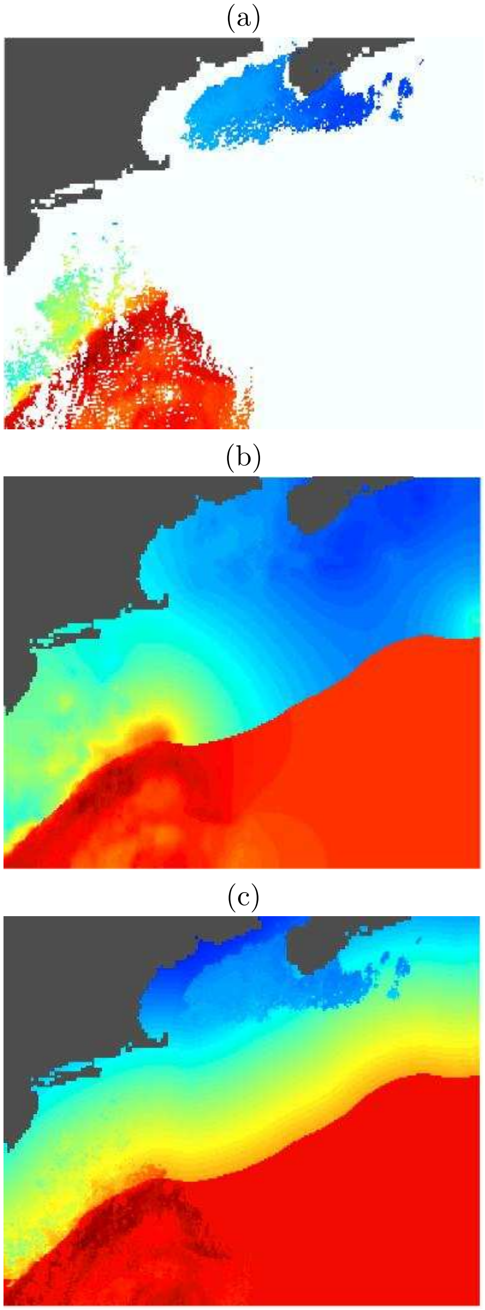


Figure 3.14. Visual comparison of Mumford-Shah with MMS. (a) Observed data. (b) Field estimate of observed data using Mumford-Shah. (c) Field estimate using MMS.

Chapter 4

Recursive Segmentation of Deformable Objects through Learning the Dynamics

WHILE the previous chapter dealt with estimation within a static framework, this chapter introduces the use of temporal information in the estimation of boundaries. Incorporating dynamics adds value when temporally-adjacent frames are correlated with each other. In such situations, knowing the segmentation of an object from a previous or future frame can aid in the segmentation of the object in the current frame. We formulate the problem of estimating a boundary based on a temporal sequence of images using a principled recursive estimation framework. To solve this estimation problem, we first propose a method for learning the dynamics of a deformable object using an information-theoretic technique. Having learned the system dynamics, we then apply temporal object segmentation techniques to the specific problem of locating the left ventricle (LV) across a cardiac cycle from MR data. We note that this methodology is not restricted to segmenting the LV nor this particular imaging modality. In addition, this technique can be adapted to and used for non-medical temporal segmentation problems.

Having accurate LV segmentations across a cardiac cycle provides quantitative and qualitative information for diagnosis of certain heart conditions. Existing LV segmentation techniques are founded mostly upon algorithms for segmenting static images. In order to exploit the dynamic structure of the heart in a principled manner, we approach the problem of LV segmentation as a recursive estimation problem. In our framework, LV boundaries constitute the dynamic system state to be estimated, and a sequence of observed cardiac images constitute the data. By formulating the problem as one of state estimation, the segmentation at each particular time is based not only on the data observed at that instant, but also on predictions based on past segmentations, often referred to as a *filtering* problem (in Chapter 5, we base our estimates on both past and future segmentations as we formulate the problem as one of *smoothing*). This requires a dynamical system model of the LV which we propose to learn from training data through an information-theoretic approach. To incorporate the learned dynamic model into our segmentation framework and obtain predictions, ideas from particle filtering are used. We demonstrate the effectiveness of the proposed approach on a set of cardiac images. From the results, our approach appears to provide more accurate segmentations than those obtained from static image segmentation techniques, especially when the observed data are of limited quality.

In Section 4.1, we explain why we examine the problem of LV segmentation across a cardiac cycle. The framework and methodology of the segmentation technique we propose is described in Section 4.2. In Section 4.3, we describe how we use sample-based methods to recursively estimate the LV boundaries. In Section 4.4, we explain how the dynamics of the system are learned. Experimental results are shown in Section 4.5 for both low and high SNR examples, and we summarize the chapter in Section 4.6.

■ 4.1 Motivation for and Recent Work on Cardiac Segmentation

Of the cardiac chambers in the heart, the left ventricle is quite frequently analyzed because its proper function, pumping oxygenated blood to the entire body, is vital for normal activity. One quantitative measure of the health of the LV is ejection fraction (EF). This statistic measures the percentage volume of blood transmitted out of the LV in a given cardiac cycle. To compute EF, we need to have segmentations of the LV at multiple points in a cardiac cycle; namely, at end diastole (ED) and end systole (ES). In addition, observing how the LV evolves throughout an entire cardiac cycle allows physicians to determine the health of the myocardial muscles. Segmented LV boundaries can also be useful for further quantitative analysis. For example, past work [53,128] on extracting the flow fields of the myocardial wall assumes the availability of LV segmentations throughout the cardiac cycle.

Automatic segmentation of the left ventricle in bright blood cardiac magnetic resonance (MR) images is non-trivial because the image intensities of the cardiac chambers vary due to differences in blood velocity [182]. In particular, blood that flows into the ventricles produces higher intensities in the acquired image than blood which remains in the ventricles [71]. Locating the LV endocardium is further complicated by the fact that the right ventricle and aorta often appear jointly with the LV in many images of the heart. Similarly, automatic segmentation of low signal-to-noise ratio (SNR) cardiac images (e.g. body coil MR or ultrasound) is difficult because intensity variations can often obscure the LV boundary.

Early work on cardiac segmentation focused on the static problem. Goshtasby and Turner [71] apply a two-stage approach to extract the left and right ventricles. First, they apply intensity thresholding for a rough estimate and then find locally maximal gradient magnitude points to determine the final segmentation. Weng *et al.* [190] consider a similar approach for thresholding, while Geiger *et al.* [65] apply dynamic programming (DP) to locate strong gradients near the initialization. Previously (in Section 2.2.3), we explained how Chakraborty *et al.* [25] was one of the first to combine gradient and region techniques for image segmentation. Their work focuses on the LV

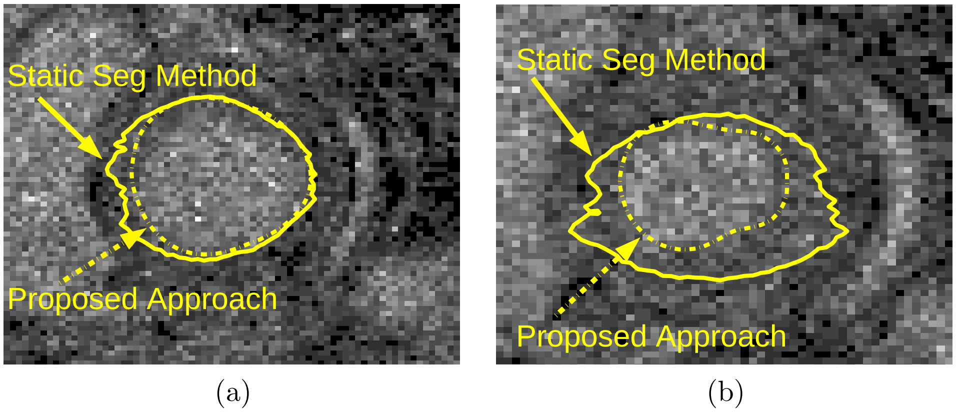


Figure 4.1. Sample segmentations of the LV, one near end diastole (a) and the other near end systole (b), using both a static segmentation and our recursive framework.

segmentation problem. Similarly, Paragios [141] also proposes the combination of edge and region-based techniques to segment the LV endocardium and epicardium.

During a single cardiac cycle, which lasts approximately 1 second, the heart contracts from end diastole (ED) to end systole (ES) and expands back to ED. Over this time, MR systems can acquire approximately 20 images of the heart. Because adjacent frames are imaged over a short time period (approximately 50 ms), the LV boundaries exhibit strong temporal correlation. Consequently, previous LV boundaries may provide information regarding the location of the current LV boundary. Using such information is particularly useful for low SNR images, where the observation from a single frame alone may not provide enough information for a good segmentation. For illustration, Figure 4.1 shows the results of static segmentations compared with results obtained using our proposed approach described in this chapter. Our method exploits the dynamic structure of the heart and incorporates information from past frames. These examples demonstrate the potential value of incorporating dynamics.

Previous work incorporating some form of dynamics includes that of Chalana *et al.* [26] and Jolly *et al.* [86]. Each took the segmentation from the most recent frame and performed a local search for strong gradients in the segmentation of the current frame (one can interpret this as a simple dynamic where the prediction is the estimated state from the previous frame). More recent work by Zhou *et al.* [201] and Senegas *et al.* [156] take the process a step further by incorporating non-trivial dynamics on the previous segmentations to provide better predictions. Zhou *et al.* [201] consider LV shape tracking by using linear system dynamics, assumed known, to produce a prediction. Then, this prediction is incorporated with the observation to generate an estimate. The prediction and observation are treated as noisy measurements with associated covariances and are fused by obtaining the best linear unbiased estimate (BLUE) of the state assuming that the two sources are independent [7, 115]. Their technique uses landmark points to represent the LV boundaries, thus introducing the issue of correspondence. Furthermore, all uncertainties are assumed to be Gaussian.

Senegas *et al.* [156] use a Bayesian framework for segmenting a temporal sequence of images using a sample-based approach. Their formulation is most similar to our approach because they use sample-based methods and recursively estimate the LV. Using sequential Monte Carlo sampling, they use particles to approximate the posterior at each time. They define their shape space using spherical harmonics (for two-dimensional polar coordinates, this amounts to Fourier decomposition). For the dynamics of the system, the mean shape at end diastole (ED) and end systole (ES) are determined from training data. In a test sequence, the prediction is assumed Gaussian with the mean shape given by a linear combination of the estimated shape of the previous frame and a fraction of the shape difference between the mean shapes at ED and ES.

In their work, the assumption of a linear transition between end diastole and end systole may be somewhat limiting. In our work, we learn a more general dynamic model from the training data. The dynamics are learned on a low-dimensional state which is coupled to the LV boundary through the process of curve evolution. We use curve evolution to find the most probable location of the LV boundary based on a combination of the observation at the current frame and the prediction density rather than taking a representative sample from the density estimate of the state. In addition, although we solve the recursive estimation problem similarly as a filtering problem in this chapter, we extend the estimation in the next chapter to include observations from the entire sequence rather than just observations past and present.

■ 4.2 Framework and Methodology

We formulate the LV segmentation and filtering problem as an estimation of the posterior distribution of the boundary at each discrete time t_0 based on data from $t = 1$ to $t = t_0$. Let y_t be the image data which are noisy measurements of the blood and tissue intensity field f_t and define X_t as the dynamic system state which contains information about the LV boundary $\vec{C}_t$ at t.

■ 4.2.1 Low-Dimensional Representation of the Boundary

Figure 4.2 shows the graphical model representation of the problem. One can observe from the model that the LV boundary $\vec{C}_t$ depends on both the observation y_t and the current state X_t. The state X_t in our model is a simple, parametric, low-dimensional approximation of the LV boundary (details are discussed in Section 4.4.1). We choose a low-dimensional model because we do not want to overfit the training data (i.e. to avoid capturing individual patient idiosyncrasies). Our particular choice of representation captures the low frequency features of the boundary. As a by-product, such a representation provides regularization in the sense that curves represented by X_t are smooth. We statistically learn the dynamics, using the representation to capture the salient parts of the evolution. The low-dimensionality allows the problem of learning the dynamics to be computationally tractable. Furthermore, given a finite set of training

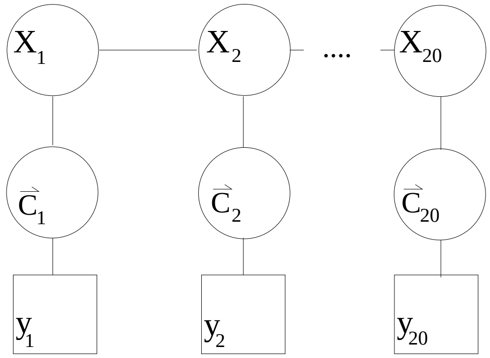


Figure 4.2. Graphical model representing our framework. The chain has length twenty because a single cardiac cycle has twenty temporal frames.

data, the dimensionality cannot be too large or else there may be insufficient richness in the data to obtain a good statistical model.

■ 4.2.2 Bayesian Formulation

In our formulation, we make use of the Markovianity of X_t and the conditional independence of the observations $y_{1:20} = [y_1, y_2, \cdots, y_{19}, y_{20}]$ in the graphical model shown in Figure 4.2. First, applying Bayes' Theorem to the posterior density $p(X_t|y_{1:t})$ and observing that $p(y_{1:t-1})$ and $p(y_{1:t})$ do not depend on X_t, we obtain

$$p(X_t|y_{1:t}) \propto \int_{f_t} \int_{\vec{C}_t} p(y_t|f_t, \vec{C}_t) p(f_t|\vec{C}_t) p(\vec{C}_t|X_t) p(X_t|y_{1:t-1}) d\vec{C}_t df_t \tag{4.1}$$

where f_t is the underlying blood and tissue intensity field,[1] $p(y_t|f_t, \vec{C}_t)$ is a likelihood function, $p(f_t|\vec{C}_t)$ is the field prior, $p(\vec{C}_t|X_t)$ is the curve prior, and $p(X_t|y_{1:t-1})$ is the prediction density. Substituting

$$p(X_t|y_{1:t-1}) = \int_{X_{t-1}} p(X_t|X_{t-1}) p(X_{t-1}|y_{1:t-1}) dX_{t-1} \tag{4.2}$$

into Equation (4.1), we have

$$p(X_t|y_{1:t}) \propto \int_{f_t}\int_{\vec{C}_t} [p(y_t|f_t, \vec{C}_t) p(f_t|\vec{C}_t) p(\vec{C}_t|X_t) \int_{X_{t-1}} p(X_t|X_{t-1}) p(X_{t-1}|y_{1:t-1}) dX_{t-1}] d\vec{C}_t df_t. \tag{4.3}$$

Note that the recursive nature of the problem (i.e. $p(X_t|y_{1:t})$ is written as a function of $p(X_{t-1}|y_{1:t-1})$) can be seen in Equation (4.3).

---

[1] f_t is a nuisance parameter that we need to estimate in order to determine the LV boundary.

Before proceeding, we make two approximations to this framework. First, we approximate the marginalization of f_t and $\vec{C}_t$ by choosing the $(f_t, \vec{C}_t)$ pair which maximizes the integrand of Equation (4.1). This maximization should ideally be done for each value of X_t. As a second approximation, we seek to maximize the $(f_t, \vec{C}_t)$ pair which maximizes the integrand of Equation (4.1) integrated over all possible values of X_t. That is, we want

$$(f_t^*, \vec{C}_t^*) = arg \max_{f_t, \vec{C}_t} \int_{X_t} p(y_t|f_t, \vec{C}_t) p(f_t|\vec{C}_t) p(\vec{C}_t|X_t) p(X_t|y_{1:t-1}) dX_t. \tag{4.4}$$

Using these approximations, we may rewrite Equation (4.1) as

$$p(X_t|y_{1:t}) \propto p(y_t|f_t^*, \vec{C}_t^*) p(f_t^*|\vec{C}_t^*) p(\vec{C}_t^*|X_t) p(X_t|y_{1:t-1}). \tag{4.5}$$

To implement temporal segmentation based on Equation (4.5), we need to overcome two challenges: the presence of arbitrary, non-Gaussian densities and non-trivial dynamics. To handle the former, Section 4.3 describes how we approximate the posterior density using a sample-based approach to represent the non-parametric densities. For the latter, Section 4.4 explains how we learn the dynamics.

■ 4.3 Approximating the Posterior Density

In this section, we describe how the posterior $p(X_t|y_{1:t})$ is recursively estimated using sample-based methods. Suppose that at time $t-1$ we approximate the posterior represented by an equally-weighted set of N samples $x_{t-1}^{(i)}$ (i $\in \{1, 2, \cdots, N\}$) using a kernel density estimate, so that

$$p(X_{t-1}|y_{1:t-1}) \approx \sum_{i=1}^{N} \frac{1}{N} k(X_{t-1}; x_{t-1}^{(i)}), \tag{4.6}$$

where $k(x; x^0)$ represents a kernel centered at x^0. We next sample the forward density $p(X_t|X_{t-1} = x_{t-1}^{(i)})$ (assumed known for now) for each $x_{t-1}^{(i)}$, obtaining MN equally-weighted samples $x_{t|t-1}^{(i,j)}$ (where $x_{t|t-1}^{(i,j)}$ for $j = 1, 2, \ldots, M$ are the M samples obtained from $x_{t-1}^{(i)}$) that serve as an approximate representation for the prediction density $p(X_t|y_{1:t-1})$. Thus, we can write

$$p(X_t|y_{1:t-1}) \approx \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{1}{MN} k(X_t; x_{t|t-1}^{(i,j)}). \tag{4.7}$$

We now want to use curve evolution to determine the $(f_t^*, \vec{C}_t^*)$ pair of Equation (4.4). First, define a functional $E(f_t, \vec{C}_t)$ by

$$E(f_t, \vec{C}_t) = -\log[\int_{X_t} p(y_t|f_t, \vec{C}_t) p(f_t|\vec{C}_t) p(\vec{C}_t|X_t) p(X_t|y_{1:t-1}) dX_t], \tag{4.8}$$

where the term inside the logarithm is what we seek to maximize. Since the log function is monotonic, minimizing $E(f_t, \vec{C}_t)$ will provide us with the same $(f_t^*, \vec{C}_t^*)$ pair. This fact is important because we have a method of minimizing $E(f_t, \vec{C}_t)$ using curve evolution. Before we describe how this is achieved, we must first examine the terms involved individually for the particular application of LV segmentation.

■ 4.3.1 Likelihood Term

For the MR data, we apply a simple observation model which assumes that the intensities are piecewise constant with a bright intensity representing blood (the region inside the LV boundary) and a darker one representing the myocardium (the region immediately outside the LV boundary). Intensity variations in the observation, such as those due to differences in blood velocity [71], are modeled through a multiplicative random field (other choices of noise models can be handled in our framework, with the result being a different observation model). Mathematically, the observation model is

$$y_t(z) = \begin{cases} f_t^{R_{in}(\vec{C}_t)} \cdot n(z) & , z \in R_{in}(\vec{C}_t) \\ f_t^{R_{out}(\vec{C}_t)} \cdot n(z) & , z \in R_{out}(\vec{C}_t), \end{cases} \tag{4.9}$$

where $f_t^{R_{in}(\vec{C}_t)}$ and $f_t^{R_{out}(\vec{C}_t)}$ are the constant, but unknown, field intensities for the blood pool region inside, R_{in}, and the myocardial region immediately outside (within five pixels - to provide context, the average area of the LV at end diastole is 1889 pixels while that at end systole is 380 pixels), R_{out}, of the LV boundary, respectively, and $n(z)$ is spatially independent, identically distributed lognormal random field with log $n(z)$ a Gaussian random field having zero mean and variance σ_n^2 (used as a rough approximation to speckle noise). Note that we explicitly indicate the dependence of the regions on $\vec{C}_t$. We choose $R_{out}(\vec{C}_t)$ to have width five in an effort to ensure that this region contains only myocardial muscle, while providing a large enough region to have meaningful statistics. Given the field intensity $f_t^{R(\vec{C}_t)}$ and the observation model of Equation (4.9), log $y_t(z)$ is normally distributed with mean log $f_t^{R(\vec{C}_t)}$ and variance σ_n^2. Consequently, the likelihood term can be written as

$$p(y_t|f_t, \vec{C}_t) \propto \tag{4.10}$$

$$exp(\quad -\int\limits_{z\in R^{in}(\vec{C}_t)} \frac{(\log\, y_t(z) - \log f_t^{R_{in}(\vec{C}_t)})^2}{2\sigma_n^2}\, dz - \int\limits_{z\in R^{out}(\vec{C}_t)} \frac{(\log y_t(z) - \log f_t^{R_{out}(\vec{C}_t)})^2}{2\sigma_n^2}\, dz).$$

■ 4.3.2 Field Prior Term

For the MR data, the mean log intensity inside and that immediately outside the LV boundary are each approximately constant across a cardiac cycle. Given this fact, we can estimate the mean and variance of the log intensity inside (u and σ_u^2, respectively) and that immediately outside the curve (v and σ_v^2, respectively) through a training set (we provide a detailed description of the exact data used in Section 4.5). Having estimated the parameters, we can represent the field prior as

$$p(f_t|\vec{C}_t) \propto exp(-\frac{(\log f_t^{R_{in}(\vec{C}_t)} - u)^2}{2\sigma_u^2})exp(-\frac{(\log f_t^{R_{out}(\vec{C}_t)} - v)^2}{2\sigma_v^2}). \tag{4.11}$$

■ 4.3.3 Curve Prior Term

As described in Section 4.2, the state X_t is a low-dimensional approximate representation of the boundary. The state is used to incorporate information from neighboring frames. Thus, we expect X_t to be a reasonable but inexact representation of the LV boundary $\vec{C}_t$. To allow for variability around the curve represented by X_t, we model the curve prior to be

$$p(\vec{C}_t|X_t) \propto exp(-D^2(\vec{C}_t, X_t)), \tag{4.12}$$

where $D^2(\vec{C}, X_t)$ measures the deviation of $\vec{C}_t$ from X_t by the following formula[2]

$$D^2(\vec{C}_t, X_t) \equiv \int_{\vec{C}_t} d_{X_t}^2(s)ds, \tag{4.13}$$

where $d_{X_t}(s)$ is the distance of point s from the curve X_t.

Note that other distance measures, such as the ones proposed by Kim [101], could be used as well. In addition, if one were interested in a curve length penalty, such a regularization could be incorporated within the curve prior term.

■ 4.3.4 Prediction Term

Next, we want to provide a model for the prediction term. Earlier in this section, we represented $p(X_t|y_{1:t-1})$ by a kernel density estimate with an arbitrary kernel. To obtain a simple form for the integral over X_t (see Equation (4.15)), we approximate this kernel using the Dirac delta function, so

$$p(X_t|y_{1:t-1}) \approx \frac{1}{MN}\sum_{(i,j)} \delta(X_t - x_{t|t-1}^{(i,j)}), \tag{4.14}$$

where MN is the number of samples of $x_{t|t-1}^{(i,j)}$.

Combining the prediction term with the curve prior and integrating over X_t, we can represent the integration over X_t in Equation (4.8)[3] by

$$\int_{X_t} p(\vec{C}_t|X_t)p(X_t|y_{1:t-1})dX_t \propto \frac{1}{MN}\sum_{(i,j)} \exp(-D^2(\vec{C}_t, x_{t|t-1}^{(i,j)})). \tag{4.15}$$

---

[2] This deviation is not a metric because it is not symmetric (in general, $D(X,Y) \neq D(Y,X)$) from its definition, but it provides a reasonable curve prior.

[3] The likelihood and field prior can be moved outside the integrand as they do not depend on X_t.

■ 4.3.5 Curve Evolution

Now, incorporating the specific densities into the functional in Equation (4.8) and defining $F_t^{in}(\vec{C}_t) = \log f_t^{R_{in}(\vec{C}_t)}$ and $F_t^{out}(\vec{C}_t) = \log f_t^{R_{out}(\vec{C}_t)}$, we have

$$E(f_t, \vec{C}_t) = -\log p(y_t|f_t, \vec{C}_t) - \log p(f_t|\vec{C}_t) - \log \int_{X_t} p(\vec{C}_t|X_t) p(X_t|y_{1:t-1}) dX_t$$

$$= (\int\limits_{z \in R_{in}(\vec{C}_t)} \frac{(\log y_t(z) - F_t^{in}(\vec{C}_t))^2}{2\sigma_n^2} dz + \int\limits_{z \in R_{out}(\vec{C}_t)} \frac{(\log y_t(z) - F_t^{out}(\vec{C}_t))^2}{2\sigma_n^2} dz)$$

$$+(\frac{(F_t^{in}(\vec{C}_t) - u)^2}{2\sigma_u^2} + \frac{(F_t^{out}(\vec{C}_t) - v)^2}{2\sigma_v^2}) + \log[\frac{1}{MN} \sum_{(i,j)} exp(-D^2(\vec{C}_t, x_{t|t-1}^{(i,j)}))]. \qquad (4.16)$$

Since the functional $E(f_t, \vec{C}_t)$ depends on two variables, we numerically solve the problem the same way we solved the MMS problem in Section 3.3.2 using coordinate descent. For the f_t step, we fix the boundary $\vec{C}_t$ and compute the field. Using the assumption that f_t is piecewise constant, the f_t which maximizes $E(f_t, \vec{C}_t)$ for a given $\vec{C}_t$ is

$$f_t^{R_{in}(\vec{C}_t)} = exp(\frac{\int_{R_{in}(\vec{C}_t)} \log y_t(z) dz}{\int_{R_{in}(\vec{C}_t)} dz})$$

$$f_t^{R_{out}(\vec{C}_t)} = exp(\frac{\int_{R_{out}(\vec{C}_t)} \log y_t(z) dz}{\int_{R_{out}(\vec{C}_t)} dz}) \qquad (4.17)$$

For a given f_t, we apply curve evolution to find the $\vec{C}_t$ that minimizes $E(f_t, \vec{C}_t)$. To accomplish this, we compute the first variation of $E(f_t, \vec{C}_t)$ with respect to $\vec{C}_t$ and move in that direction. The formula for the first variation is

$$\frac{\partial \vec{C}_t}{\partial \tau}(z) = -[\alpha(F_t^{out}(\vec{C}_t) - F_t^{in}(\vec{C}_t))(2\log y_t(z) - F_t^{out}(\vec{C}_t) - F_t^{in}(\vec{C}_t))$$

$$+2\beta_u \frac{(F_t^{in}(\vec{C}_t) - u)}{A_{in}}(F_t^{in}(\vec{C}_t) - \log y_t(z)) + 2\beta_v \frac{(F_t^{out}(\vec{C}_t) - v)}{A_{out}}(F_t^{out}(\vec{C}_t) - \log y_t(z))$$

$$+\frac{1}{Q} \sum_{(i,j)} exp(-D^2(\vec{C}_t, x_{t|t-1}^{(i,j)}))(\nabla d^2(z, x_{t|t-1}^{(i,j)}) \cdot \vec{N} + d^2(z, x_{t|t-1}^{(i,j)})\kappa(z))]\vec{N}, \qquad (4.18)$$

where $\alpha = \frac{1}{2\sigma_n^2}$, $\beta_u = \frac{1}{2\sigma_u^2}$, $\beta_v = \frac{1}{2\sigma_v^2}$, $Q = \sum_{(i,j)} exp(-D^2(\vec{C}_t, x_{t|t-1}^{(i,j)}))$, A_{in} is the area of R_{in} and A_{out} is the area of R_{out}, $\kappa(z)$ is the curvature of $\vec{C}$ at z, $\vec{N}$ is the unit outward normal of $\vec{C}$ at z, and τ is an iteration-time parameter used during the curve evolution process. The computation of the first variation relies on four separate derivations of curve flows [29,30,101,197].

4.3.6 Using Weighted Samples to Approximate the Posterior

Now, using Equation (4.5), we obtain a representation for $p(X_t|y_{1:t})$ by taking the equally-weighted samples $x_{t|t-1}^{(i,j)}$ which represent $p(X_t|y_{1:t-1})$ and re-weighting them by the likelihood, field prior, and curve prior terms so that

$$p(X_t|y_{1:t}) = \sum_{i=1}^{N}\sum_{j=1}^{M} w_t^{(i,j)} k(X_t; x_{t|t-1}^{(i,j)}), \tag{4.19}$$

where

$$w_t^{(i,j)} = \frac{1}{Z} p(y_t|f_t^*, \vec{C}_t^*) p(f_t^*|\vec{C}_t^*) p(\vec{C}_t^*|X_t = x_{t|t-1}^{(i,j)}), \tag{4.20}$$

with Z chosen so that the weights sum to 1. Finally, we can resample this density to N equally-weighted samples, setting the quantity of samples points to a manageable number (the resampling effectively discards samples with very small weights and creates multiple samples near samples having large weights). Thus, from equally-weighted samples representing the posterior at $t-1$, we have arrived at equally-weighted samples for the posterior at t, completing the algorithm for recursion. Through this discussion, however, we have assumed that $p(X_t|X_{t-1})$ is known. In Section 4.4, we explain how $p(X_t|X_{t-1})$ is estimated by learning the system dynamics. Before we examine the dynamics, we discuss an alternate approach to the recursive estimation problem when the posterior $p(X_t|y_{1:t})$ can be reasonably approximated as a Gaussian.

4.3.7 Gaussian Posterior

In problems in which the assumption of a unimodal posterior is reasonable, a simplification to the sample-based method can be used as an approximation. In particular, instead of representing the posterior with the sample points and weights as described in Section 4.3.6, we determine the mean and covariance of X_t in order to create a Gaussian posterior. We then take samples from this distribution.

In Section 4.3.5, we obtained the curve $\vec{C}_t^*$ which maximizes $E(f_t, \vec{C}_t)$ given a set of samples $x_{t|t-1}^{(i,j)}$ representing the prediction. Assuming that $p(\vec{C}_t|X_t)$ is symmetric about X_t, as is the case for our curve prior of Equation (4.12), the projection of $\vec{C}_t^*$ onto the space of curves spanned by X_t is an approximation to the MAP estimate of X_t. Using the Gaussian assumption, the MAP estimate is the mean parameter used for our posterior. Since there is no easy method to determine the posterior covariance, we approximate it to be a diagonal matrix with individual variances determined empirically from the shape variability in the training data. So, using this mean and covariance approximation, we can represent a Gaussian posterior and easily generate samples from this distribution.

4.4 Learning the Dynamics

A number of approaches can be taken to learn the dynamics of an evolving system. First, we can consider purely physics-based models to constrain and explain the dynamics of a given problem [116–118, 127, 157, 162]. The drawback is that systems that accurately model physics may require high dimensional states and/or a complex set of differential equations that model the interaction between adjacent masses in the system. Alternatively, we may attempt to learn the dynamical model statistically where the estimated densities can either be parametric or non-parametric. For a parametric model, the challenge is to find a model that matches the problem structure well and captures the statistical variability inherent in the problem. For richer modeling capacity, one can turn to non-parametric models, which can be computationally difficult. In Section 4.4.2, we propose a non-parametric, yet computationally tractable approach to learning the dynamics of LV boundaries. Before discussing this method, we first provide a description of the system state X_t.

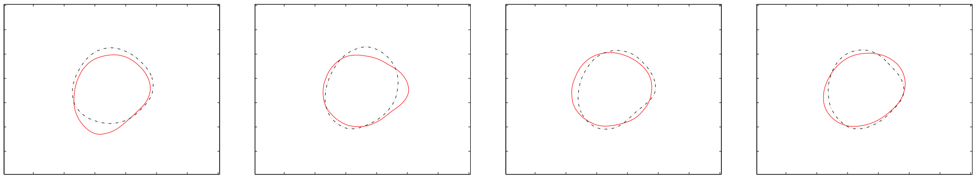
4.4.1 Implicit Parametric Shape Model and State Representation

The set of LV boundaries have different internal areas and different shapes across a cardiac cycle and between patients. We want to represent these boundaries in a simple, low-dimensional, yet accurate, manner. To accomplish this, we use principal components analysis (PCA) on the signed distance functions (SDFs) to obtain a basis for the shapes [113]. We then represent the SDF of each shape by a linear combination of the basis elements. With this representation, the estimation of dynamics reduces to learning the time evolution of the coefficients of the basis elements.

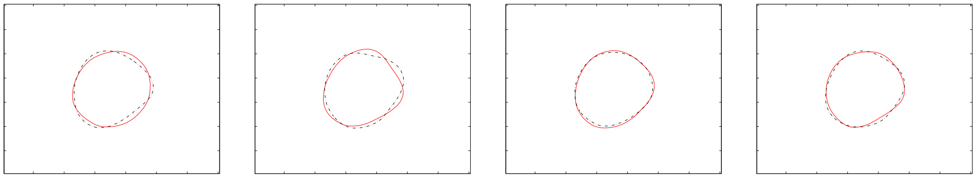
Starting with a training set of manually segmented and registered data, we determine the area of each LV. Normalizing with respect to area, we create signed distance functions whose zero level sets are the shapes [159]. Following Leventon's PCA modeling of shapes [113], we obtain a mean shape $\bar{\psi}$ and the primary modes of variability ψ_i (for i=1,2, ..., K, where K is the number of shapes in the dataset) across the entire training set. In effect, we use a single basis to represent the shapes across the entire cardiac cycle. Figure 4.3 shows the eight primary modes of variability from the training set used in the experimental results presented in Section 4.5. For a given signed distance function ψ in the training set,

$$\psi = \bar{\psi} + \sum_{i=1}^{K} \alpha^{(i)} \psi_i, \tag{4.21}$$

where $\alpha^{(i)}$'s are a set of coefficients. It is known that for shapes which do not vary greatly, the primary few modes of variability can explain the majority of the variability of the data. In our training set, the first eight modes explain 97% of the variability in our specific training set of data. Thus, we approximately represent each ψ by the eight element vector $\vec{\alpha} = [\alpha^{(1)}; \alpha^{(2)}; \ldots; \alpha^{(8)}]^T$. By using PCA, a given curve (LV segmentation) can be approximately represented by a vector containing its area A and



Mode 1 Mode 2 Mode 3 Mode 4



Mode 5 Mode 6 Mode 7 Mode 8

Figure 4.3. Illustration of LV shape variability. $\bar{\psi} \pm \sigma_i \psi_i$ for the first eight primary modes of variability ($i = 1, 2, \cdots, 8$). Solid curve represents $\bar{\psi} + \sigma_i \psi_i$ while dashed represents $\bar{\psi} - \sigma_i \psi_i$.

$\vec{\alpha}$. Thus, assuming a first-order system for all experiments in the thesis, the state is $X_t = [A_t; \vec{\alpha}_t^T]^T$, a nine-dimensional vector.

We conclude by noting that the use of PCA to determine the representation of the state is just one choice among many to represent the shape of the LV. Jaggi *et al.* [83], for instance, use an ellipsoid representation for the LV. For two-dimensional slices, the parameters they use to represent the boundary are the two semiaxis lengths and an angle of rotation. Our formulation can be used to learn the dynamics given any type of representation, as the description of the Bayesian formulation in Section 4.2.2 does not require a specific form for the state X_t.

■ 4.4.2 A Maximally-Informative Statistic

We propose learning the dynamics from a training set of data based on a technique [81] which produces a non-parametric density estimate of $p(X_t|X_{t-1})$. This estimate is obtained by using an information-theoretic criterion to maximize the predictive power of the observations. To minimize the computational complexity, we consider only the portion of the state X_{t-1} that is statistically pertinent to the prediction of X_t. Thus, we introduce a function $q_{t-1}(X_{t-1})$ which seeks to reduce dimensionality yet capture all of the information in X_{t-1} that relates to X_t. This is achieved exactly only when $I(X_t; X_{t-1}) = I(X_t; q_{t-1}(X_{t-1}))$, where $I(X_t; X_{t-1})$ is the mutual information between X_t and X_{t-1}. From this, we can create an estimate of $p(X_t|q_{t-1}(X_{t-1}))$ as an equally-informative yet simpler representation of $p(X_t|X_{t-1})$.

Practically, however, information is lost when we introduce the function q_{t-1}. As a result, we choose the parameters of q_{t-1} such that $I(X_t; q_{t-1}(X_{t-1}))$ is maximized but do not require equality to $I(X_t; X_{t-1})$ (by the data processing inequality [38], $I(X_t; q_{t-1}(X_{t-1})) \leq I(X_t; X_{t-1})$, with equality when there is no loss of information). This makes $q_{t-1}(X_{t-1})$ a maximally-informative statistic instead of a sufficient statistic.

Since the mutual information between X_t and $q_{t-1}(X_{t-1})$ is defined as (with $H(x)$ being the entropy of random variable x)

$$I(X_t; q_{t-1}(X_{t-1})) = H(X_t) + H(q_{t-1}(X_{t-1})) - H(X_t, q_{t-1}(X_{t-1})), \tag{4.22}$$

we must first describe how we estimate entropies in order to compute the gradient of $I(X_t; q_{t-1}(X_{t-1}))$. We estimate the entropy using leave one out resubstitution [10]. In particular, given N equally-weighted samples $x_t^{(i)}$ of $p(X_t|y_{1:t})$, suppose we approximate the posterior using a kernel density estimate with Gaussian kernels (define $k(X; x_t^{(i)}, \sigma^2)$ to be a Gaussian kernel with mean $x_t^{(i)}$ and variance σ^2, where σ^2 is determined by a method such as that described in [77]). Then, the entropy estimate of $q_t(X_t)$ is

$$H(q_t(X_t)) = -\frac{1}{N}\sum_{i=1}^{N} \log(\frac{1}{N-1}\sum_{j\neq i} k(q_t(x_t^{(i)}); q_t(x_t^{(j)}), \sigma^2)). \tag{4.23}$$

Taking the derivative with respect to any parameter a of the function q_t yields

$$\frac{\partial H(q_t(X_t))}{\partial a} = -\frac{1}{N}\sum_{i=1}^{N}[\frac{1}{\sum_{j\neq i} k(q_t(x_t^{(i)}); q_t(x_t^{(j)}), \sigma^2)} \cdot$$

$$\sum_{j\neq i} -\frac{q_t(x_t^{(i)}) - q_t(x_t^{(j)})}{\sigma^2} k(q_t(x_t^{(i)}); q_t(x_t^{(j)}, \sigma^2))(\frac{\partial q_t(x_t^{(i)})}{\partial a} - \frac{\partial q_t(x_t^{(j)})}{\partial a})]. \tag{4.24}$$

By applying Equation (4.24) to the second term of Equation (4.22) (and using a similar derivation to find the derivative of the joint entropy of the third term), we can determine the gradient of $I(X_t; q_{t-1}(X_{t-1}))$. At each iteration, we proceed to move in the direction of the gradient, continuing until convergence.

■ 4.4.3 Learning the Forward Density through Training

In the discussion thus far, we have discussed finding a time-varying q_t to determine the forward density. In order to have an accurate estimate of the parameters of q_t, there must be sufficient training data. Practically, we may not have enough data to learn a different q_t for each t. For our particular training set, we learn the dynamics separately in the two distinct phases of the cardiac cycle (q_S for the systolic phase, when oxygenated blood leaves the LV, and q_D is for the diastolic phase, when the LV fills itself with blood) and then use these two dynamics in the corresponding regions and a third (a mixture of q_S and q_D) in the region where we are uncertain (in the test data) which phase we are in.

The reason for the inclusion of the transition region is simply that the location of end systole (the point of minimum area, when the cardiac motion changes from the systolic phase to the diastolic phase) varies from patient to patient. To illustrate, Figure 4.4 plots the cross-sectional area of the LV across the cardiac cycle for each

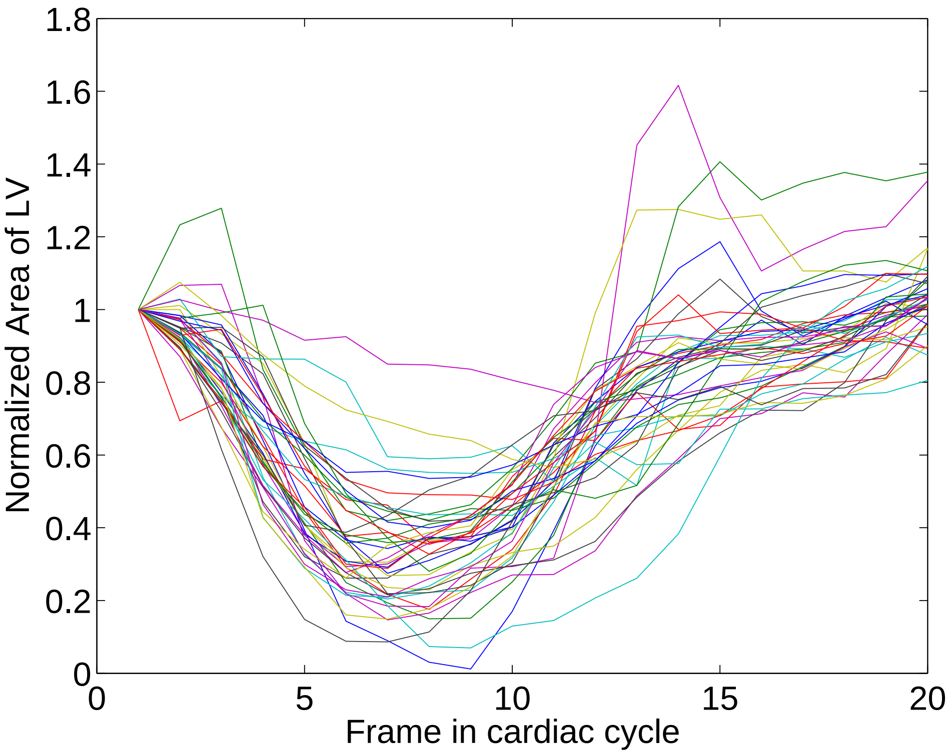


Figure 4.4. Plot of normalized area (normalized by the area at end diastole ($t = 1$)) as a function of time for the 42 training cardiac cycles. In addition to the wide range of area variability across each cardiac cycle, note that the location of end systole, the time in the cardiac cycle when the LV area is a minimum, occurs at different times for different patients between frames 6 and 12.

Frame	Frequency	Percentage
6	6	14.3%
7	16	38.1%
8	12	28.6%
9	3	7.1%
10	2	4.8%
11	2	4.8%
12	1	2.4%

Table 4.1. Frame number for end systole from the 42 cardiac cycles used in the training set.

of the cardiac cycles in the training set (each cardiac cycle consists of 20 frames). Furthermore, Table 4.1 lists the frequency for the location of end systole across our training set. Based on these data, we assume that for transitions between frames 1 to 6, we are in the systolic phase. Similarly, for the transitions between frames 12 to 20, we are in the diastolic phase. For the frames in between, we apply a combination of the two dynamics according to the mixture probability determined by the data in Table 4.1.[4] For instance, 14.3% of the training examples reach end systole at frame 6 and transition to the diastolic phase from that frame onward. So, in our test set, 14.3% of the samples of $p(X_7|q_6(X_6))$ come from the forward density using q_D while 85.7% use the forward density of q_S (although we do not do this, one could also attempt to estimate the probability of being in the systolic (or diastolic) phase through the course of recursively estimating X_t, possibly providing more accurate mixture percentages than historical data). To demonstrate the usefulness of the mixture model, Figure 4.5(a) shows predictions obtained for the eighth frame in a test sequence using a combination of samples from the two forward densities. Note the bimodal nature of the predictions due to sampling from the two densities.

Empirically, we know that this particular test sequence is in the systolic phase between frames 7 and 8. From the experiment, we observe that the black curves are samples from $p(X_8|q_D(X_7))$ while the green curves are samples from $p(X_8|q_S(X_7))$. As expected, the most likely estimates (shown in green in Figure 4.5(b) with the truth shown by the red dash-dotted curve) based on the predicted samples of Figure 4.5(a) all come from samples of $p(X_8|q_S(X_7))$, the dynamic from the systolic phase. By providing predictions from both phases, we allow the likelihood and field prior terms to properly weight the more likely samples, thus resulting in reasonable estimates without having to make a strict decision on which phase we are in.

For this transition region, we could alternatively learn a third mixture model q_{MIX} from the samples between frames 6 and 12 in the training data. We would then apply q_S between frame 1 and 6, q_{MIX} between frames 6 and 12, and q_D between frames 12

---

[4] We acknowledge that the small empirical dataset serves to provide only a coarse estimate for the distribution of the location of end systole.

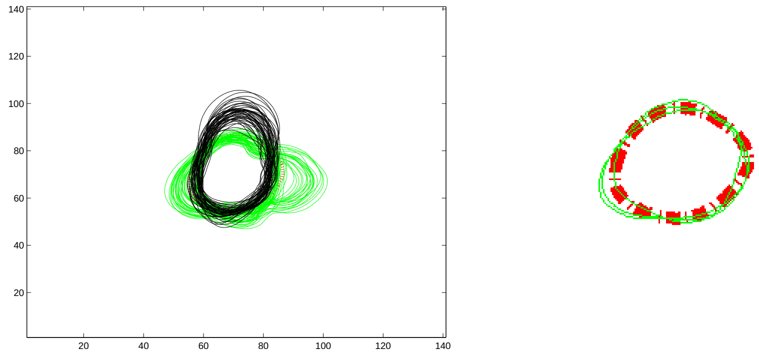

(a) (b)

Figure 4.5. (a) Predictions for frame 8 in a test sequence. Samples in black are made from p_D while those in green are made from p_S. (b) Most likely samples upon reweighting after incorporating the observation at frame 8 superimposed upon the truth (red dash-dotted curve). In this example, the predictions are bimodal but a reasonable set of curve estimates are achieved after reweighting the samples with the likelihood and priors.

and 20 on the test data. The results were found to be similar using either approach.

4.4.4 Linear Assumption for Statistic

In the experiments, we assume that q_t is a linear function. We can write $q_{t-1}(X_{t-1}) = Q_{t-1}X_{t-1}$, where

$$Q_{t-1}X_{t-1} = \begin{bmatrix} Q_A & Q_{A\vec{\alpha}} \\ Q_{\vec{\alpha}A} & Q_{\vec{\alpha}} \end{bmatrix}_{t-1} X_{t-1}. \tag{4.25}$$

With this assumption, learning the dynamics reduces to finding the parameters of Q_t. It should be noted that the linearity assumption for q_t does not make the dynamics linear. The forward density $p(X_t|q_{t-1}(X_{t-1}))$ is still a non-parametric density. For Q_t, we additionally assume that there is no interaction between the area and shape of the object. So, $Q_{A\vec{\alpha}} = 0$ and $Q_{\vec{\alpha}A} = 0$ in Equation (4.25).[5] Thus, the learning of the area dynamics can be separated from that of the shape dynamics.

As a trade-off between computational complexity and being maximally-informative, we choose the statistic Q_t to be three-dimensional (two dimensional for shape and one for area). Thus, Q_A is scalar and $Q_{\vec{\alpha}}$ is 2×8. From the training data, we learn the 17 parameters (1 for Q_A, 16 for $Q_{\vec{\alpha}}$) for each dynamic function Q. We choose a one-dimensional statistic for the area. For the shape statistic, we empirically determined that going from one to two dimensions led to a big improvement, while subsequent increases in dimension did not lead to substantial gains. Thus, we settled on a 2-D

---

[5] Given a rich enough training data set, one can allow interactions between the area and shape when learning Q_t, but for now, we assume a block diagonal Q_t matrix.

statistic for the shape.

■ 4.5 Experimental Results

We apply the proposed technique on 2-D mid-ventricular slices of data, but we also note that we can apply the method easily to 3-D data. The dataset we use contains twenty-frame time sequences of breath-hold cardiac MR images, each representing a single cardiac cycle with the initial frame gated (synchronized) with an electrocardiogram (EKG) signal. We do not consider arrhythmias because only patients having sustained and hemodynamically-stable arrhythmias, a rare situation, can be practically imaged and analyzed. Anonymized data sets were obtained from the Cardiovascular MR-CT Program at Massachusetts General Hospital. Our training set consists of 42 cardiac cycles of 20 frames each for a total of 840 images. We perform the experiments on eight different test sets (distinct from the training sets), with representative results of full cardiac cycles shown. Note that all of the training and test sets come from healthy patients.

In Section 4.5.1, we explain the metrics that we use to determine the accuracy of the segmentations. In Section 4.5.2, we examine the filtering problem using a sample-based approach without any assumptions for the posterior. In the analysis, we show that a time-varying model produces estimates that are reasonably accurate and also better than those obtained using a time-invariant model, as one may expect. In Section 4.5.3, we examine results assuming a Gaussian posterior. We show that for this particular problem, such an assumption is reasonable. In Section 4.5.4, we compare and contrast the results with and without the Gaussian posterior assumption. The results of our proposed approach are then compared with two static segmentation methods in Section 4.5.5 to illustrate the benefit of using information from previous frames.

■ 4.5.1 Metrics

Quantitatively, we measure accuracy of a segmentation by computing the dice coefficient [49] between this segmentation and a manual segmentation, which we treat as the ground truth. This coefficient is commonly used for evaluation of segmentations in medical imaging. Introduced in the previous chapter (but repeated here for convenience), the dice coefficient measures the closeness of two regions. If boundary $\vec{C}_1$ encloses a region R_1 and boundary $\vec{C}_2$ encloses region R_2, we define the dice measure between the two regions as

$$Dice(R_1, R_2) \equiv \frac{2A(R_1 \cap R_2)}{A(R_1) + A(R_2)}, \tag{4.26}$$

where $R_i \cap R_j$ represents the intersection of regions R_i and R_j and $A(R_i)$ is the area of region R_i. The dice measure evaluates to 1 when two regions are a perfect match and 0 when the regions are a complete mismatch.

Comparison of LV Boundary Estimates

To determine the accuracy of a set of boundary estimates (e.g. across a set of cardiac cycles), we compute the dice measure between each of the estimates and the corresponding ground truth and then average over all of the available dice coefficients. For instance, if we have 4 cardiac cycles of segmentations to examine, we compute the dice measure for each of the 80 frames (recall, each cardiac cycle has 20 frames) and then determine the average of these eighty dice coefficients. Henceforth, we refer to this average as the dice boundary coefficient.

Comparison of Samples from Posterior Estimate of the State

In addition to examining the accuracy of the boundary estimates, it is instructive to look at the quality of the samples of the posterior estimates. To quantitatively examine samples, we can determine the dice coefficient between each sample $x_t^{(i)}$ from a posterior density $p(X_t|y_{1:t})$ and ground truth. Since none of these samples are expected to be as accurate as the LV boundary estimate $\vec{C}_t^*$, we expect the average dice coefficient from the samples to be smaller than the dice coefficient for the LV boundary estimate at the same frame. To determine the accuracy of samples, we compute the average across the samples in all of the frames analyzed. For example, if we have 4 cardiac cycles of posterior estimates with each frame having 50 samples, then we compute the dice measure for all 4000 samples and then determine the average of these dice coefficients. Henceforth, we refer to this average as the dice sample coefficient.

■ 4.5.2 Estimation with a Non-Parametric Posterior Density

In this section, we present results using the sample-based approach described in Section 4.3.6 for LV boundary estimation. We examine the results of the filtering problem using a piecewise time-invariant model (different dynamics for the different phases of the cardiac cycle) as described in Section 4.4.3 which we henceforth refer to as the PTI dynamic model. The reasonable accuracy of the results suggests that the choice of model sufficiently trades-off the need for time-varying dynamics and the richness of available data.

For initialization, we want the posterior at $t = 0$ (for use in the prediction of X_1), but we practically do not have any observation or information about the frame preceding the first frame at $t = 1$. To approximate the initial posterior, we assume that the cardiac cycle is quasi-periodic (the periodicity will be exploited more systematically in the next chapter, but for now we use this to obtain an initialization), and use a given segmentation of the last frame ($t = 20$) to approximate $p(X_0)$. This segmentation can be approximately obtained using an automated static method, an expert hand-segmentation, or a segmentation of a spatially-neighboring 2-D slice from the same time. Starting with $p(X_0|y_0) = p(X_0)$ (since we don't have an observation from $t = 0$), we recursively estimate the posterior for each subsequent frame in the cardiac cycle.

Results - High SNR

Using the high SNR bright blood MR measurements, Figure 4.6 shows the 50 most likely samples of the posterior $p(X_t|y_{1:t})$ using the PTI dynamic model. To illustrate that a time-invariant model is insufficient, Figure 4.7 shows the 50 most likely samples using a time-invariant model. In both cases, the posteriors display some variability around the boundary. However, note that results from the PTI dynamic model show samples of the posterior that more accurately estimate the true area of the LV around end systole (see frames 8 through 11) whereas the time-invariant model seems to over-estimate the area. Figure 4.8 shows the estimate of $\vec{C}_t^*$ obtained using curve evolution[6] with predictions from the PTI model while Figure 4.9 shows the estimates of $\vec{C}_t^*$ obtained using curve evolution with predictions from the time-invariant model. As is the case in the samples of the posterior, the decrease in area in the end systolic region (frames 7 to 11) is not captured well by the time-invariant dynamic model.

For the PTI model, the dice boundary coefficient is 0.8654. For the time-invariant dynamics, the dice boundary coefficient is 0.8210.

Results - Low SNR

In order to test the algorithm on different quality data despite having only one type of observation (bright blood MR), we synthetically create a noisy test set by adding independent, lognormal multiplicative noise to the MR images. Henceforth, we refer to this as low SNR data. Figure 4.10 compares the mostly likely samples of the posterior using the PTI dynamic model (yellow curves) on the test set with samples obtained from the time-invariant model (red curves). We observe that the areas of the samples again appear to be more accurate for the PTI model near end systole. This fact is perhaps more visible when comparing the estimates of the LV boundary using the PTI dynamic model with those using time-invariant dynamics (Figure 4.12).

Although the results using low SNR data yield less accurate segmentations than those for the high SNR data, we see a similar behavior with the LV boundary estimates when comparing the PTI and time-invariant models. Namely, the time-invariant model tends to over-estimate the area of the LV boundary around end systole while the PTI model seems to provide a reasonable estimate. Quantitatively, the dice boundary coefficient is 0.8123. However, the dice boundary coefficient from the time-invariant model is 0.7996.

This result is perhaps not surprising, as a time-invariant model cannot accurately capture both the contraction and expansion of the LV across a cardiac cycle. The results also indicate that dividing the dynamic model into cardiac phases (systolic and diastolic) appears sufficient to obtain reasonable boundary estimates.

---

[6] As described in Section 4.2.2, we find the $\vec{C}_t^*$ that maximizes Equation (4.4) over all samples $x_t^{(i,j)}$.

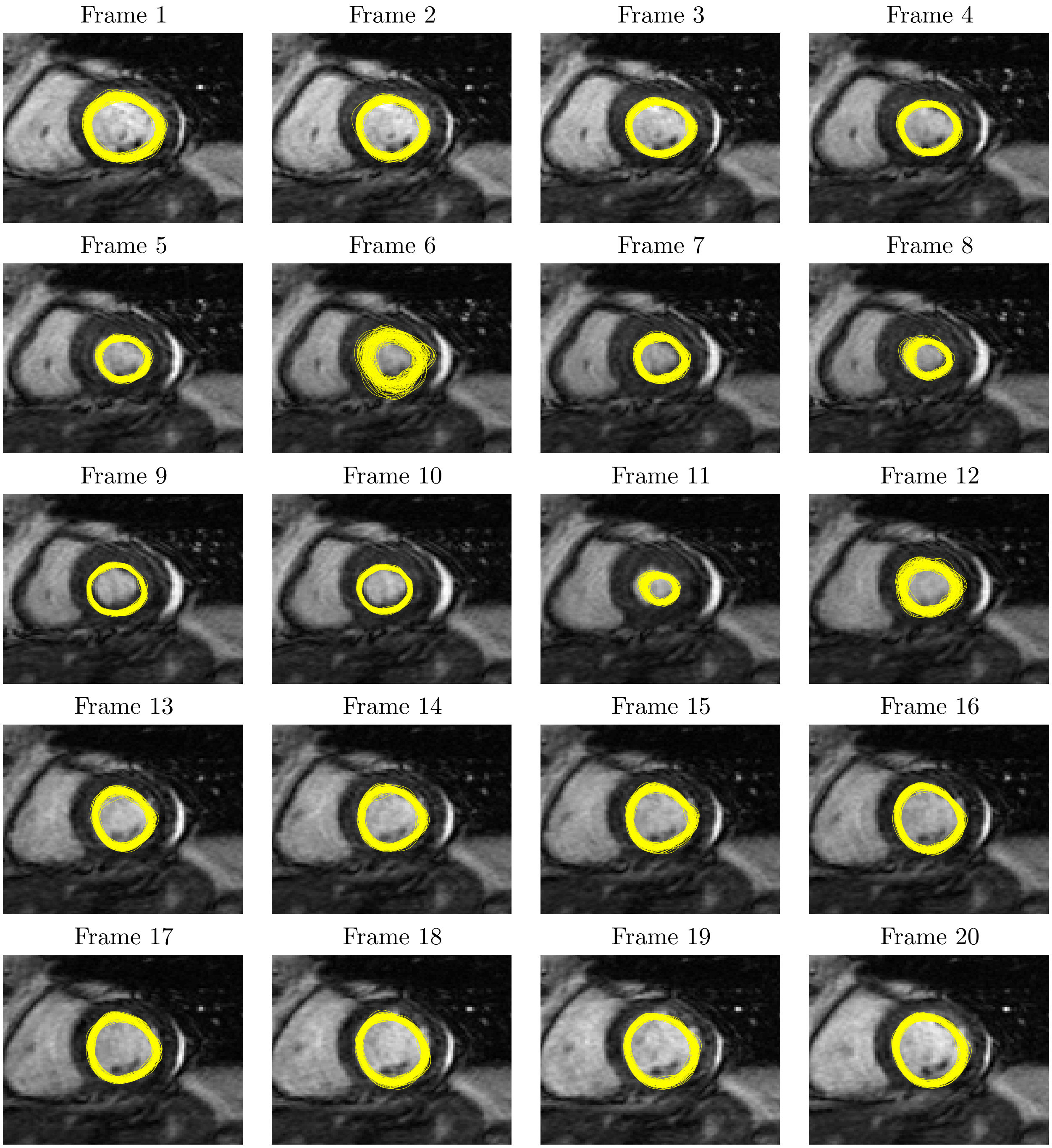


Figure 4.6. The 50 most likely samples of the posterior across a cardiac cycle (20 frames) using the PTI dynamic model overlaid on the observation. The recursive estimation was performed using the general particle filtering approach.

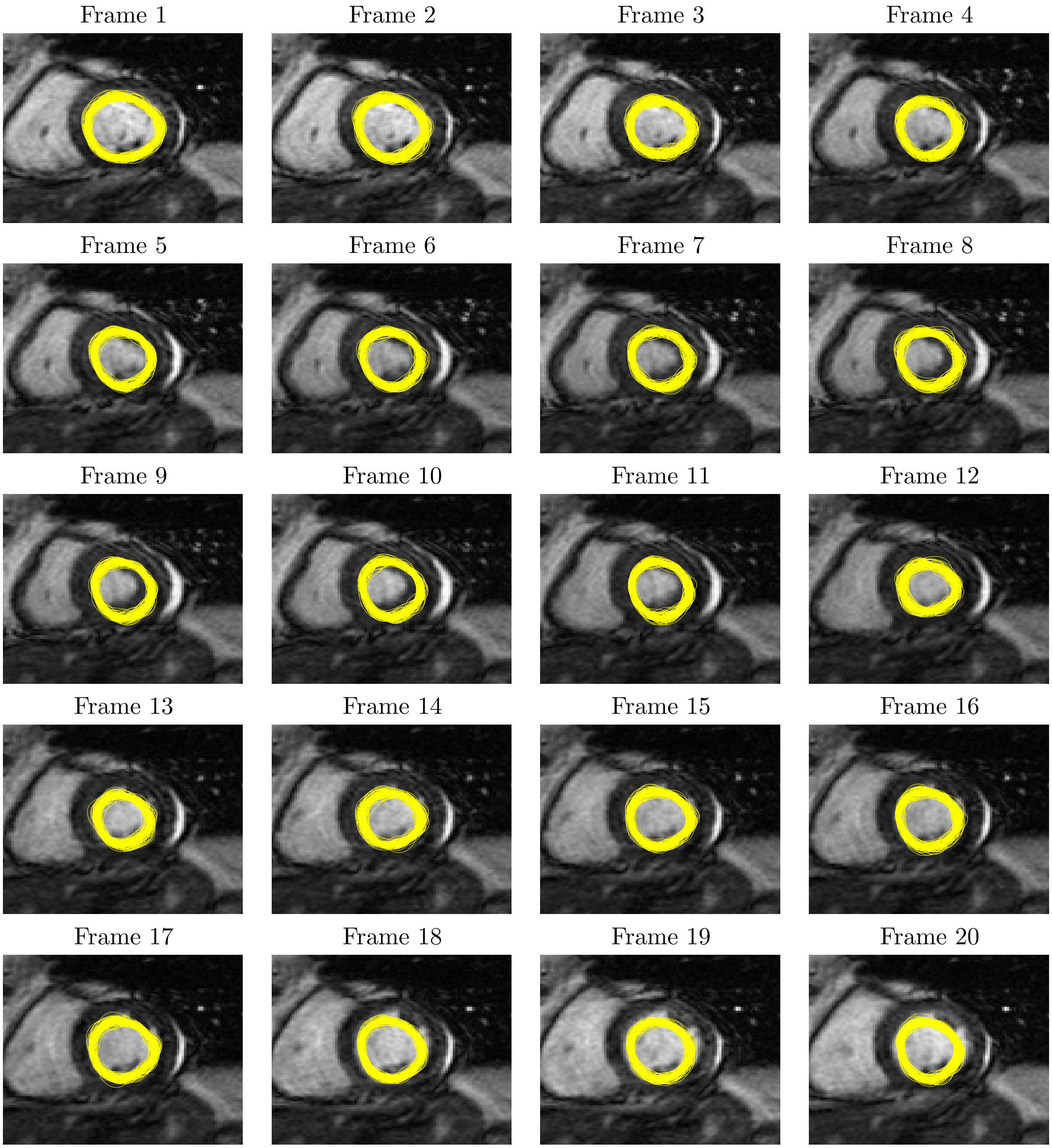
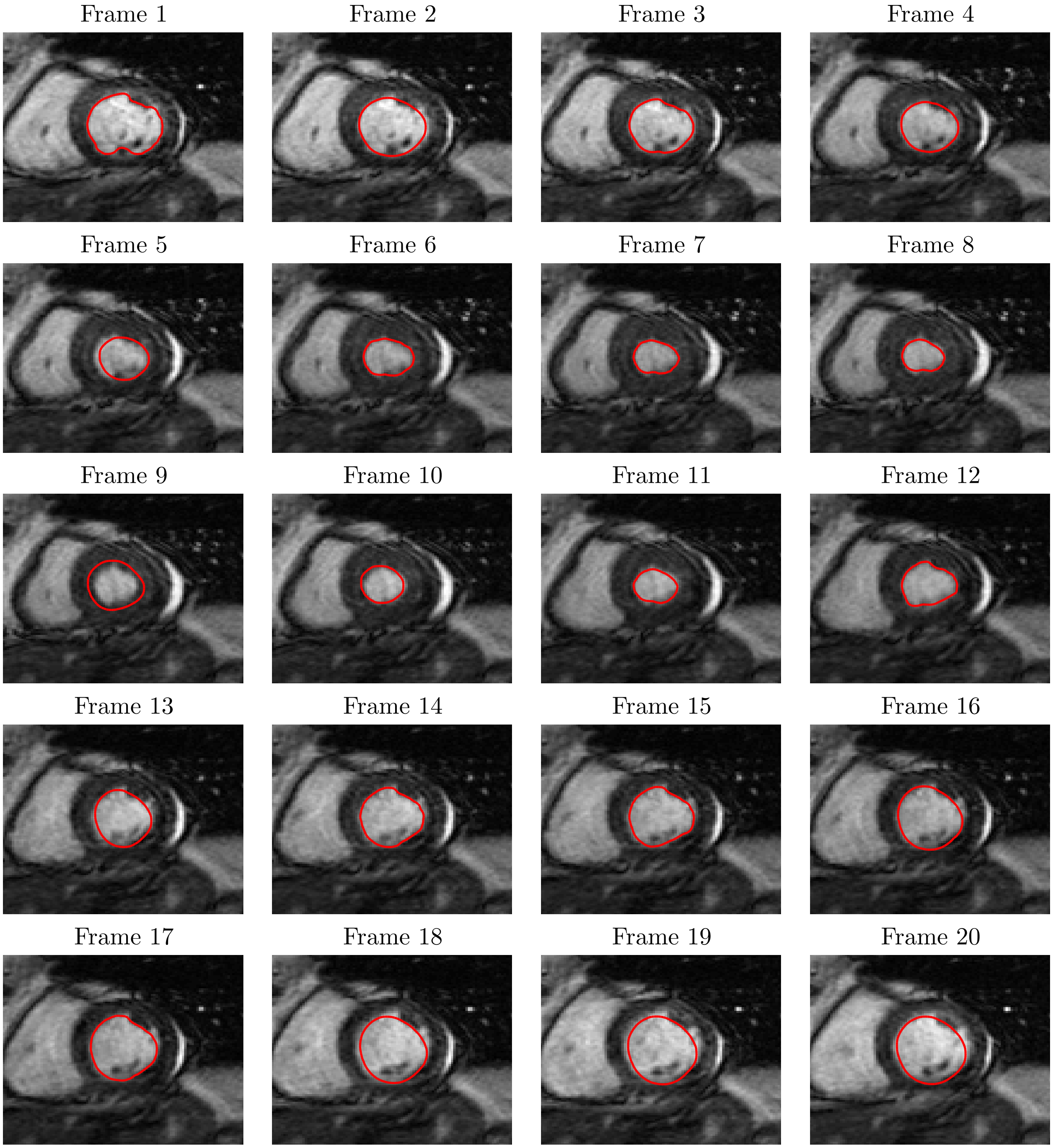


Figure 4.7. The 50 most likely samples of the posterior across a cardiac cycle using only a time-invariant dynamic model overlaid on the observation. Note how the estimates for frames 8 through 11, inclusive, tend to overestimate the LV boundary (compare with the same frames in Figure 4.6). The recursive estimation was performed using the general particle filtering approach.



Frame 1 Frame 2 Frame 3 Frame 4

Frame 5 Frame 6 Frame 7 Frame 8

Frame 9 Frame 10 Frame 11 Frame 12

Frame 13 Frame 14 Frame 15 Frame 16

Frame 17 Frame 18 Frame 19 Frame 20

Figure 4.8. Illustration of LV segmentations superimposed on the MR data. Estimates of the LV boundary obtained using the PTI dynamic model.

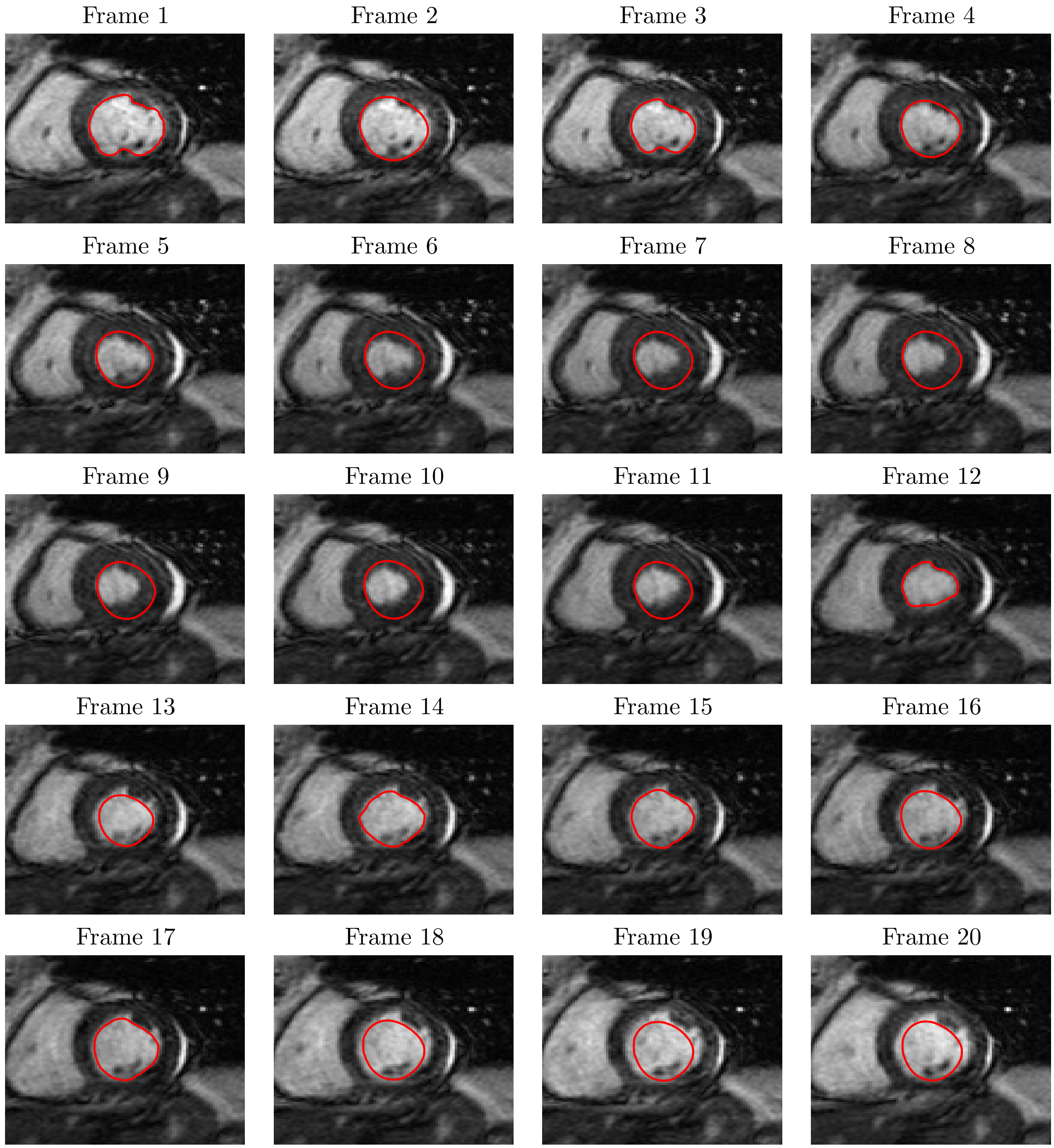
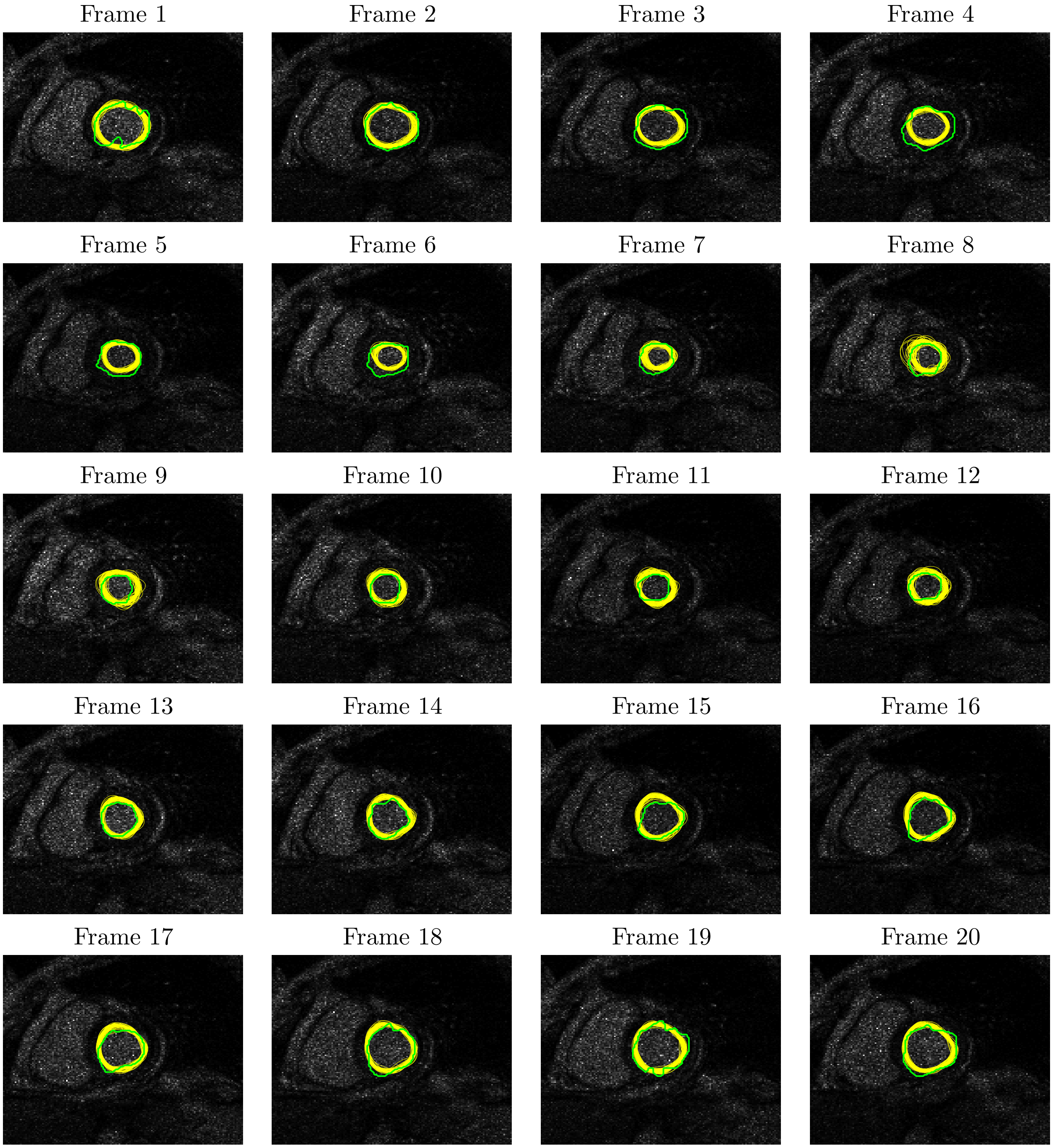
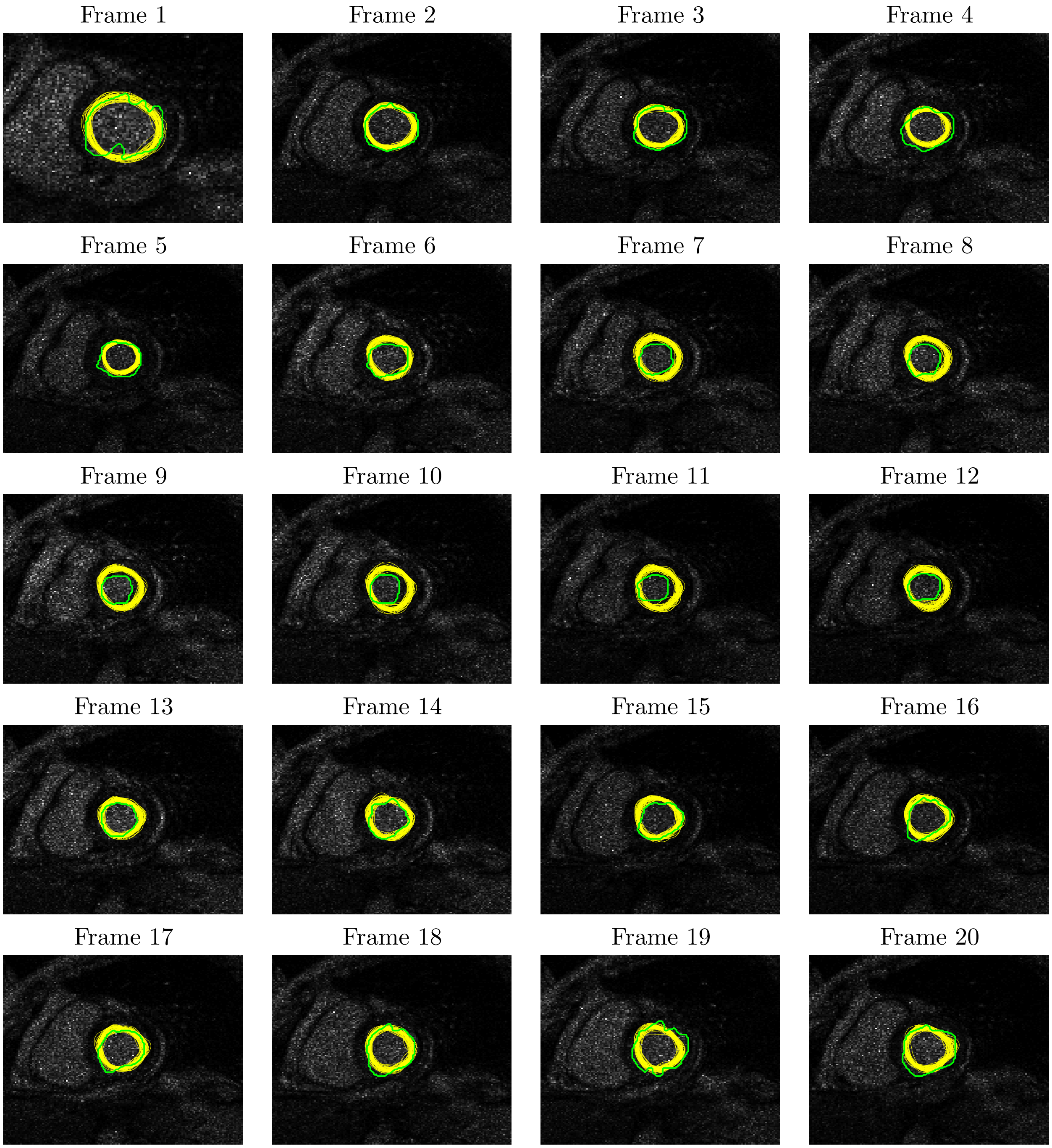

Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 4.9. Illustration of LV segmentations superimposed on MR data. Estimates of the LV boundary obtained using the time-invariant dynamic model.



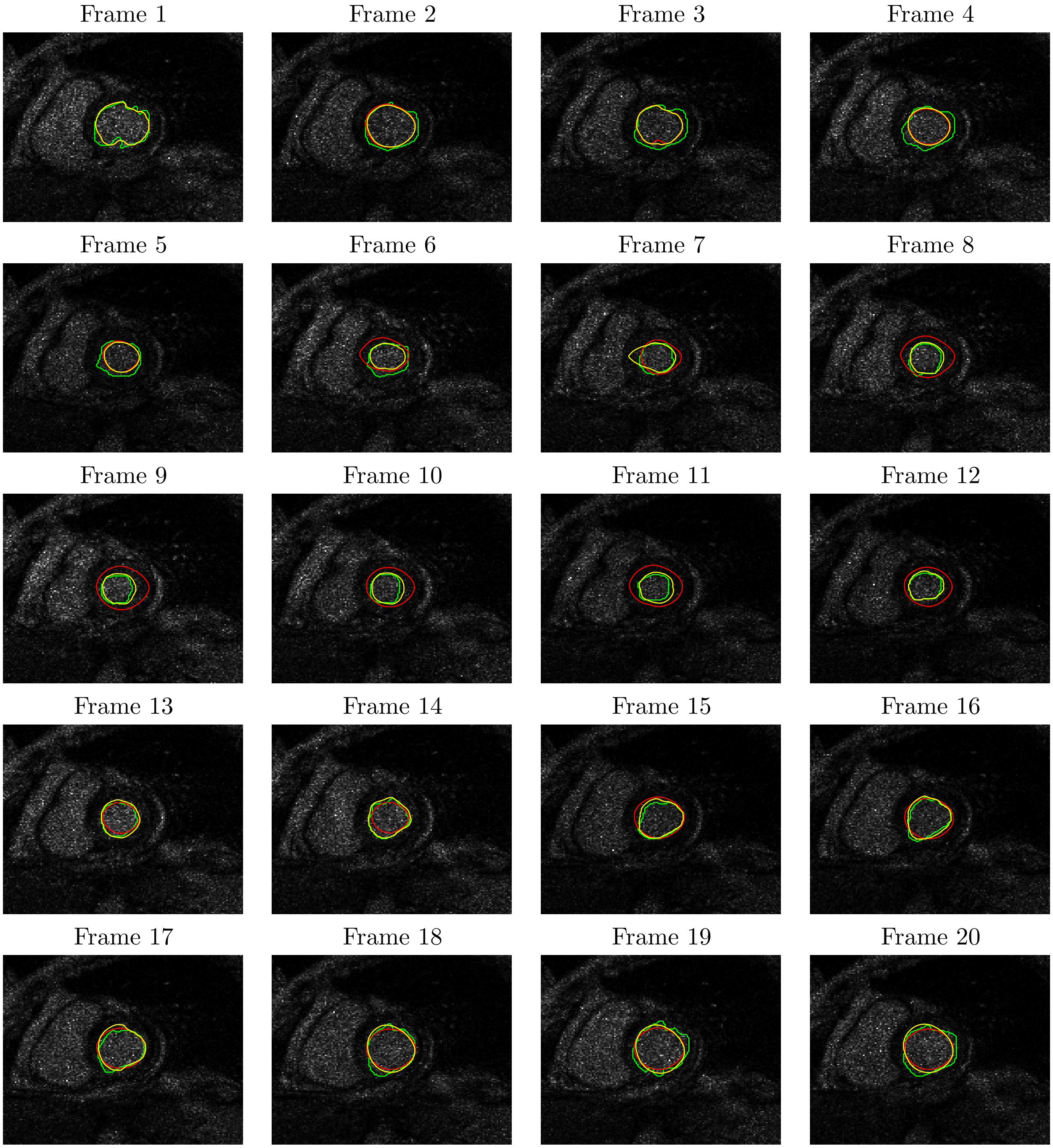
Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 4.10. The 50 most likely samples (yellow) of the posterior across a cardiac cycle (20 frames) using the PTI dynamic model overlaid on the low SNR observation. The recursive estimation was performed using the general particle filtering approach. Since the LV boundary is not as obvious in low SNR data, the manually segmented truth is overlaid (in green).



Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 4.11. The 50 most likely samples (yellow) of the posterior across a cardiac cycle using a time-invariant dynamic model overlaid on the low SNR observation. The recursive estimation was performed using the general particle filtering approach. Since the LV boundary is not as obvious in low SNR data, the manually segmented truth is overlaid (in green). As with the high SNR samples (Figures 4.7 and 4.6), the samples using a time-invariant model again tend to overestimate the LV boundary (e.g. in frames 8 through 12, inclusive, almost all of the samples curves are larger than the manual segmentation).



Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 4.12. Illustration of LV segmentations superimposed on the low SNR data. Segmentations in yellow are obtained by curve evolution using the posterior shown in Figure 4.10 (PTI dynamic model). Those in red are obtained using the posterior shown in Figure 4.11 (time-invariant model). The manually segmented truth is shown in green. Note how the time-invariant model overestimates the LV boundary in frames 8 through 12, inclusive.

4.5.3 Estimation with a Gaussian Posterior Assumption

We now consider the special case of a Gaussian posterior. The hypothesis is that for certain problems, such as LV boundary estimation, where the distribution of the state may be unimodal, we may simplify the recursive estimation process (by constraining the posterior) without sacrificing accuracy in the results. For all future experiments, we use the PTI dynamic model (q_S in the systolic phase, q_D in the diastolic phase, and a mixture of the two in between) and a twenty-frame test sequence for $t = 1, 2, ..., 20$. The procedure for using the test data is similar to that in the previous experiments. We again initialize the problem by approximating $p(X_0)$ from segmentation information of frame 20 because of the quasi-periodicity assumption. Since the posterior is Gaussian, we need to additionally approximate the covariance as described in Section 4.3.7 to obtain our initial Gaussian distribution $p(X_0)$. We recursively estimate the posterior for each subsequent frame assuming a Gaussian posterior.

Results - High SNR

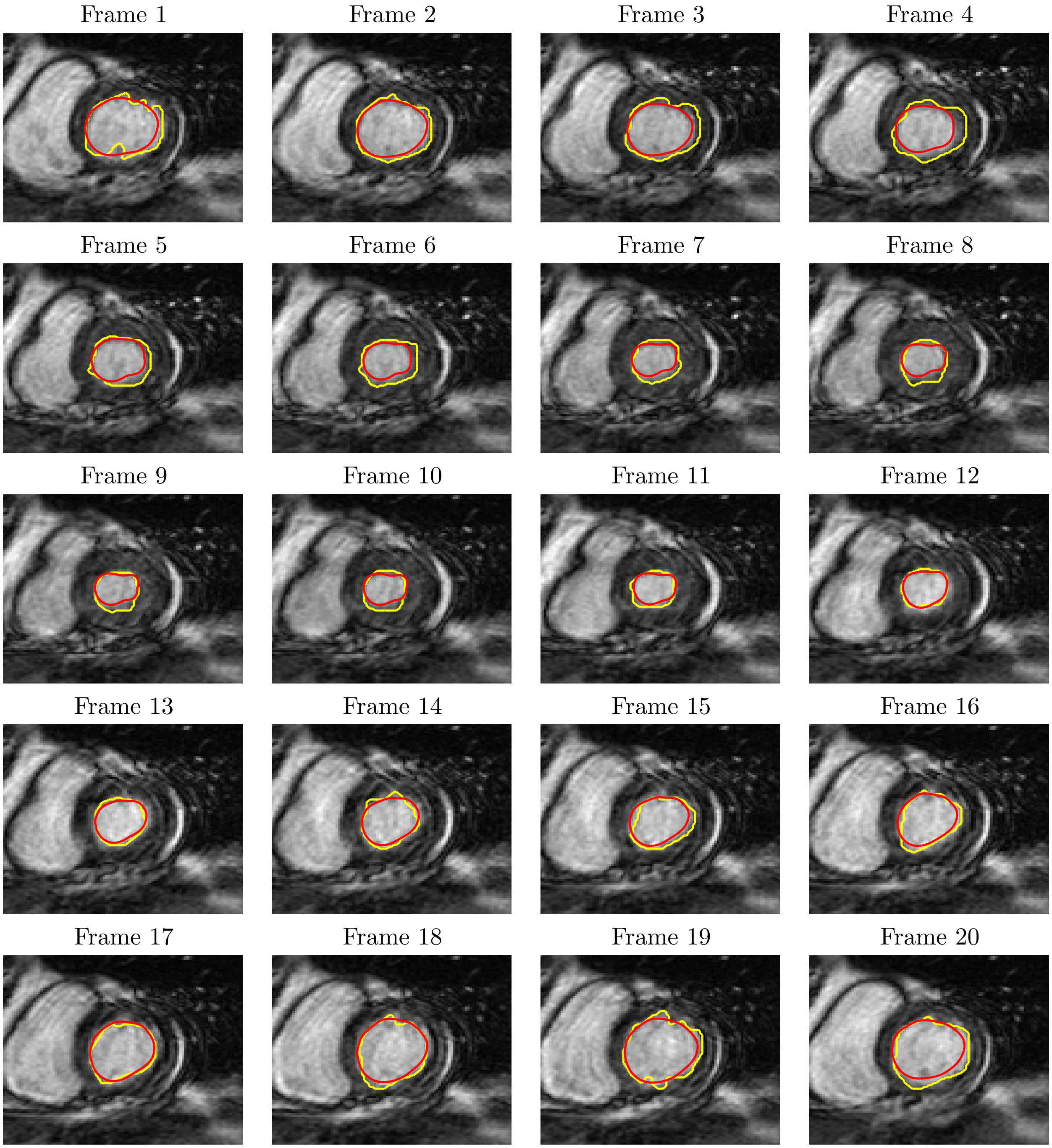
In Figure 4.13, we show LV boundary estimates $\vec{C}_t^*$ for frames 1 to 20 in the cardiac cycle. This estimate is obtained and provides what qualitatively appears to be a reasonable segmentation of the LV boundary. Quantitatively, the dice boundary coefficient is 0.8581. This result is of similar quality to the dice boundary coefficient obtained for the sample-based approach in the previous section (the dice boundary coefficient for the PTI case was 0.8654). In Section 4.5.5, we compare this result with that obtained from a static segmentation. Using the projection of the segmentation onto the space of X_t as described in Section 4.3.7, we determine the posterior density $p(X_t|y_{1:t})$ for each t. In Figure 4.14, we show equally-weighted samples of the posterior. Since the LV boundary estimates are close to the truth, the samples from the Gaussian posterior are distributed near the correct segmentation.

Results - Low SNR

We again consider results using low SNR images as data. Using the same initialization and trained dynamics as before, we recursively estimate the posterior and the LV boundary. Figure 4.15 shows LV boundary estimates for a full cardiac cycle. Visually, the segmentations appear to accurately localize the LV boundary. Figure 4.16 shows samples of the posterior density. Again, because of the Gaussian posterior assumption, the samples are accurately centered around the LV boundary. Quantitatively, the dice boundary coefficient is 0.8189. Table 4.2 summarizes and compares the results.

4.5.4 Comparison of Gaussian and Non-Gaussian Posterior Assumption

The results assuming a Gaussian posterior (dice coefficient of 0.8581 for high SNR and 0.8189 for low SNR) are similar to those for general non-parametric estimates of the posterior (0.8654 and 0.8123, respectively), both visually and quantitatively. Possible explanations for why this is true include (but may not be limited to):



Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 4.13. LV boundary segmentations (red) using high SNR images as data with the ground truth shown in yellow. The estimates are based on a Gaussian posterior and predictions obtained using a PTI model.

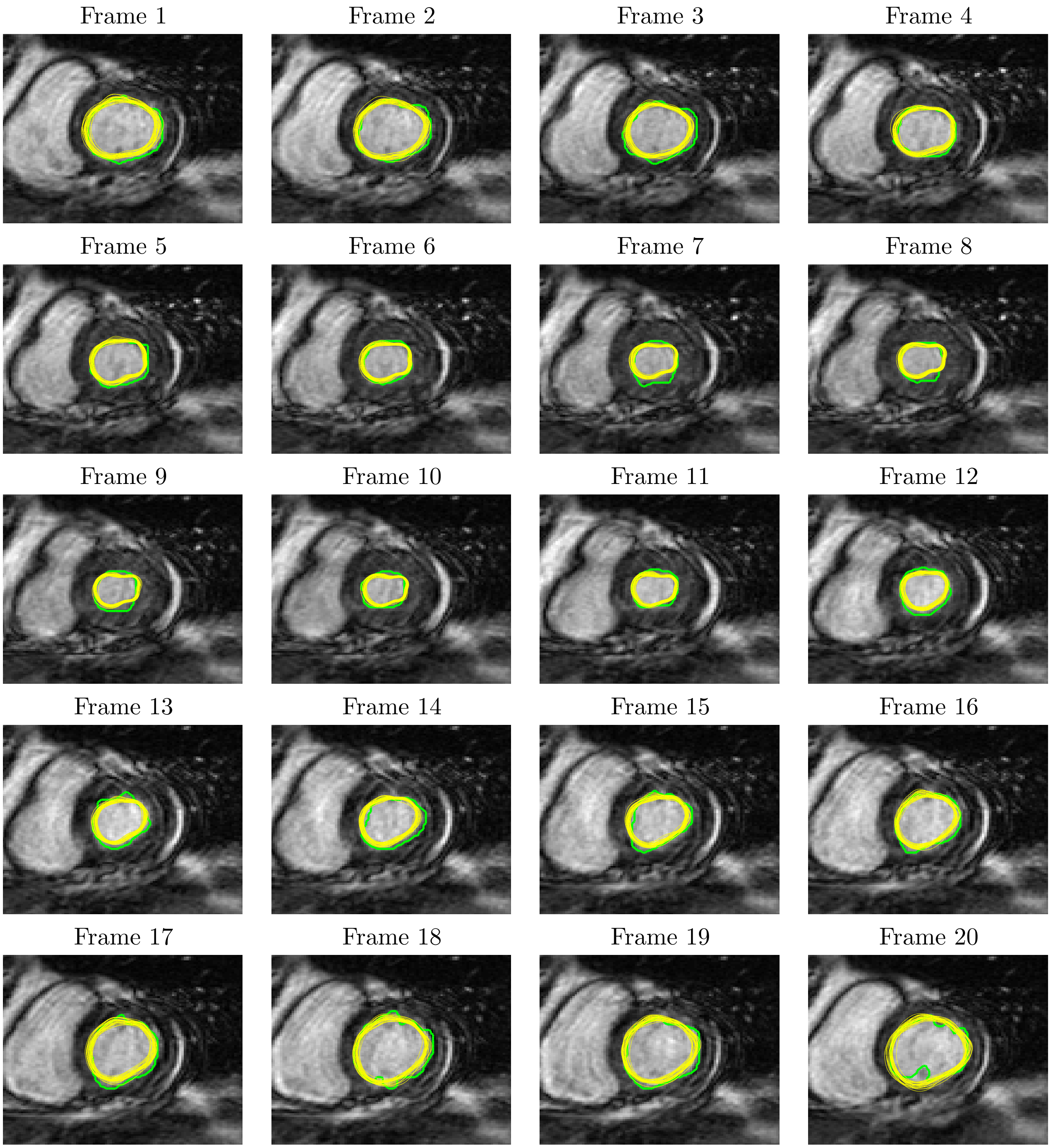
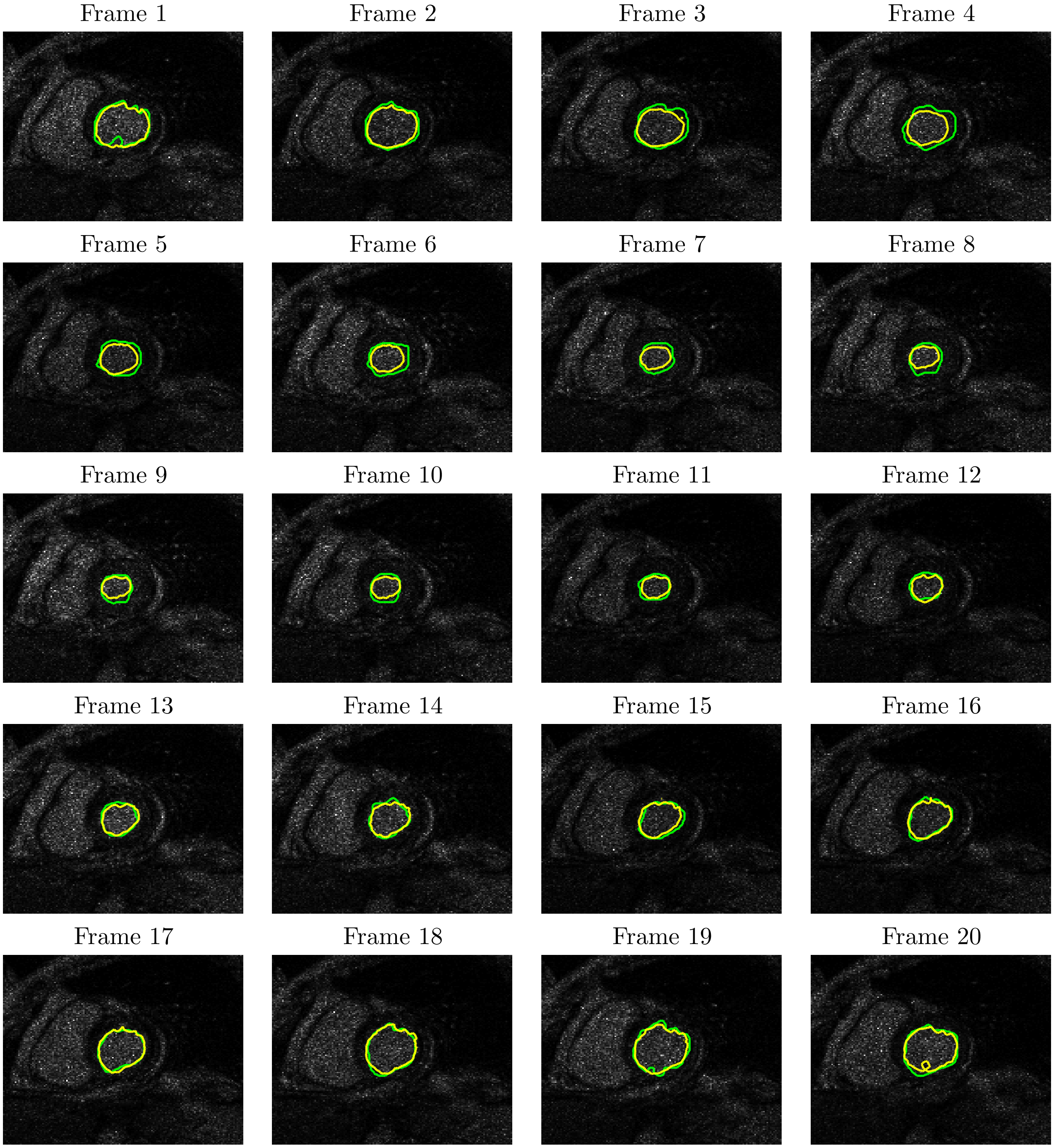


Figure 4.14. Curves (yellow) representing samples of the Gaussian posterior density $p(X_t|y_{1:t})$ having means equal to the segmentations shown in Figure 4.13. For reference, the ground truth is shown in green.



Frame 1 Frame 2 Frame 3 Frame 4

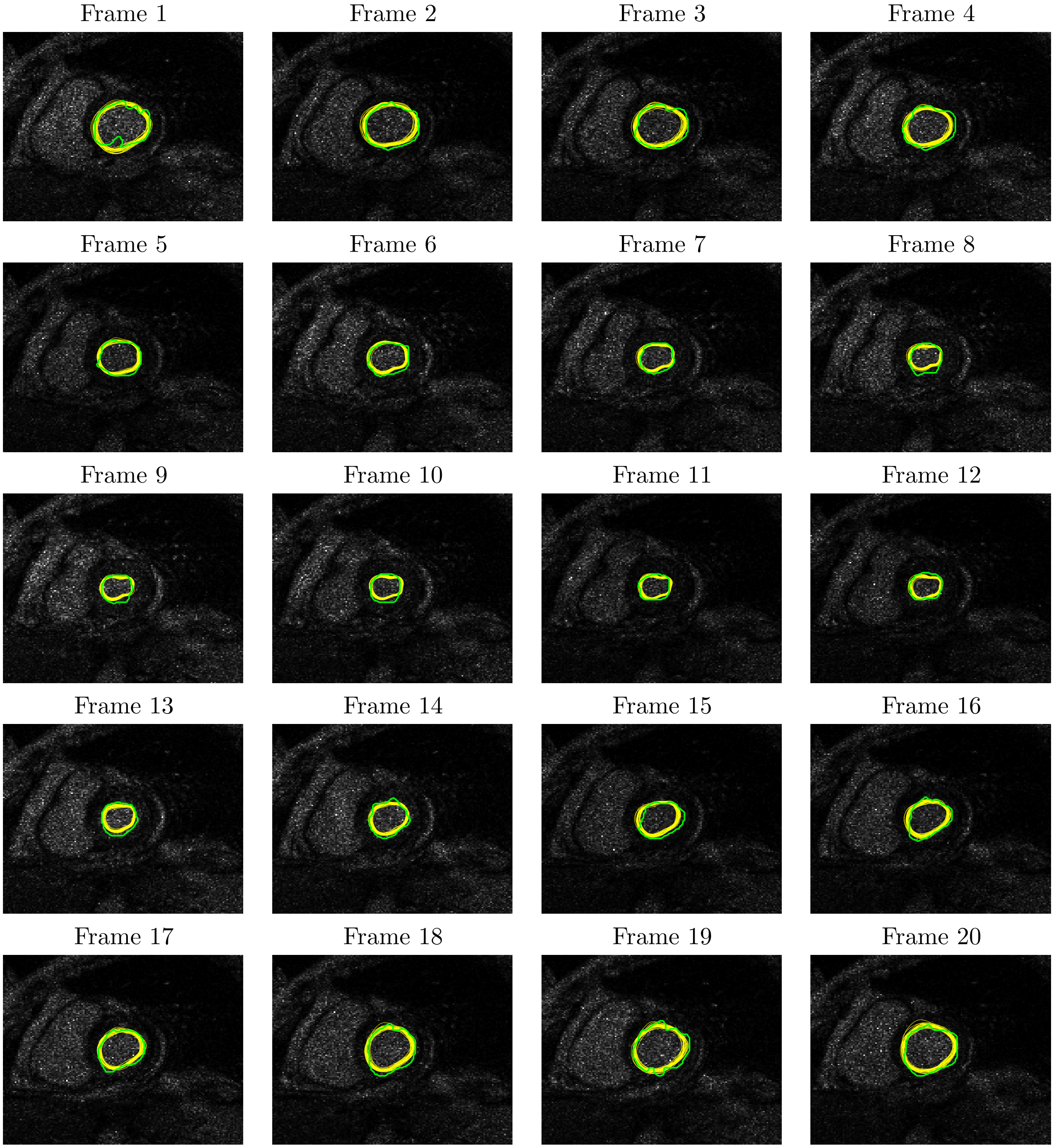
Frame 5 Frame 6 Frame 7 Frame 8

Frame 9 Frame 10 Frame 11 Frame 12

Frame 13 Frame 14 Frame 15 Frame 16

Frame 17 Frame 18 Frame 19 Frame 20

Figure 4.15. LV boundary segmentations (yellow) using low SNR images as data. The estimates are based on a Gaussian posterior and predictions obtained using the PTI model. Ground truth is shown in green.



Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 4.16. Curves representing samples of the Gaussian posterior density $p(X_t|y_{1:t})$ having means shown in Figure 4.15. Ground truth is shown in green.

	Filtering PTI	Filtering TI	Filtering PTI & Gaussian
High SNR	0.8654	0.8210	0.8581
Low SNR	0.8123	0.7996	0.8189

Table 4.2. Comparison of the dice boundary coefficient for the general sample-based method for both PTI and time-invariant cases as well as the Gaussian posterior formulation. (PTI = piecewise time-invariant, TI = time-invariant)

- the true posterior distribution of the LV boundary based on the training data is unimodal, making the Gaussian assumption reasonable, or
- the training data is limited, so the Gaussian posterior provides a reasonable estimate of the true non-parametric density.

The segmentations using either approach are similar as well. This can be attributed to the fact that in both cases, we apply curve evolution over a set of predictions which are similar enough that similar local minima are found when minimizing $E(f_t, \vec{C}_t)$.

However, there exist applications where one expects the posterior distribution to be multi-modal. For instance, the segmentation and tracking of multiple sclerosis lesions is made difficult because of the possibility of lesions splitting, appearing, and disappearing over a sequence of time [74, 100, 160]. Multiple sclerosis is a disease characterized by demyelination lesions in the brain's white matter. Lesions appear and disappear periodically with accurate knowledge of their occurrence and frequency being important for the diagnosis of a patient. In producing a prediction of a lesion which may split, one would expect the density estimate to be bimodal, with one mode representing a single lesion and another mode representing the presence of multiple lesions resulting from the separation of the single lesion. Thus, a Gaussian posterior would not work in this application.

■ 4.5.5 Comparison of the Proposed Approach and Static Segmentation Methods

It is important to show that the proposed recursive estimation algorithm provides improvements over traditional static segmentation methods. To demonstrate this, we provide a comparison between our approach and two static segmentation methods.

Static Segmentation with a Curve Length Prior

The first static segmentation method we use employs a curve length penalty for the prior term (i.e. shorter curves are preferred). Figure 4.17 shows the comparison between the segmentations obtained using a Gaussian posterior (shown in yellow) and those using static segmentations. For the static segmentation using a curve length prior (green curves), we take the boundary which minimizes the functional of Equation (4.16) except

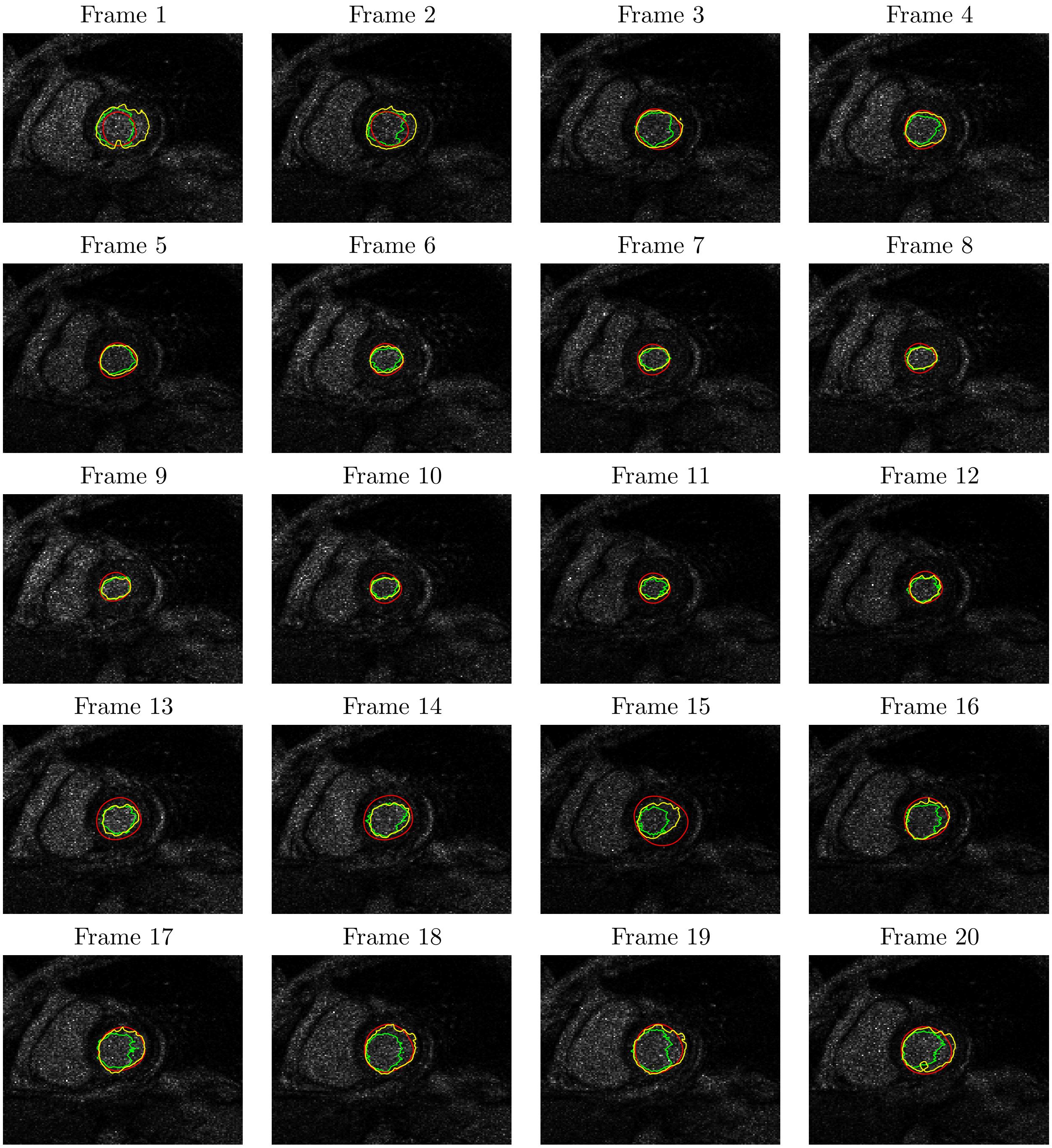


Figure 4.17. Comparison of the LV boundary estimates using a Gaussian posterior on PTI dynamics (yellow) (as shown in Figure 4.15) with static segmentations. Static segmentations using a curve length prior are shown in green while those using a shape prior are shown in red.

	Filtering PTI	Filtering TI	Filtering PTI & Gaussian	Static - Curve Length Prior	Static - Shape Prior
High SNR	0.8654	0.8210	0.8581	0.7741	0.8516
Low SNR	0.8123	0.7996	0.8189	0.6844	0.8035

Table 4.3. Dice boundary coefficient of the LV boundary estimates and the ground truth for different methods proposed in this chapter. (PTI = piecewise time-invariant, TI = time-invariant)

that the curve prior and prediction term ($-\log \int_{X_t} p(\vec{C}_t|X_t)p(X_t|y_{1:t-1})dX_t$) is chosen to be a curve length penalty ($\oint ds$) which uses no information from previous frames and is thus similar to the region-based segmentation methods described in Section 2.2.3.

As one may observe through the illustrations, incorporating dynamics into the segmentation process using the approach we propose results in better estimates than those obtained using a static segmentation method. Without a prediction, the static segmentation occasionally locates a region of high intensity which is only a subset of the region of interest (e.g. frames 1 and 2). Incorporating the prediction also results in boundary estimates that appear more like the shape of the LV (smooth and elliptical). Quantitatively, the dice boundary coefficient (based on low SNR data) is 0.8189 while that for static segmentation is 0.6844.

Static Segmentation with a Shape Prior

Alternatively, we can consider a more advanced static segmentation in which we use a prior based on a shape space determined by our training set. Namely, we employ Tsai *et al.*'s [178] approach for finding the best segmentation within the space of shapes defined by the mean and eight primary modes of variability (i.e. the space spanned by the shape terms of X_t in our approach) using the mean shape as an initialization.

The red curves in Figure 4.17 represent the segmentations obtained using a shape prior. Quantitatively, the dice boundary coefficient for the shape prior-based static segmentation is 0.8035. The accuracy using the shape prior is worse than that obtained in our filtering approach (yellow curves in Figure 4.17 which have a dice boundary coefficient of 0.8189). The fact that the filtering results are only somewhat better than the shape prior-based approach suggests that something more advanced should be considered in our formulation. Table 4.3 summarizes the dice boundary coefficients for the different results in this chapter.

■ 4.6 Conclusion

This chapter presents a principled approach to dynamically estimating a boundary (the left ventricle) across a temporal sequence of images (a cardiac cycle). To accomplish this, we propose a method for learning the dynamics of a deformable shape using an information-theoretic technique.

In the training phase, we learn the dynamics of the LV by obtaining a non-parametric density estimate for the system dynamics. We propose a system which has piecewise time-invariant dynamics depending on the phase of the cardiac cycle. The choice of learning a dynamic for the systolic and diastolic phases provides a compromise between the limitations of learning a fully time-varying dynamic model given the quantity and richness of the training data available and the desire to fully capture the non-stationary of the system across a cardiac cycle. In the experiments, we showed that this choice of dynamics is better than using a time-invariant dynamic model across the cardiac cycle.

After the dynamics are learned, we use them on test sets of low and high SNR data to recursively estimate the LV boundary. We show that incorporating segmentations from previous frames produces estimates which are slightly better than that obtained using a static segmentation (see Table 4.3). When training data are limited or when distributions are unimodal, we show that constraining the posterior to be Gaussian reduces the computational complexity yet produces reasonable results when compared to the general particle filtering approach. In this chapter, we have posed the problem as a forward recursive filter. The next chapter considers improving the estimates through the use of fixed-interval smoothing.

Chapter 5

Smoothing in Markov Chains and Single Cycle Graphs for Deformable Objects

IN the previous chapter, we proposed the recursive estimation of the left ventricle (LV) boundary using a filtering framework. In that formulation, the state estimate was based on past and current data. For problems where causal processing is required, the filtering framework is appropriate. However, in the application of cardiac imaging, LV boundary estimates can also be made assuming the knowledge of an entire cardiac cycle as well as additional information from temporally-adjacent cardiac cycles. For such problems, we generalize the filtering framework posed in Chapter 4 first by formulating the problem as one of fixed-interval smoothing on a Markov chain and then as one of inference on a single cycle graph.

A single cardiac cycle of the human heart takes place in approximately one second. Since it takes longer for a cardiologist or radiologist to analyze the data, causal processing is generally not necessary. In addition, because physical constraints force adjacent LV segmentations to be highly correlated, the knowledge of future segmentations is potentially as useful as that of past segmentations. Thus, using all data in a cardiac cycle to segment the LV is an attractive extension to the filtering problem. One way we use segmentations from future states is through smoothing on a Markov chain. In this chapter, we first estimate the LV boundary using this approach. However, the assumption of a simple chain does not capture the quasi-periodicity of the LV. In order to incorporate this, our second approach involves formulating the problem as a loopy graph to enforce the dependency between the beginning and end of each cardiac cycle. We apply the method of non-parametric belief propagation (NBP) to approximate the boundary estimates for this model.

In Section 5.1, we provide the basic framework for smoothing on a Markov chain, describing how the problem is solved first using NBP and then using a forward-backward approach. In Section 5.2, we exploit the periodic nature of the cardiac cycle and formulate the estimation problem as an inference problem on a graph with a single loop. We provide an overview for estimation by approximate inference on a single loop. Results are shown and compared with the results obtained using smoothing on a simple chain. Finally, Section 5.3 summarizes the chapter.

■ 5.1 Fixed-Interval Smoothing on a Markov Chain

In Chapter 4, we proposed learning the complex dynamics of the LV for the recursive estimation of LV boundaries in a filtering framework. In this chapter, we extend this work by examining fixed-interval smoothing on a Markov chain.

■ 5.1.1 Smoothing Formulation

Past work involving fixed-interval smoothing on a simple chain includes analysis on linear systems [1,2,134,135] as well as more general non-linear systems [51,58,67,104]. In this section, our contribution to the well-analyzed smoothing problem includes learning the dynamics, both forward $p(X_{t+1}|X_t)$ as described in Chapter 4 as well as backward $p(X_t|X_{t+1})$,[1] and using an implicit state representation X_t to estimate a deformable boundary in a fixed-interval smoothing framework. To estimate the states in a Markov chain, we apply the forward-backward algorithm of Doucet *et al.* [51] as well as inference using NBP [171].

The posterior density which we estimate with smoothing is $p(X_t|y_{1:T})$, where $y_{1:T} = [y_1, y_2, ..., y_T]$ is an entire set of observations over some pre-determined fixed interval. As before, the system is a hidden Markov model with the unobserved state X_t an implicit representation of the boundary, y_t the observed image intensity, and f_t the underlying field intensity.

The formulation for fixed-interval smoothing is similar to that shown in Equation (4.1) for filtering, except that the posterior is conditioned on the entire dataset, so

$$p(X_t|y_{1:T}) \propto \int_{f_t} \int_{\vec{C}_t} p(y_t|f_t, \vec{C}_t) p(f_t|\vec{C}_t) p(\vec{C}_t|X_t) p(X_t|y_{(1:T)\backslash t}) d\vec{C}_t df_t, \tag{5.1}$$

where $y_{(1:T)\backslash t} \equiv \{y_{1:t-1}, y_{t+1,T}\}$ (all y_τ for $\tau \in \{1, \cdots, T\}$ excluding y_t). Note that the likelihood, field prior, and curve prior terms are the same as in the previous chapter, but the prediction term is now conditioned on $y_{(1:T)\backslash t}$. Given this framework, we again approximate the integrals over f_t and $\vec{C}_t$ by finding the $(f_t^*, \vec{C}_t^*)$ pair which maximizes the integrand of Equation (5.1) when integrated over all X_t. Namely, we solve

$$(f_t^*, \vec{C}_t^*) = arg \max_{f_t, \vec{C}_t} \int_{X_t} p(y_t|f_t, \vec{C}_t) p(f_t|\vec{C}_t) p(\vec{C}_t|X_t) p(X_t|y_{(1:T)\backslash t}) dX_t. \tag{5.2}$$

This maximization is performed using curve evolution as described in Section 5.1.2. Using these approximations, Equation (5.1) can be written as

$$p(X_t|y_{1:T}) \propto p(y_t|f_t^*, \vec{C}_t^*) p(f_t^*|\vec{C}_t^*) p(\vec{C}_t^*|X_t) p(X_t|y_{(1:T)\backslash t}). \tag{5.3}$$

---

[1] In addition to the forward dynamics learned for filtering, the belief propagation-based solution requires the learning of the backward dynamics.

Given this framework, we explore two methods of estimating the posterior $p(X_t|y_{1:T})$, one using NBP and the other applying a forward-backward smoothing algorithm.

■ 5.1.2 Smoothing by Non-Parametric Belief Propagation

We consider the use of NBP for estimating the posterior $p(X_t|y_{1:T})$ for each t. Defining the observation, or node, potentials by

$$\psi_t(X_t, y_t) \equiv p(y_t|f_t^*, \vec{C}_t^*)p(f_t^*|\vec{C}_t^*)p(\vec{C}_t^*|X_t)p(X_t), \tag{5.4}$$

where we make the assumption that $p(X_t)$ is an uninformative prior. We define the pair-wise, or edge, potentials as

$$\psi_{t,\tau}(X_t, X_\tau) \equiv \frac{p(X_t, X_\tau)}{p(X_t)p(X_\tau)}. \tag{5.5}$$

The posterior $p(X_t|y_{1:T})$ can be written as a product of terms which includes the message products (described in Section 2.6). In particular,

$$p(X_t|y_{1:T}) \propto \int_{f_t} \int_{\vec{C}_t} p(y_t|f_t, \vec{C}_t)p(f_t|\vec{C}_t)p(\vec{C}_t|X_t) \prod_{\tau \in N(t)} m_{\tau t}(X_t) d\vec{C}_t df_t, \tag{5.6}$$

which we approximate by

$$p(X_t|y_{1:T}) \propto p(y_t|f_t^*, \vec{C}_t^*)p(f_t^*|\vec{C}_t^*)p(\vec{C}_t^*|X_t) \prod_{\tau \in N(t)} m_{\tau t}(X_t), \tag{5.7}$$

where each message is defined as

$$m_{\tau t}(X_t) \propto \int_{X_\tau} \psi_{t,\tau}(X_t, X_\tau)\psi_\tau(X_\tau, y_\tau) \prod_{s \in N(\tau)\backslash t} m_{s\tau}(X_\tau) dX_\tau. \tag{5.8}$$

Comparing Equation (5.6) with Equation (5.1), we see that the *prediction* $p(X_t|y_{(1:T)\backslash t})$ of Equation (5.1) is the product of the incoming messages. The message passing process is iterative in nature (see Equation (2.36) for the iterative expression of Equation (5.8)). For inference on tree-structured discrete or Gaussian graphical models, convergence occurs once the messages from each node have propagated to every other node in the graph. In NBP, since the particle-based density representations are approximate, convergence is less predictable. Thus, we iterate until successive estimates of the posterior of Equation (5.7) become similar (as determined by a measure such as the Kullback Leibler (KL) distance). The message products in Equation (5.7) and Equation (5.8) are computed using the efficient local Gibbs sampling algorithm explained in [171]. In the case of a simple chain, the product of the messages in Equation (5.8) reduces to a single term (for $1 < \tau < T$) or the identity (for $\tau = 1, T$) since $N(\tau)$ consists of at most two elements, $\{\tau - 1, \tau + 1\}$.

Backward Dynamics

Since $p(X_t|X_\tau)p(X_\tau)$ and $p(X_\tau|X_t)p(X_t)$ are equivalent forms of the joint density $p(X_t, X_\tau)$, the pair-wise potential shown in Equation (5.5) can be written either as $\frac{p(X_t|X_\tau)}{p(X_t)}$ or $\frac{p(X_\tau|X_t)}{p(X_\tau)}$. We write the potentials in terms of conditional densities because, as described in Section 4.4, we have a tractable technique for learning conditional densities from training data. For the Markov chain, at some non-terminal node t (i.e. for $t \neq 1, T$), messages are passed to this node from nodes $t-1$ and $t+1$. For messages passed forward (i.e. $m_{(t-1)t}$), the pair-wise potential used in Equation (5.8) is

$$\psi_{t-1,t}(X_{t-1}, X_t) = \frac{p(X_t|X_{t-1})}{p(X_t)} \tag{5.9}$$

which requires the forward conditional density. Meanwhile, for messages passed backward in time (i.e. $m_{(t+1)t}$), we write the pair-wise potential as

$$\psi_{t+1,t}(X_{t+1}, X_t) = \frac{p(X_t|X_{t+1})}{p(X_t)} \tag{5.10}$$

which requires the backward conditional density. Thus, we need both the forward dynamics $p(X_{t+1}|X_t)$ and the backward dynamics $p(X_t|X_{t+1})$ for NBP.

In Chapter 4, we discussed how the forward dynamics are learned. Similarly, to determine the backward dynamics, we first define a function $q'_{t+1}(X_{t+1})$, where $q'(\cdot)$ represents the statistic learned for backwards dynamics (as opposed to q for the forward dynamics), that again seeks to capture as much information from X_{t+1} pertaining to X_t while reducing the dimensionality for tractability. Again, we assume a linear function for q'_t and find the parameters which maximize the mutual information between X_t and $q'_{t+1}(X_{t+1})$. The result allows us to produce a kernel density estimate of $p(X_t|q'_{t+1}(X_{t+1}))$. As with learning the forward dynamics, we learn a different backward dynamic in each of the different stages of the cardiac cycle. Namely, we learn q'_S in the systolic phase and q'_D in the diastolic phase.

Note that the forward and backward dynamics are directly related by

$$\frac{p(X_t|X_{t-1})}{p(X_t)} = \frac{p(X_{t-1}|X_t)}{p(X_{t-1})}. \tag{5.11}$$

However, since taking quotients of densities which we approximate by particles is non-trivial and given that we have a framework to estimate the conditional densities, we rely on learning the forward and backward dynamics separately from the training data, and do not explicitly enforce the equality of Equation (5.11). Such an approximation is not unprecedented. Sigal *et al.* [163], for instance, learn the conditional densities separately in the process of tracking motion.

Iterative Process of NBP

We now describe how we iterate the parallel message passing (i.e. messages are passed along every edge in each iteration) process of NBP to obtain estimates of the posterior. At a particular iteration k, the product of all incoming messages represents

$\hat{p}^k(X_t|y_{(1:T)\setminus t})$ (we use the $\hat{p}^k$ notation to indicate that this is an intermediate estimate at iteration k), which is approximated by samples[2] $x^{(i)}_{t|T\setminus t}[k]$ having corresponding weights $w^{(i)}_{t|T\setminus t}[k]$ (this represents the weight of the prediction density, whereas that of the posterior density is written as $w_{t|T}$). To obtain an approximation for the posterior at each iteration, we find the $(f^*_t[k], \vec{C}^*_t[k])$ pair which maximizes the integrand of Equation (5.6) using curve evolution.

Curve Evolution to Determine f^*_t and $\vec{C}^*_t$

The curve evolution process is similar to that explained in Section 4.3.5 for the filtering problem. In particular, the likelihood, field prior, and curve prior terms are the same. Assuming $\hat{p}^k(X_t|y_{(1:T)\setminus t}) \equiv \sum_i w^{(i)}_{t|T\setminus t}[k]\delta(X_t - x^{(i)}_{t|T\setminus t}[k])$, we obtain a functional of the form

$$\begin{aligned} E_k(f_t, \vec{C}_t) &= -\log p(y_t|f_t, \vec{C}_t) - \log p(f_t|\vec{C}_t) - \log \int_{X_t} p(\vec{C}_t|X_t)\hat{p}^k(X_t|y_{(1:T)\setminus t})dX_t \\ &= (\int\limits_{z \in R_{in}(\vec{C}_t)} \frac{(\log y_t(z) - F^{in}_t(\vec{C}_t))^2}{2\sigma_n^2}\, dz + \int\limits_{z \in R_{out}(\vec{C}_t)} \frac{(\log y_t(z) - F^{out}_t(\vec{C}_t))^2}{2\sigma_n^2} dz) \\ &+ (\frac{(F^{in}_t(\vec{C}_t) - u)^2}{2\sigma_u^2} + \frac{(F^{out}_t(\vec{C}_t) - v)^2}{2\sigma_v^2}) + \log[\frac{1}{MN}\sum_{(i)} w^{(i)}_{t|T\setminus t}[k] exp(-D^2(\vec{C}_t, x^{(i)}_{t|T\setminus t}[k]))], \end{aligned} \tag{5.12}$$

where, as in the previous chapter, we define $F^{in}_t(\vec{C}_t) = \log f_t^{R_{in}(\vec{C}_t)}$ and $F^{out}_t(\vec{C}_t) = \log f_t^{R_{out}(\vec{C}_t)}$, with $f_t^{R_{in}(\vec{C}_t)}$ and $f_t^{R_{out}(\vec{C}_t)}$ unknown constant field intensities for the blood pool region inside (R_{in}) and the myocardial region immediately outside (R_{out}) the LV boundary, u and v are the prior means inside and outside the boundary, respectively, and D is as defined in Equation (4.13).

Using Field and Curve Parameters to Update the Observation Potential

Once the values of $f^*_t[k]$ and $\vec{C}^*_t[k]$ are found to minimize this functional (and thus maximize the posterior), we update the observation potential $\psi^k_t(X_t, y_t)$ (in our formulation, ψ_t changes in each iteration) using Equation (5.4). Practically, we take equally-weighted samples $x^{(i)}_{\psi_t}[k]$ of $p(X_t)$ and weight them according to

$$w^{(i)}_{\psi_t}[k] \propto p(y_t|f^*_t, \vec{C}^*_t)p(f^*_t|\vec{C}^*_t)p(\vec{C}^*_t|X_t = x^{(i)}_{\psi_t}[k]) \tag{5.13}$$

so that

$$\psi^k_t(X_t, y_t) \approx \sum_i w^{(i)}_{\psi_t}[k]k(X_t; x^{(i)}_{\psi_t}[k]), \tag{5.14}$$

---

[2] These samples are indexed by k since they are different from iteration to iteration. The notation $x_{t|T\setminus t}$ represents samples of $\hat{p}^k(X_t|y_{(1:T)\setminus t})$.

where $k(x;x^0)$ represents a kernel centered at x^0. Using a Gibbs sampler, the message products of Equation (5.7) are recomputed. The posterior $\hat{p}^k(X_t|y_{1:T})$ is obtained from this equation by sampling the message products to obtain a new set of samples $x^{(i)}_{t|T}[k]$ and re-weighting them according to

$$w^{(i)}_{t|T}[k] \propto p(y_t|f^*_t, \vec{C}^*_t)p(f^*_t|\vec{C}^*_t)p(\vec{C}^*_t|X_t = x^{(i)}_{t|T}[k]). \tag{5.15}$$

We re-sample the distribution to remove samples having small weights and create multiple samples for heavily-weighted ones, resulting in an equally-weighted set of samples.

■ 5.1.3 Results on Markov Chain using NBP

As in Chapter 4, the data are two sets of twenty-frame sequences, one high SNR and the other low, of cardiac MR images. All of the experiments in this chapter assume a first-order system model (i.e. $X_t = [A_t; \vec{\alpha}^T_t]^T$, a nine-dimensional vector containing the area and coefficients for the eight primary modes of variability). Again, we use a training set of hand segmentations of the left ventricles to learn the dynamics. The forward dynamics are found as in Section 4.4, with piecewise time-invariant dynamics learned depending on the phase of the cardiac cycle (systolic or diastolic). For the backward dynamics $p(X_t|X_{t+1})$, we also learn the dynamics on the same two phases. Using NBP, we iterate the parallel message passing algorithm until convergence. Figure 5.1 shows the 50 most likely samples of the posterior when the data are of high quality. Similarly, Figure 5.2 shows the 50 most likely samples when we use low SNR data. As the results shown in Figures 5.1 and 5.2 make clear, the samples exhibit more variability than those obtained from filtering in the previous chapter. Although the exact reason for this is uncertain, we conjecture that the greater uncertainty determined by NBP is perhaps a more accurate indication of the variability inherent in the posterior densities. Thus, the estimates made by NBP should not be considered as less accurate. Further evidence that this is the case is the fact that the boundary estimates using NBP on the Markov chain are more accurate than those obtained in filtering, as discussed in the next paragraph.

Associated with these posteriors, curve evolution is used to find a single estimate $\vec{C}^*_t$ of the LV boundary using Equation (5.2). Figure 5.3 and 5.4 show the LV boundary estimates $\vec{C}^*_t$ for both high and low SNR data, respectively. As in the previous chapter, we use the dice measure (see Section 4.5.1) to evaluate the accuracy of the estimates. The dice boundary coefficient in the high SNR case is 0.9214 while that for the low SNR case is 0.8909. The segmentations using inference on the Markov chain are more accurate than those based on filtering (dice boundary coefficient of 0.8654 for high SNR and 0.8210 for low SNR) as we would expect because the posterior estimates are based on more data. Table 5.1 summarizes the results presented in this chapter, the piecewise time-invariant (PTI) filtering results from the previous one, and the static segmentation results using a shape prior.

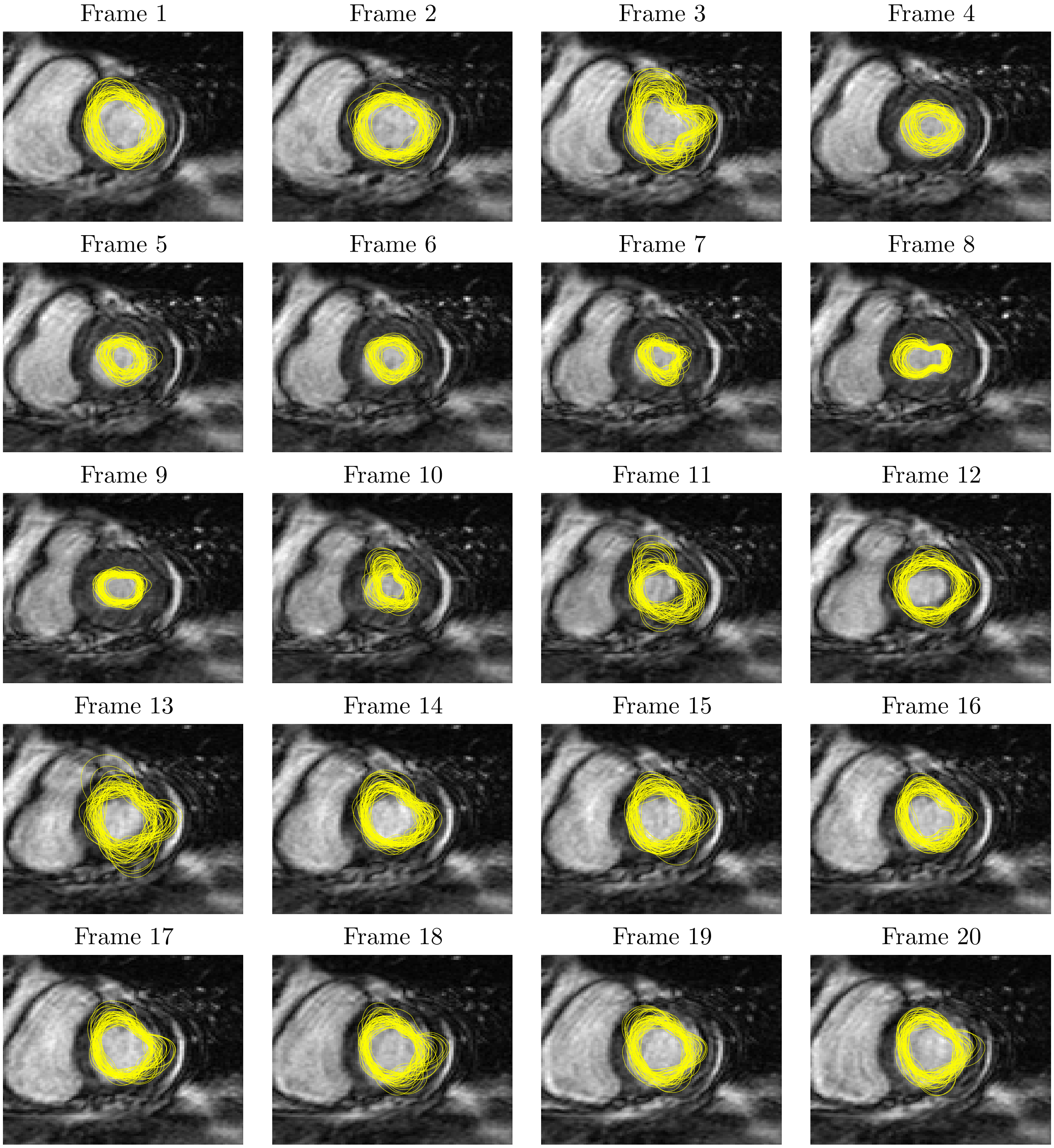


Figure 5.1. Samples of the posterior obtained from smoothing using NBP. The high SNR observations are shown in the background.

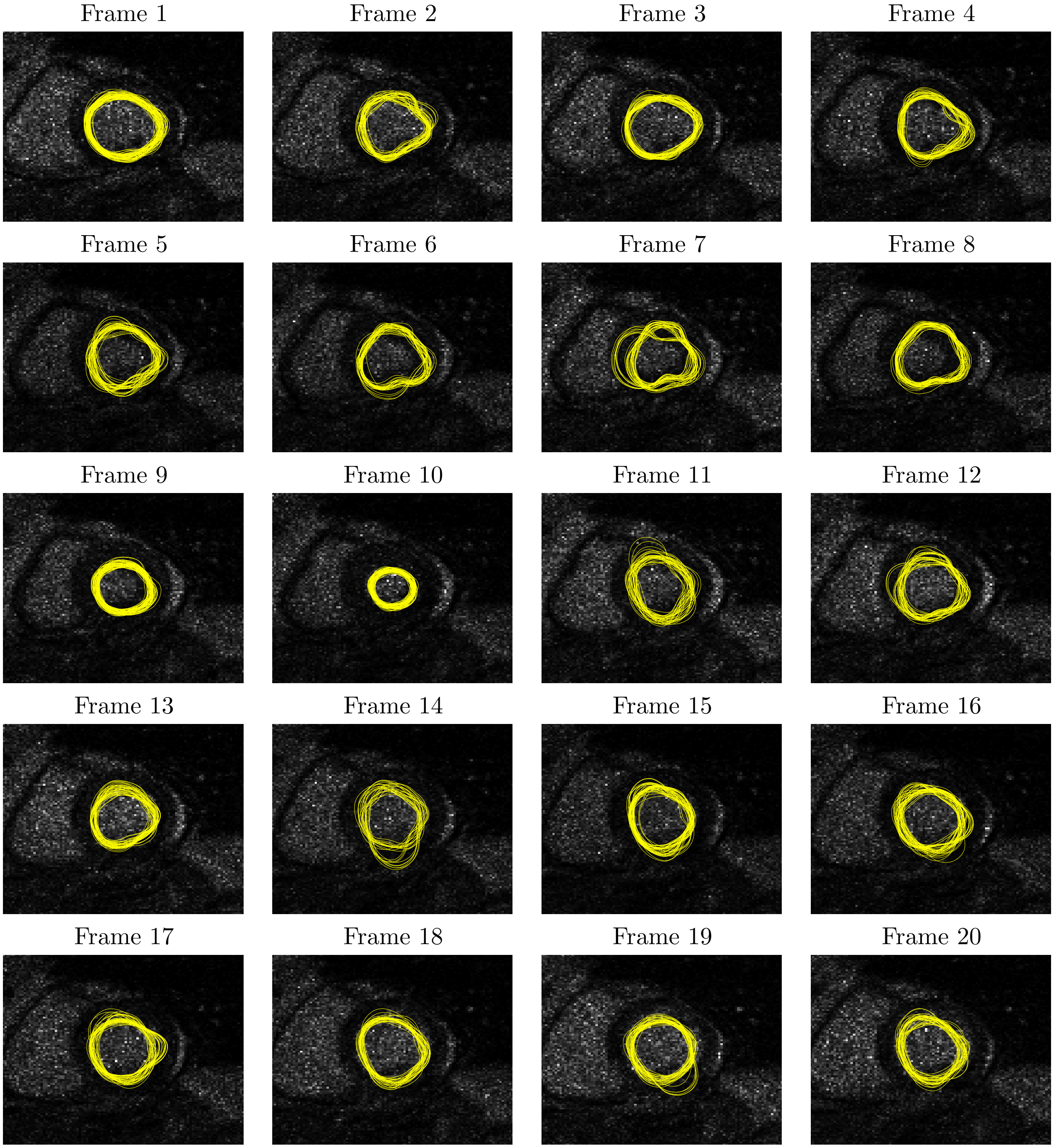

Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 5.2. Samples of the posterior obtained from smoothing using NBP on low SNR data. The observation for each frame is shown in the background.

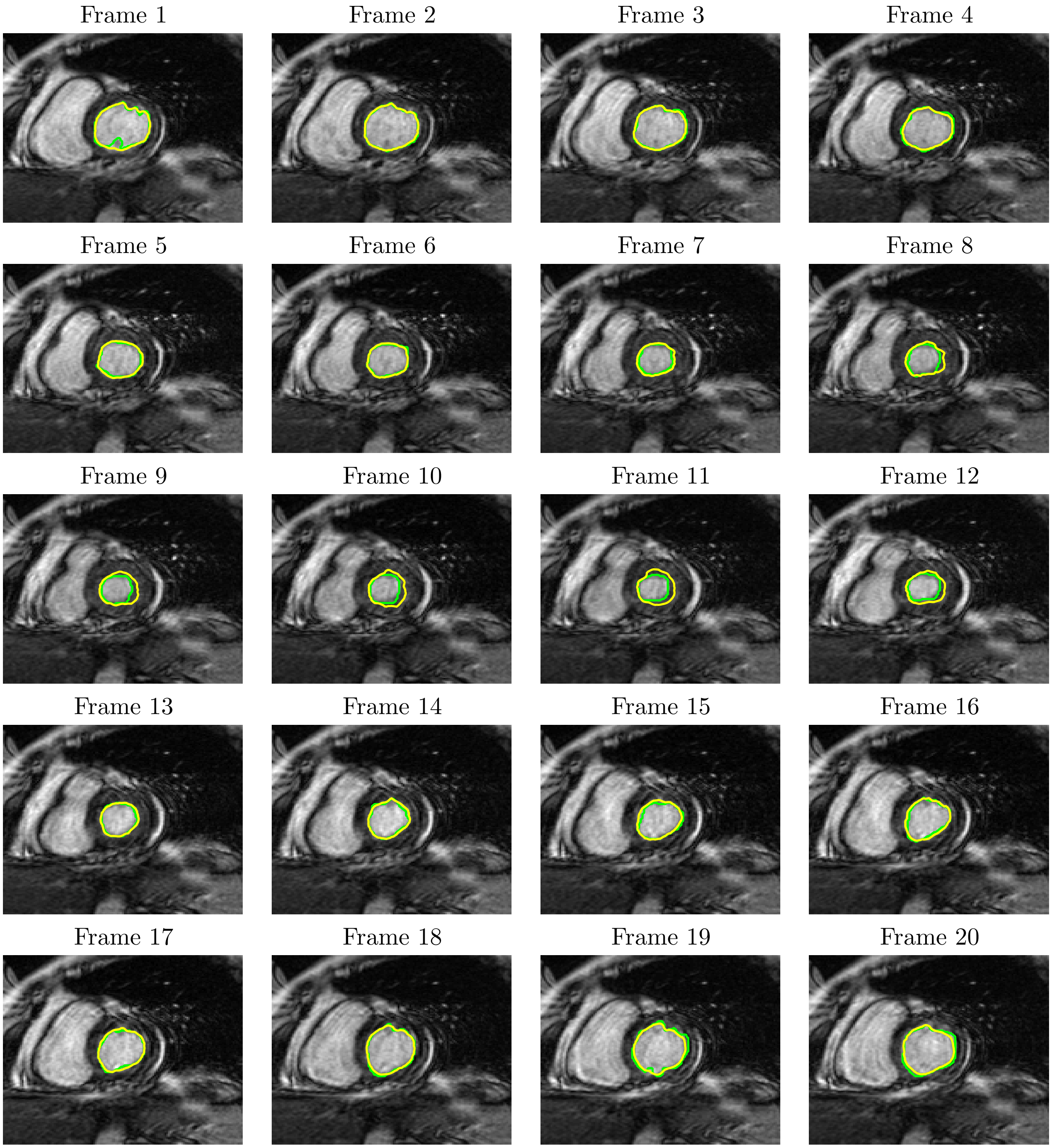


Figure 5.3. LV segmentations (yellow) obtained from curve evolution based on the posterior (having samples shown in Figure 5.1) obtained from smoothing using NBP on high SNR observations. Ground truth shown in green.

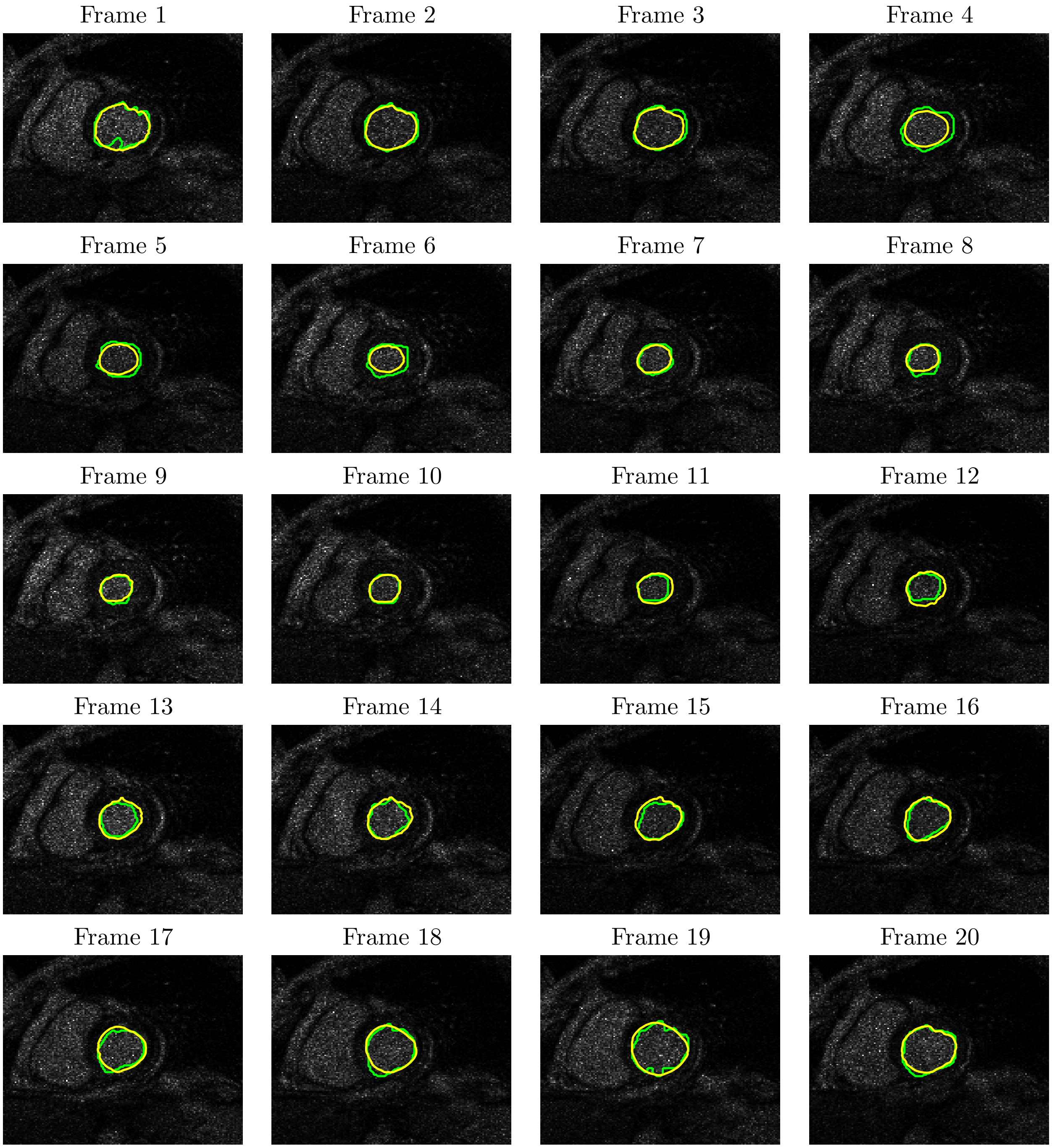


Figure 5.4. LV segmentations (yellow) obtained from curve evolution based on the posterior (having samples shown in Figure 5.2) obtained from smoothing using NBP on low SNR observations. Ground truth shown in green.

	Static - Shape Prior	Filtering PTI	Smoothing by NBP
High SNR	0.8516	0.8654	0.9214
Low SNR	0.8035	0.8210	0.8909

Table 5.1. Dice boundary coefficient for Markov chain smoothing compared with filtering results from the previous chapter and static segmentation results using a shape prior. (PTI = piecewise time-invariant)

■ 5.1.4 Forward-Backward Method for Markov Chain

Using an alternate approach to estimate the posteriors, we propose a method based on Doucet *et al.*'s [51] forward-backward algorithm which involves re-weighting the particles that were obtained through the forward filtering process. One advantage of this method is that the backward dynamics do not need to be computed, so issues regarding the consistency of the forward and backward dynamics (i.e. enforcement of Equation (5.11)) do not arise. However, unlike the previous method which re-samples the posterior density at each iteration, this algorithm uses the samples obtained from the forward pass (i.e. the filtering process). This restricts the ability of the smoothing step to change the representation of the density (i.e. it cannot create new samples).

The forward step of the algorithm uses particle filtering as explained in Section 4.3. For each time t, a set of samples $x_t^{(i)}$ having associated weights $w_t^{(i)}$ represents the posterior density $p(X_t|y_{1:t})$.

By considering the smoothing formula

$$p(X_t|y_{1:T}) = \int p(X_{t+1}|y_{1:T}) \frac{p(X_t|y_{1:t})p(X_{t+1}|X_t)}{p(X_{t+1}|y_{1:t})} dX_{t+1}, \tag{5.16}$$

estimates of $p(X_t|y_{1:T})$ can be performed going backward in time. In particular, suppose at time $t+1$ we have $x_{t+1}^{(j)}$ having weights $\tilde{w}_{t+1}^{(j)}$, where the $\tilde{w}_{t+1}^{(j)}$'s correspond to the weights at time $t+1$ for the points $x_{t+1}^{(j)}$ that represent the smoothing posterior $p(X_{t+1}|y_{1:T})$. Using Equation (5.16), we obtain the weights $\tilde{w}_t^{(i)}$ from $\tilde{w}_{t+1}^{(j)}$ by

$$\tilde{w}_t^{(i)} = \frac{1}{Z} \sum_j \tilde{w}_{t+1}^{(j)} w_t^{(i)} p(X_{t+1} = x_{t+1}^{(j)} | X_t = x_t^{(i)}), \tag{5.17}$$

where the summation is an approximation to the integral and Z represents a normalization constant. We note that the denominator $p(X_{t+1}|y_{1:t})$ of Equation (5.16) is not reflected in the re-weighting because this term is constant given $p(X_{t+1}|y_{1:T})$.

■ 5.1.5 Results on Markov Chain Using Forward-Backward Method

We now present results for smoothing on a Markov chain using the forward-backward method. In addition, we compare these results with those obtained through filter-

	Static - Shape Prior	Filtering PTI	Smoothing by NBP	Smoothing by Forward-Backward
High SNR	0.8516	0.8654	0.9214	0.9210
Low SNR	0.8035	0.8210	0.8909	0.8836

Table 5.2. Dice boundary coefficient for Markov chain smoothing using both the forward-backward method and NBP compared with filtering results from the previous chapter and static segmentation results using a shape prior. (PTI = piecewise time-invariant)

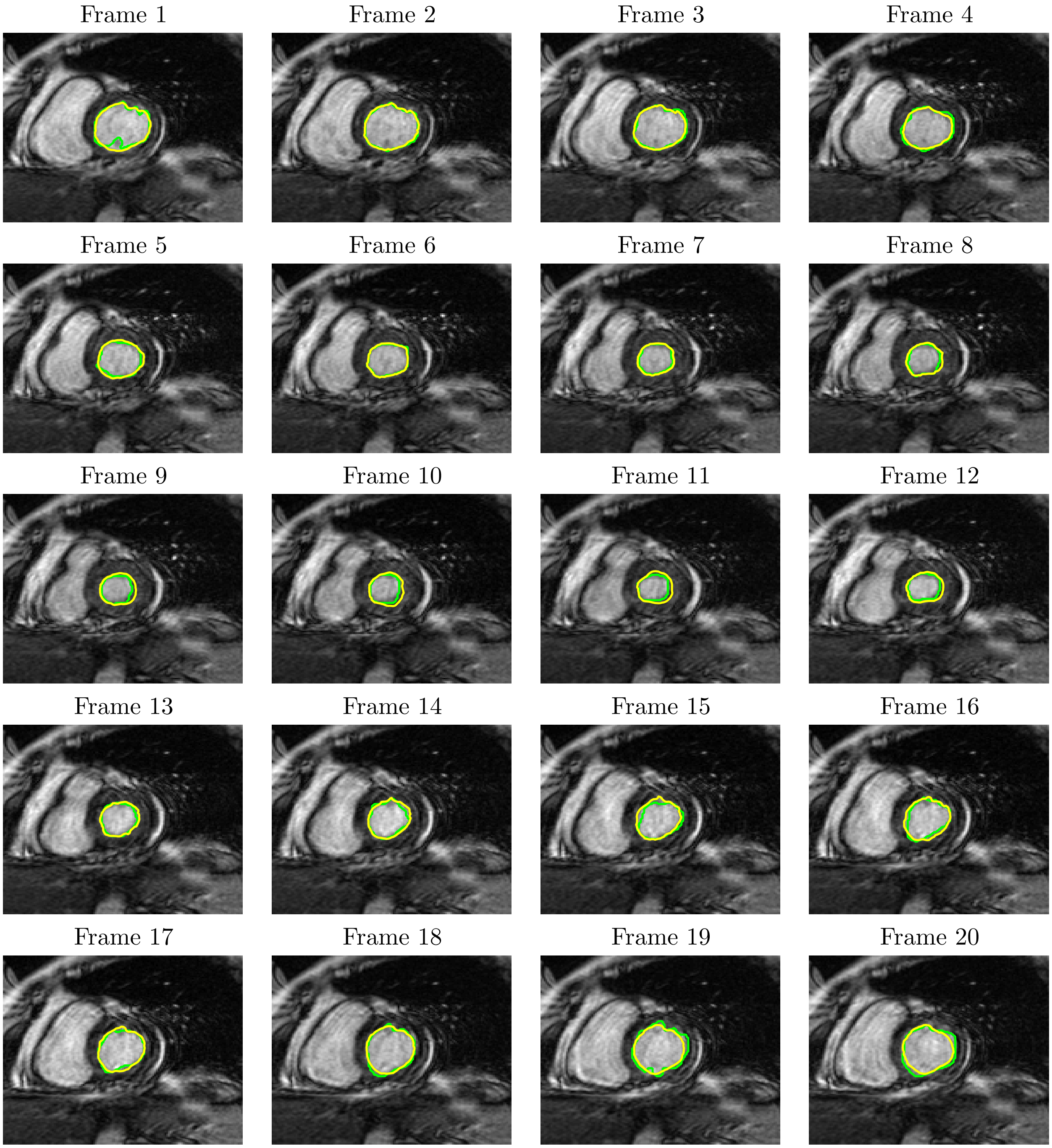
ing alone for the LV boundary estimate and the posterior estimate (i.e. we compare $p(X_t|y_{1:t})$ with $p(X_t|y_{1:T})$).

Comparison between Smoothing and Filtering on High SNR Data

As with NBP on the Markov chain, we obtain segmentations on the forward-backward method by using curve evolution to find $\vec{C}_t^*$ according to Equation (5.2). Figure 5.5 shows the segmentations obtained for the high SNR data. The dice boundary coefficient is 0.9210. The accuracy is similar to that obtained in smoothing by NBP (dice boundary coefficient of 0.9214) and better than that obtained from filtering (dice boundary coefficient of 0.8654). Table 5.2 compares these results with previously discussed approaches. In addition, we find it instructive to examine the accuracy of the samples of the posterior between filtering and forward-backward smoothing on the Markov chain. It is of particular interest using this approach because the forward filter is part of the algorithm for the forward-backward smoothing.

Figure 5.6 shows the 50 most likely samples of the posteriors determined using the method adapted from Doucet's forward-backward smoothing. The results can be compared with Figure 4.6 which shows the forward filtering portion of this analysis. One may observe visually that smoothing results in a posterior estimate having a smaller variance than that from filtering (e.g. compare frame 6 for the two methods). Furthermore, the samples from forward-backward smoothing appear to be more accurate estimates than those obtained from the forward filtering alone. To confirm this, we computed the dice coefficient for the 50 most likely samples for each of the two methods. For filtering, the dice sample coefficient is 0.7683 (since the average comes from a set of samples of the posterior, we expect this number to be smaller than the dice boundary coefficient). On the other hand, the dice sample coefficient for forward-backward smoothing is 0.8452. Table 5.3 shows these results.

To provide a more direct visual comparison between filtering and smoothing, Figure 5.7(a) shows the 50 most likely samples from filtering (yellow) and forward-backward smoothing (red) as compared with the manually-segmented truth (blue) for two representative frames. Since the method of smoothing used here involves a re-weighting of the filtering samples, one can observe the reduction in variance of the posterior by noting that the smallest and largest curves (by area) in these images mostly come from



Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 5.5. Segmentations (yellow) obtained through finding the MAP estimate of the forward-backward smoothing posterior given high SNR observations. Ground truth shown in green.

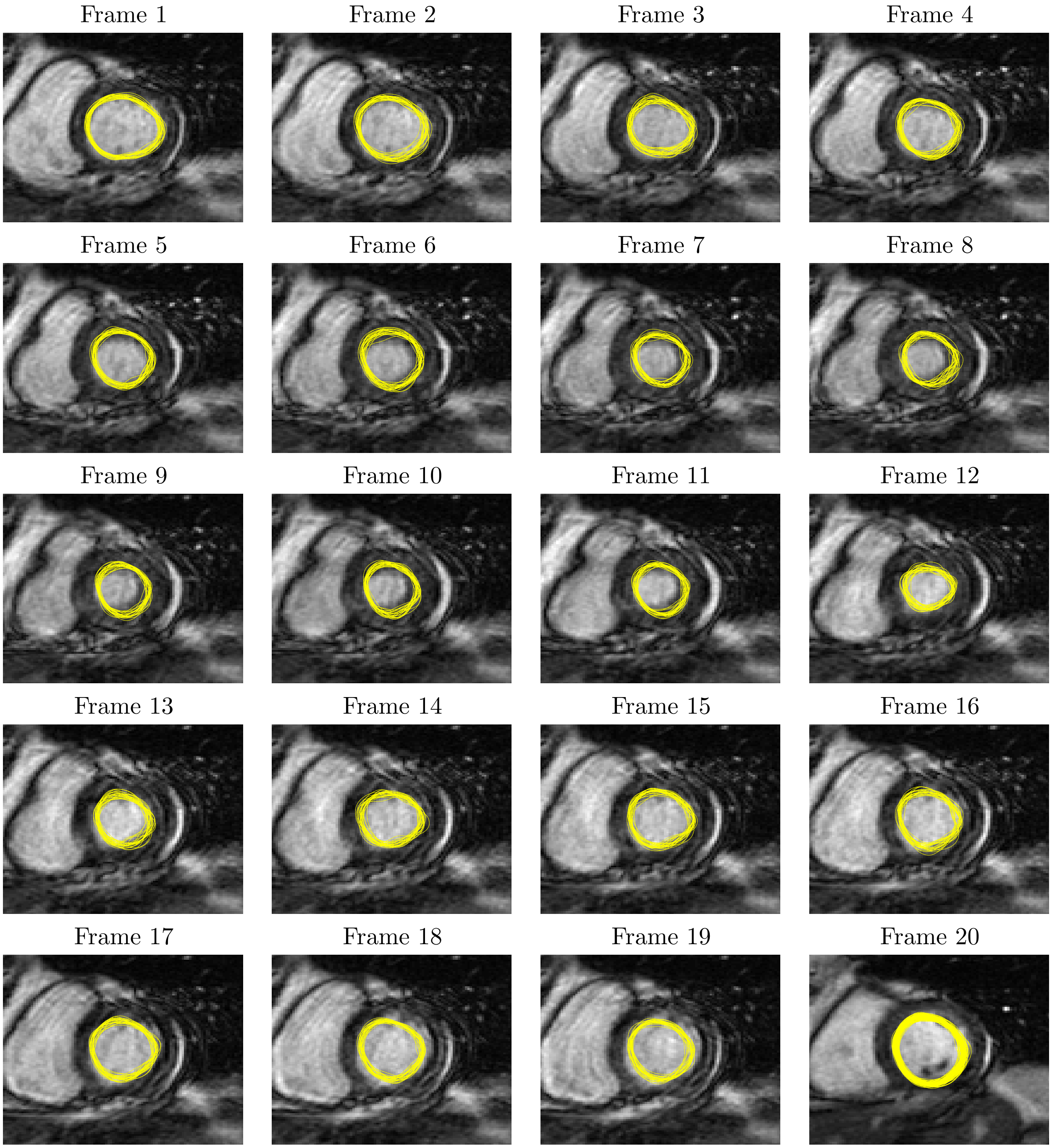

Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 5.6. Curves representing the posteriors $p(X_t|y_{1:t})$ (yellow) superimposed on the high SNR observation y_t. Posteriors obtained using forward-backward smoothing.

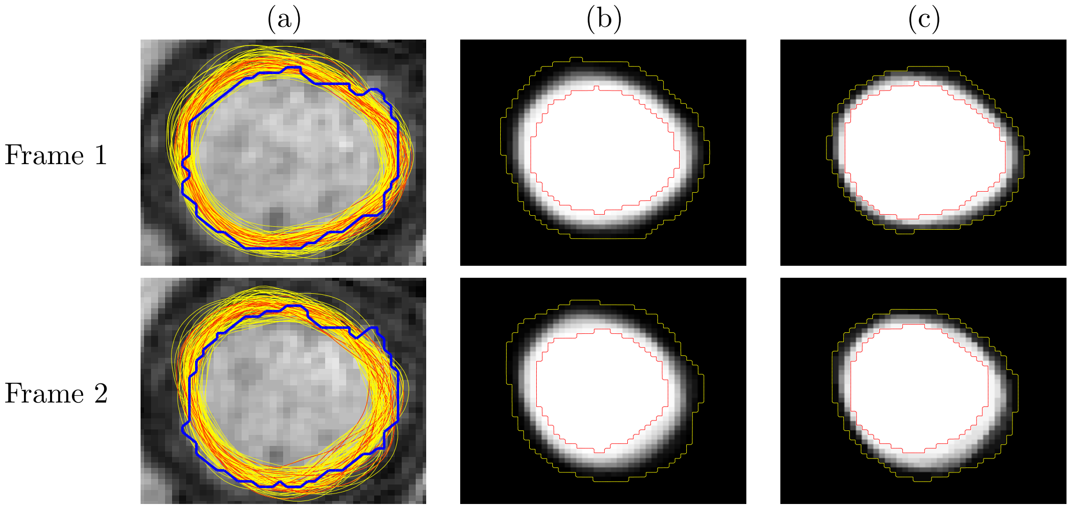


Figure 5.7. Comparison between filtering and Doucet's forward-backward smoothing on high SNR data. (a) Samples of the filtering posterior $p(X_t|y_{1:t})$ are shown in yellow, while samples of the smoothing posterior $p(X_t|y_{1:T})$ are shown in red. The manually-segmented truth is shown in blue. (b) Image which illustrates the distribution of the filtering samples, with the two curves (yellow and red) indicating the largest and smallest curves in the samples and the gray scale variations showing the frequency that a given pixel is contained within the curve samples (black indicating zero occurrences and white indicating that the point is contained in all of the sampled curves). (c) Image which shows the distribution of the smoothing samples using the same method as in (b), with the two curves again indicating the largest and smallest curves.

the posterior $p(X_t|y_{1:t})$ (filtering) rather than $p(X_t|y_{1:T})$ (smoothing). Alternatively, Figures 5.7(b) and (c) show the variability of the curves in a different way. Namely, we first create a binary map (1 inside and 0 outside) of each curve, and then aggregate all of them by plotting their average. For each image, the red curve indicates the smallest curve of the samples while the yellow curve indicates the largest. For filtering (Figure 5.7(b)), the distribution of curves (as indicated by the area between the innermost and outermost curve) is larger than that for smoothing (Figure 5.7(c)). For each pixel between the two extremes, the particular shade of gray determines the frequency in which it lies inside the set of curves (white indicates pixels inside all of the curves while black indicates points outside all of the curves).

Comparison between Smoothing and Filtering on Low SNR Data

Similarly, we can compare the forward-backward method of smoothing against filtering using low SNR data for observations. Figure 5.8 shows the segmentations using low SNR data. The dice boundary coefficient is 0.8836. This result is similar to the 0.8909 obtained by NBP on the same Markov chain and better than the 0.8210 obtained using filtering. Table 5.2 compares the results using forward-backward filtering with all of

Dice *sample* coefficient	Filtering PTI	Smoothing by Forward-Backward
High SNR	0.7683	0.8452
Low SNR	0.6951	0.7904

Table 5.3. Dice sample coefficient for Markov chain smoothing using the forward-backward method compared with that from filtering. The values represent an average of the accuracy of the top 50 samples of the posterior. (PTI = piecewise time-invariant)

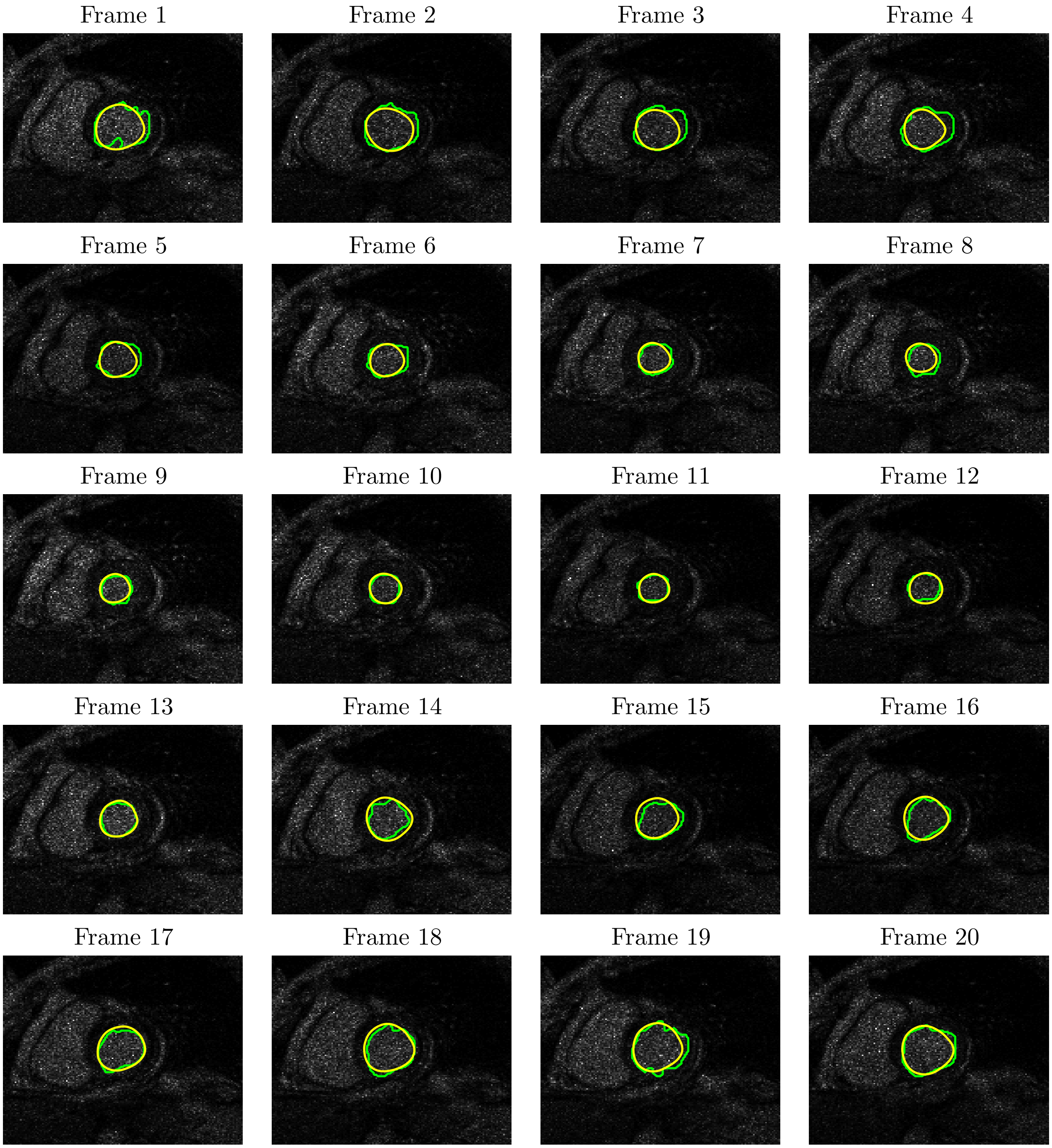
the previous results previously shown in Table 5.1. To compare posterior distributions, Figure 5.9 shows the 50 most likely samples of the posterior with the associated low SNR observations shown in the background, while Figure 4.10 shows the same results using filtering. Quantitatively, the dice sample coefficient is 0.6951 for filtering and 0.7904 for forward-backward smoothing. Table 5.3 shows the results of the dice sample coefficient for high and low SNR data.

Again, to provide a direct visual comparison between filtering and smoothing, Figure 5.10(a) shows the most likely samples from filtering (yellow) and forward-backward smoothing (red) as compared with the manually-segmented truth (blue) for two representative frames. We can again observe the reduction in variance of the posterior by noting that the smallest and largest curves (by area) in these images mostly come from the posterior $p(X_t|y_{1:t})$ (filtering) rather than $p(X_t|y_{1:T})$ (smoothing). Figures 5.7(b) and (c) show the variability of the curves in a different way for filtering and smoothing, respectively, in the same manner as that shown in Figure 5.7(b) and (c) and explained earlier.

■ 5.2 Single Cycle Graphs

In Section 4.5.2, we first exploit the quasi-periodicity of the LV to obtain the initial condition for recursive estimation. In this section, we explicitly incorporate this periodicity using a graph with a single loop. Using a simple chain, LV boundary estimates near the end are loosely coupled to those at the beginning of a cardiac cycle. Adding the dependency between the first and last frames of the cardiac cycle strengthens the propagation of information in the system, a desirable feature since the evolution of the LV is approximately periodic. Figure 5.11 shows the graph for this model obtained by adding an edge between the first and last node of the Markov chain. This graph implies that the canonical cardiac cycle returns to the same position each period. Data obtained at the same point in the cycle but from different cardiac cycles can be considered as multiple observations at a single location. This problem can be viewed as a boundary value stochastic process [1, 2]. Alternatively, it can be viewed as a loopy graphical model.

On tree-structured graphs containing only discrete or Gaussian variables, exact inference is possible using belief propagation (BP) [145]. In single cycle graphs, empirical



Frame 1 Frame 2 Frame 3 Frame 4

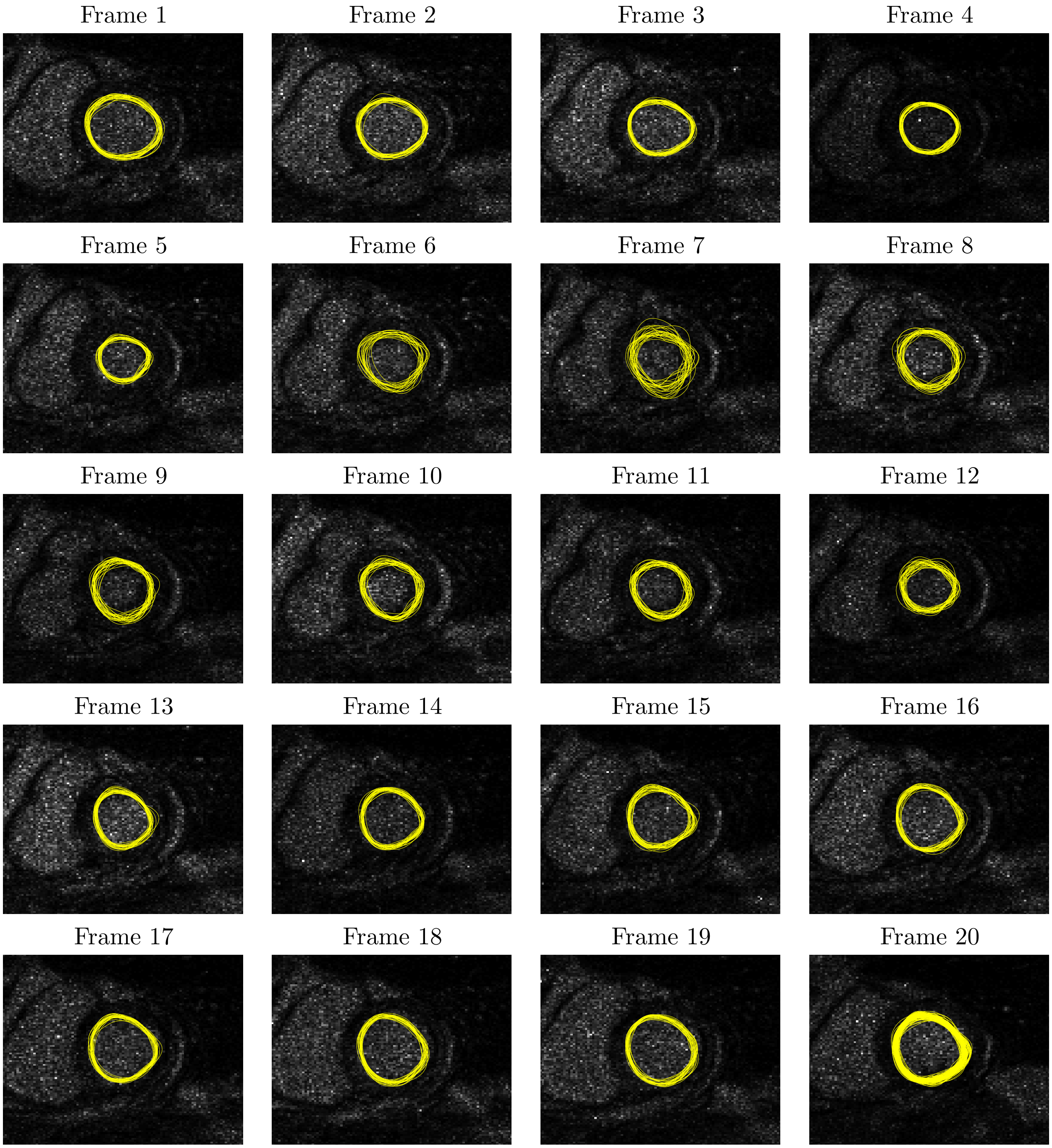
Frame 5 Frame 6 Frame 7 Frame 8

Frame 9 Frame 10 Frame 11 Frame 12

Frame 13 Frame 14 Frame 15 Frame 16

Frame 17 Frame 18 Frame 19 Frame 20

Figure 5.8. Segmentations (yellow) obtained through finding the MAP estimate of the forward-backward smoothing posterior given low SNR observations. Ground truth shown in green.



Frame 1 Frame 2 Frame 3 Frame 4

Frame 5 Frame 6 Frame 7 Frame 8

Frame 9 Frame 10 Frame 11 Frame 12

Frame 13 Frame 14 Frame 15 Frame 16

Frame 17 Frame 18 Frame 19 Frame 20

Figure 5.9. Curves representing the posteriors $p(X_t|y_{1:T})$ superimposed on the observations y_t. Posteriors obtained using forward-backward smoothing.

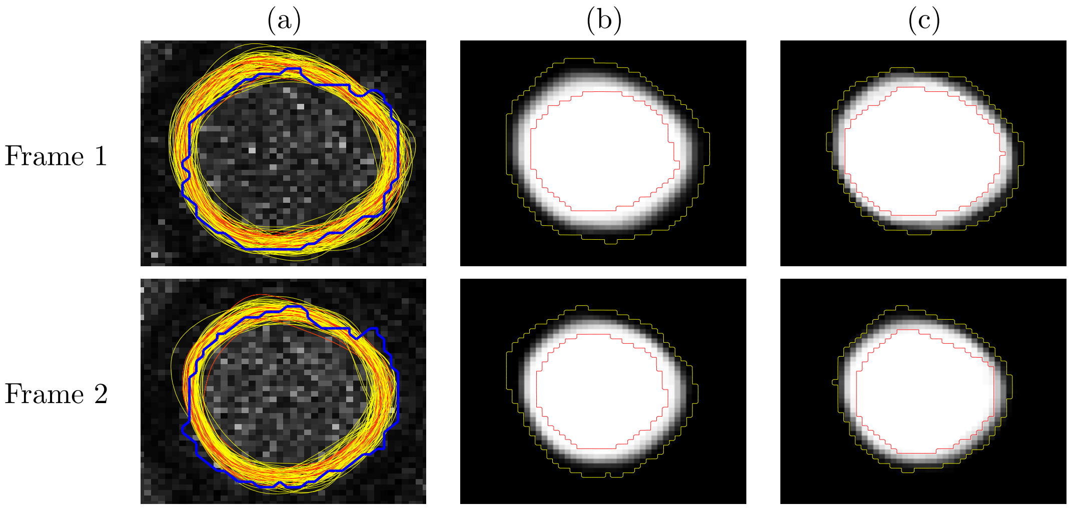


Figure 5.10. Comparison between filtering and Doucet's forward-backward smoothing on low SNR data. (a) Samples of the filtering posterior $p(X_t|y_{1:t})$ are shown in yellow, while samples of the smoothing posterior $p(X_t|y_{1:T})$ are shown in red. The manually-segmented truth is shown in blue. (b) Image which illustrates the distribution of the filtering samples, with the two curves (yellow and red) indicating the largest and smallest curves in the samples and the gray scale variations showing the frequency that a given pixel is contained within the curve samples (black indicating zero occurrences and white indicating that the point is contained in all of the sampled curves). (c) Image which shows the distribution of the smoothing samples using the same method as in (b), with the two curves again indicating the largest and smallest curves.

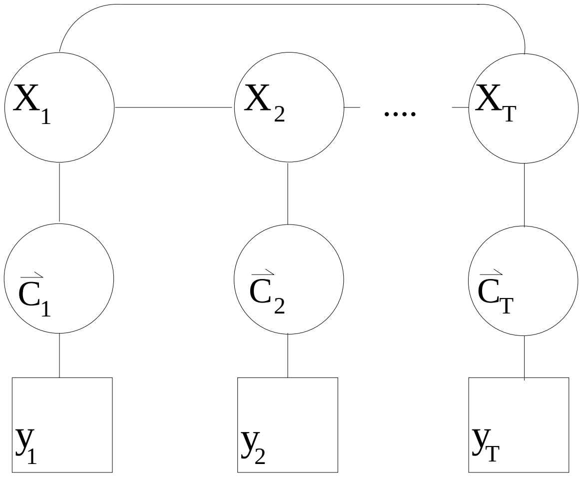


Figure 5.11. Graph with a single loop with dependencies between the state X_t and curve $\vec{C}_t$ as well as between the curve $\vec{C}_t$ and the observation y_t.

studies have shown that the local BP update equations can be used for accurate approximate inference, especially when the cycle length is long [59, 120, 187]. This result is quite intuitive because for long loops the effect of closing the loop at the ends on a simple chain will have little effect on nodes which were previously in the middle of the chain.[3] Theoretical work has established bounds on the accuracy of BP in loopy graphs [185, 195]. Furthermore, for jointly Gaussian random variables, it has been shown that, when it converges, BP converges to the correct posterior means [189]. For these reasons, we believe that a form of BP will produce accurate marginals on our single cycle graph. In particular, since our problem involves non-parametric densities, we use loopy NBP to approximate the solution to our problem.

5.2.1 NBP

The single cycle graph formulation lends itself well to applications which are periodic. In such problems, we learn the forward and backward dynamics of the system and use them in the approximate method of loopy NBP to estimate the posterior densities at each frame. In particular, we approximate X_t at each node and then determine the appropriate LV boundary estimate $\vec{C}_t^*$ using curve evolution as described before. Again, we define the observation potentials by

$$\psi_t(X_t, y_t) \equiv p(y_t | f_t^*, \vec{C}_t^*) p(f_t^* | \vec{C}_t^*) p(\vec{C}_t^* | X_t) p(X_t) \tag{5.18}$$

and the pair-wise potentials as

$$\psi_{t,\tau}(X_t, X_\tau) \equiv \frac{p(X_t, X_\tau)}{p(X_t) p(X_\tau)}, \tag{5.19}$$

where each node has exactly two edges.

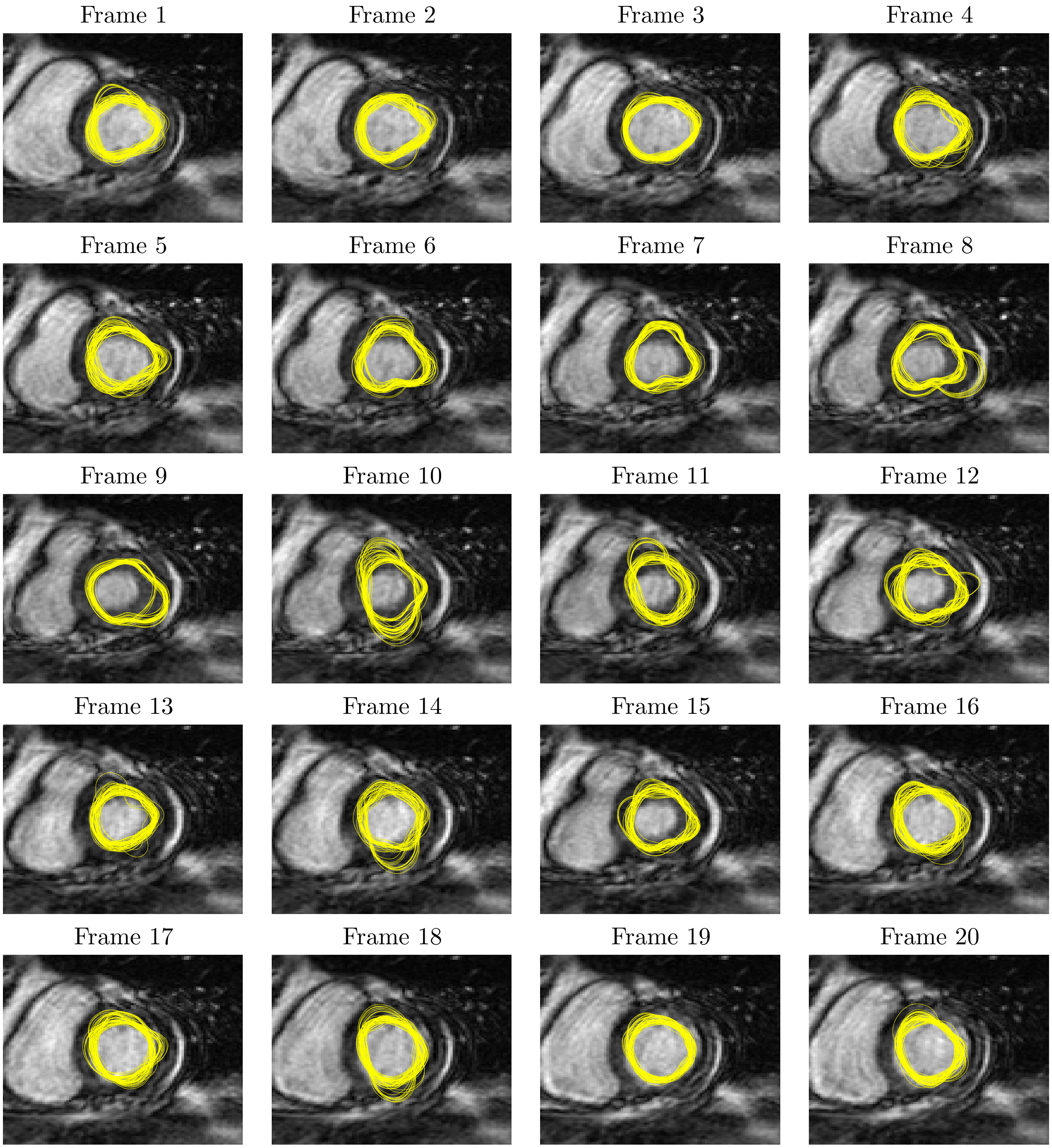
5.2.2 Results on a Single Cycle Graph

Figure 5.12 shows equally-weighted samples of the posterior obtained using NBP on high SNR data while Figure 5.13 shows the same based on low SNR data. As with previous experiments, the posterior based on the high SNR observations tend to have lower variance than that based on low SNR data. Unlike previous experiments, the loopy graph formulation induces greater correlation between states at time $t = 1$ and $t = 20$.

Figure 5.14 shows the segmentations obtained using loopy NBP on the original MR data. Figure 5.15 shows the segmentations on noisy data. The results are in general very accurate. For high SNR data, the dice boundary coefficient is 0.9292, while that for low SNR data is 0.9069. The results show an improvement of estimates over that from fixed-interval smoothing, mainly near the beginning and end of the cardiac cycle, as one may expect. Table 5.4 illustrates this fact by presenting the dice boundary

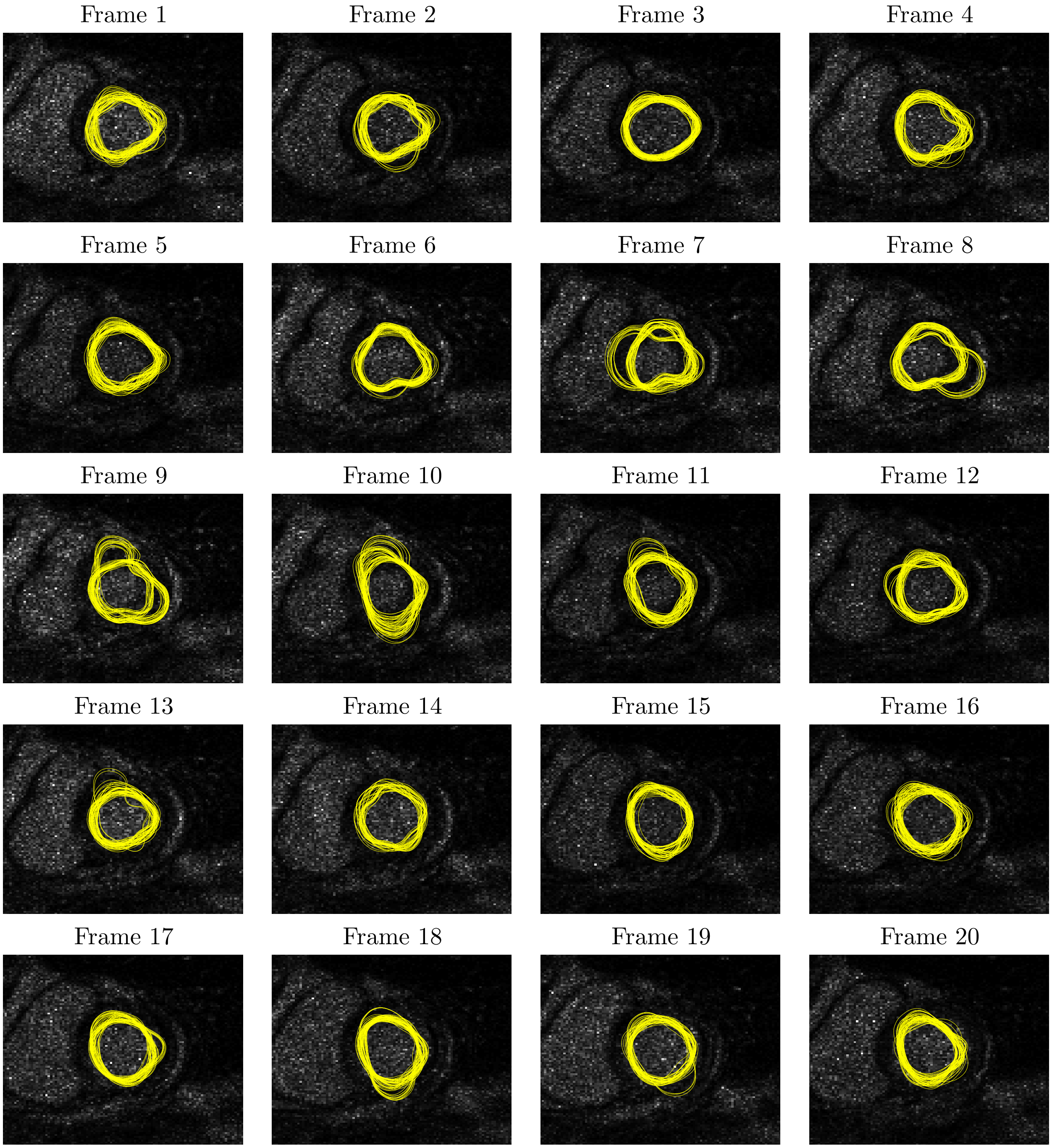
---

[3] It was also shown for discrete single cycle graphs that an adjustment could be made to the result from BP to obtain the correct marginals. [188]



Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 5.12. Samples of the posterior $p(X_t|y_{1:T})$ obtained from loopy NBP on a cardiac MR image sequence where observations have high SNR.



Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 5.13. Samples of the posterior $p(X_t|y_{1:T})$ obtained from loopy NBP on a cardiac MR image sequence where observations have low SNR.

Frame Number	Static - Shape Prior	Smoothing by Forward-Backward	Loopy NBP
1	0.7219	0.8774	0.9316
2	0.7521	0.8876	0.9375
3	0.8057	0.8547	0.8891
4	0.7565	0.8199	0.8703
5	0.7995	0.8795	0.9137
6	0.7960	0.8760	0.9281
7	0.8383	0.8788	0.8791
8	0.7558	0.8129	0.8304
9	0.7898	0.8830	0.9031
10	0.8754	0.9335	0.9387
11	0.8533	0.9336	0.9379
12	0.8448	0.8997	0.8831
13	0.7954	0.9402	0.9487
14	0.7641	0.8531	0.8436
15	0.6925	0.8428	0.8842
16	0.8267	0.8761	0.8751
17	0.8692	0.9190	0.9498
18	0.8490	0.9041	0.9598
19	0.8338	0.8822	0.9032
20	0.8491	0.9178	0.9302
Mean	0.8035	0.8836	0.9069

Table 5.4. Dice boundary coefficient across a cardiac cycle on low SNR data for static segmentation using a shape prior, smoothing using the forward-backward method, and loopy NBP. Note how the coefficient near the ends of the cardiac cycle show the most improvement between smoothing and loopy NBP. The mean boundary coefficients aggregate over the entire cardiac cycle.

	Static - Shape Prior	Filtering PTI	Smoothing by NBP	Smoothing by Forward-Backward	Loopy NBP
High SNR	0.8516	0.8654	0.9214	0.9210	0.9292
Low SNR	0.8035	0.8210	0.8909	0.8836	0.9069

Table 5.5. Dice boundary coefficient of LV boundary estimates based on static segmentation, a PTI filter, smoothing using the forward-backward method and NBP, and loopy NBP. (PTI = piecewise time-invariant)

coefficient for each frame[4] in a cardiac cycle on the low SNR data for loopy NBP, smoothing by the forward-backward method, and static segmentation using a shape prior. Meanwhile, Table 5.5 summarizes the dice boundary coefficient of the estimates from all of the different methods described in this chapter, in Chapter 4, and from the shape prior-based static segmentation. From these statistics, we observe that estimates are most accurate using the single cycle graphical model.

■ 5.2.3 Comparison with Static Segmentation using a Shape Prior

As in the previous chapter, we want to demonstrate the benefits of our approach by comparing the segmentation results with a static segmentation. Here, we compare only with the static segmentation using a shape prior, the more advanced of the two static approaches, because it is more accurate. Figure 5.16 shows segmentations obtained from loopy NBP on low SNR data (yellow curves) as well as those obtained from static segmentation with a shape prior (red curves) superimposed on the ground truth (green curves) and the low SNR data. Visually, one can observe that the estimates from our approach are closer to the ground truth than those from the shape prior-based static segmentation. Quantitatively, the loopy NBP results are better (e.g. 0.9292 vs. 0.8516 for high SNR data and 0.9069 vs. 0.8035 for low SNR data). Therefore, we have seen that loopy NBP does better than Markov chain smoothing (using either of the two approaches), which in turn does better than filtering, which is better than static segmentation. The numerical results are summarized in Table 5.5.

■ 5.3 Summary

In this chapter, we have generalized the filtering problem from Chapter 4 first to a smoothing problem and then to the problem of inference on a graph with a single loop. For the former, we have applied two different methods of particle smoothing to obtain estimates of the posterior density. For the latter, we have used approximate inference to estimate the posteriors. The single loop assumption is possible because of the quasi-periodic nature of the LV boundary.

---

[4] Instead of averaging the dice coefficient across the cardiac cycle and across patients, the computations for this table only average across patients to break down the accuracy at different points in the cardiac cycle.

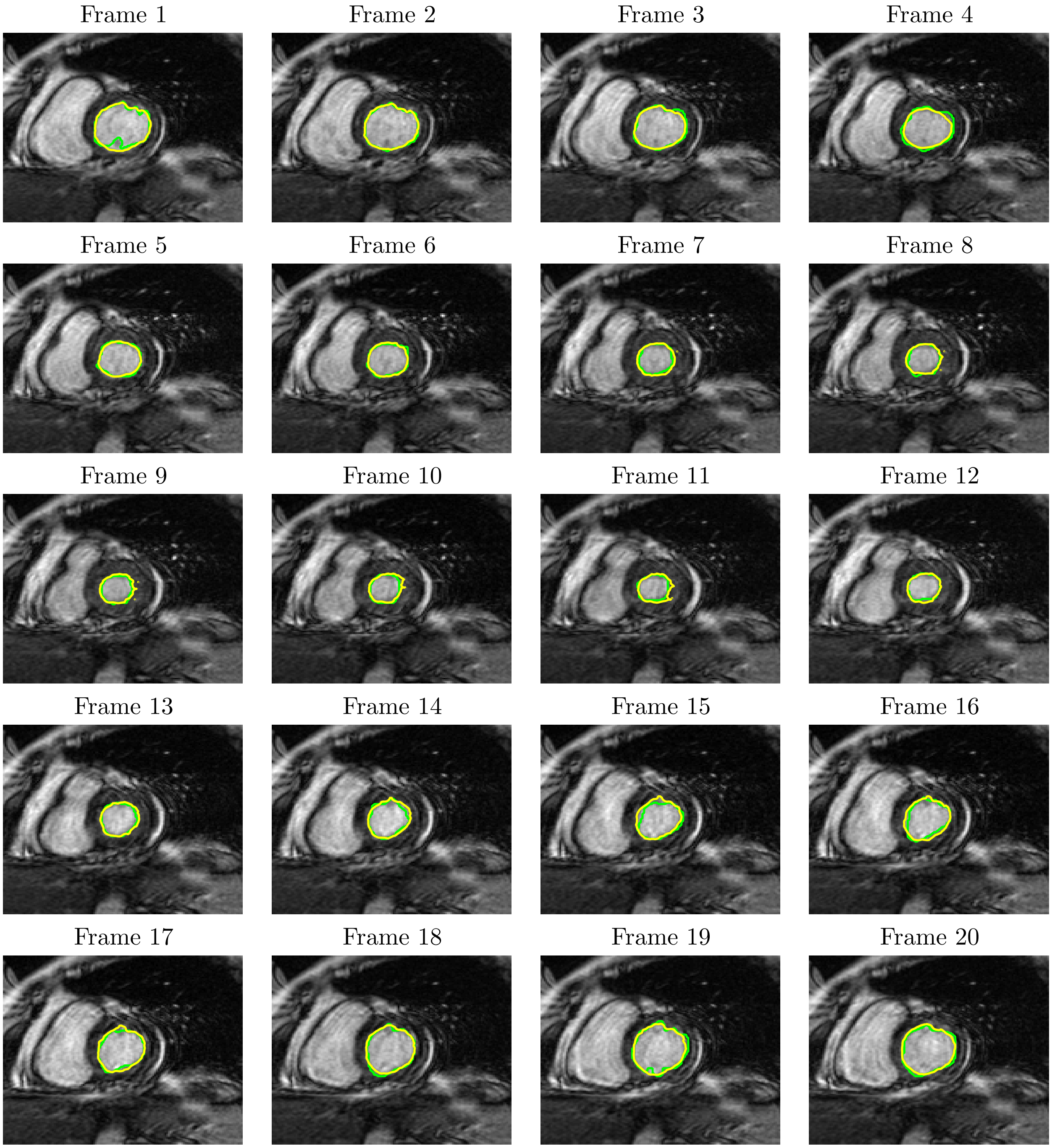
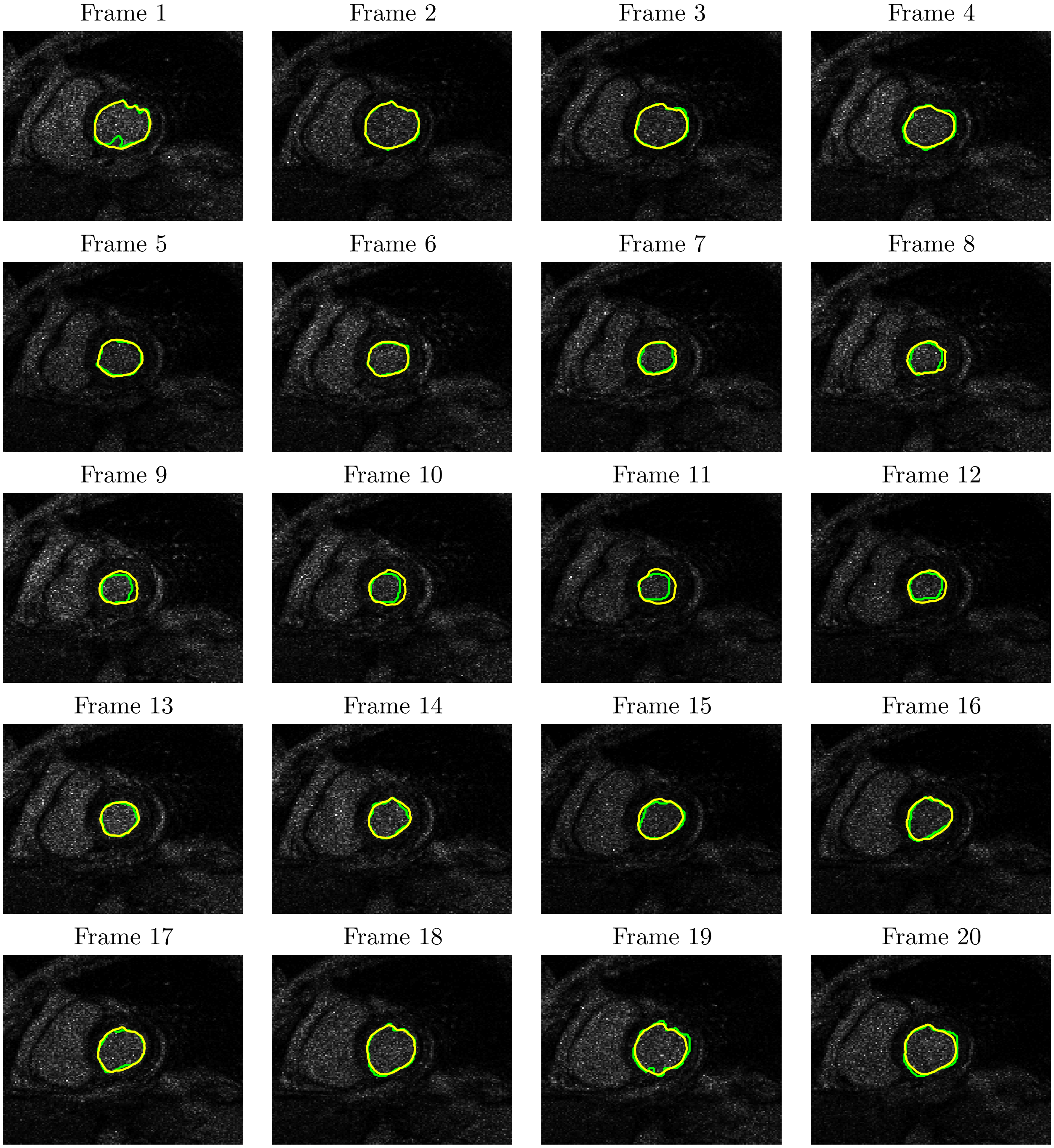


Figure 5.14. Segmentations (yellow) obtained from loopy NBP on a high SNR cardiac MR image sequence. Ground truth shown in green.



Frame 1 Frame 2 Frame 3 Frame 4

Frame 5 Frame 6 Frame 7 Frame 8

Frame 9 Frame 10 Frame 11 Frame 12

Frame 13 Frame 14 Frame 15 Frame 16

Frame 17 Frame 18 Frame 19 Frame 20

Figure 5.15. Segmentations (yellow) obtained from loopy NBP on a noisy cardiac MR sequence. Ground truth shown in green.

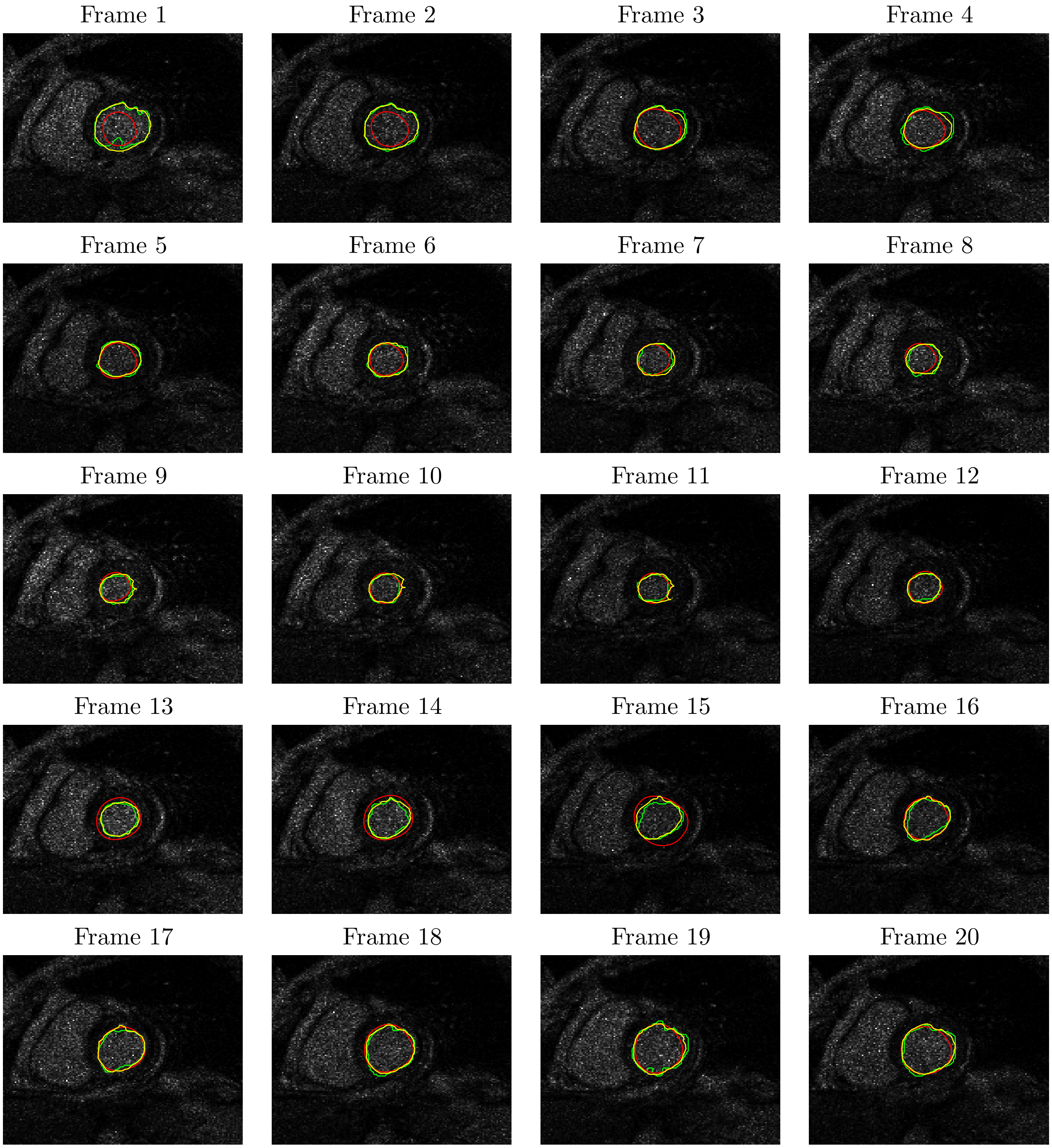

Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16
Frame 17
Frame 18
Frame 19
Frame 20

Figure 5.16. Comparison of the LV boundary estimates obtained from a loopy graph formulation (yellow) (as shown in Figure 5.15) with static segmentations using a shape prior (red). Ground truth is shown in green.

The smoothing formulation increases the amount of information which can be used to estimate the state at all frames. This is particularly useful near the beginning of the chain because in the filtering process very few observations are used in the estimation problem. From the results shown in this chapter we can see that the segmentations obtained through the smoothing formulation are closer qualitatively and quantitatively to the LV boundary.

The single cycle graph formulation lends itself well to LV boundary estimation because the cardiac cycle is periodic. Using a simple chain, LV boundary estimates at the beginning and end are not strongly coupled. Adding the dependency between the first and last frames of the cardiac cycle provides a better model for the system. As observed from the results, treating the LV as a loopy graph leads to improvements over filtering and smoothing on a chain.

Chapter 6

Contributions and Suggestions for Future Research

In this thesis, we have examined methods for object segmentation coupled with field estimation in both static and dynamic environments. In the latter, we have proposed a technique for learning the dynamics of deformable shapes and used this in the recursive estimation of object boundaries. We summarize the contributions in Section 6.1 and conclude with a discussion of future research topics in Section 6.2.

■ 6.1 Thesis Contributions

The major contributions of the thesis are summarized in this section.

■ 6.1.1 Incorporation of Spatially Varying Statistics in Joint Field and Boundary Estimation

The first major contribution of this portion of our research is the incorporation of spatially varying statistics in the joint field and boundary estimation problem. This approach is shown to yield improved estimates over existing methods such as Mumford-Shah when the field model has spatial variability. In particular, we have shown that applying our modified Mumford-Shah approach to simultaneously estimating the underlying sea surface temperature field in the North Atlantic Ocean as well as the location of the Gulf Stream's north wall results in more accurate estimates than Mumford-Shah.

Both Mumford-Shah and our approach trade off three components: the fidelity of observations to the field estimate, the field prior on either side of the boundary, and smoothness of the boundary itself. While the boundary smoothness and data fidelity components are the same, the field prior for Mumford-Shah involves a gradient penalty term, one that encourages a smooth field estimate. On the other hand, our approach incorporates a field prior that is based on general first and second order statistics of the field. One can think of modified Mumford-Shah as a generalization to Mumford-Shah.

In addition to formulating the problem, we also derive the curve evolution equation necessary to evolve the curve. This contribution allows the modified Mumford-Shah approach to be practically implemented. The minimization is done using coordinate descent on the two variables, the underlying field and the boundary estimate.

Furthermore, we introduce both our method and Mumford-Shah to geoscience problems that previously were solved using traditional interpolation methods. In particular, we compare these joint field and boundary estimation methods with methods of kriging, gradient smoothing, and smoothing splines on both sea surface temperature data as well as soil moisture maps, illustrating how the joint approach avoids the problem of smoothing across discontinuities.

■ 6.1.2 Learning the Dynamics

Another contribution of this thesis is a method for learning the dynamics of a deformable boundary. The technique involves using an information-theoretic approach to find a maximally-informative statistic. Using this statistic, we produce non-parametric estimates of conditional densities (both forward and backward). Since the approach is purely statistical, one can obtain an estimate for the dynamics given only training data (no physical model or prior information is needed). Thus, instead of using complex mathematical equations to model the low-level physics inherent in a physical system, we raise the level of abstraction to coherent structures whose dynamics are learned.

During the process of learning, we have developed a method to model complex, deformable structures using a low-dimensional state representation. In particular, in the problem of segmenting the left ventricle (LV) across a cardiac cycle, the full representation of the LV boundary is a difficult, high-dimensional problem. For computational tractability, we choose a low-dimensional state representation which captures the salient parts of the object's evolution. This state is recursively estimated using sample-based methods and, based on this information, the best estimate of the boundary is obtained using curve evolution.

■ 6.1.3 Recursive Estimation of the State

The final major contributions of the thesis involve combining curve evolution with sample-based methods for recursive state estimation, extending the problem to Markov chain smoothing and proposing two techniques to arrive at estimates, and formulating a single cycle graphical model to exploit the quasi-periodicity of the LV. In addition, through the approximation of densities using a sample-based approach, we essentially have taken the traditional method of particle filtering and applied it to a complicated shape space.

Combining Curve Evolution with Sample-Based Methods

In our framework, we combine curve evolution with a sample-based method. In particular, we choose a low-dimensional state representation which attempts to capture the temporal evolution of the boundary. Given a particle-based estimate for the prediction density for the state, we apply curve evolution to determine the best estimate of the boundary. By using this approach, computing the system dynamics is made tractable because of the low-dimensionality of the state. Furthermore, the incorporation of the curve evolution step allows us to obtain reasonably accurate estimates of the boundary.

Solutions Using Fixed-Interval Smoothing

We apply two approaches to solve the smoothing problem. First, we use non-parametric belief propagation (NBP) to determine the segmentations. Alternatively, we incorporate the forward-backward approach of Doucet *et al.* [51] to determine the segmentation. For the latter method, we use the results from filtering for the forward step and apply a re-weighting to the sample points (representing the posterior densities) in the backward step of the algorithm to obtain the final representation of the posteriors which are used to determine the segmentations. The results show that smoothing improves the segmentation obtained from filtering or a static approach.

Single Cycle Graph to Exploit Quasi-Periodicity

To incorporate the quasi-periodic nature of the LV, we next pose the segmentation problem as a graphical model having a single cycle loop. To accomplish this, we take the Markov chain and add an edge between the first and last nodes of the graph. We use NBP, a generalized version of BP, to approximate the non-parametric posterior densities. The results show an improvement over fixed-interval smoothing, with the main difference, as expected, near the beginning and end of the cardiac cycle.

■ 6.2 Suggestions for Future Research

The general framework proposed for recursively estimating deformable boundaries can be extended in a number of ways. We describe each of the possible directions for future research below, divided into a section for algorithmic advances and another for different application domains.

■ 6.2.1 Algorithmic Advances

The methods proposed in Chapters 4 and 5 provide reasonable results to the temporal segmentation of the LV. Throughout the thesis, however, we have made specific choices in the framework for purposes of computational simplicity, convenience, or limitations due to scope. Here, we discuss some alternatives which may be considered that could potentially yield better results.

Different Choice of Dynamics

Many other forms of the statistic for the dynamics may be considered. In our work, we restrict the maximally-informative statistic to be linear. More general non-linear functions may be considered, as they could provide a more informative statistic albeit at the cost of greater computational complexity. Any such change can easily be made and inserted into our framework.

State Augmentation

In addition to shape and area, other information may be added to the state to improve the learning of the dynamics. In particular, we can add components such as time, ejection fraction, or cardiac phase into our state representation.

By adding the time element, we can more accurately determine where we are within the cardiac cycle. In particular, we can consider this as a relative time component which allows us to align the systolic and diastolic phases across different patients (e.g. instead of patients reaching end systole between time frames 6 and 12, we can have the relative time element at end systole be the same for all patients). In essence, this component compresses or expands (accordingly) the dynamics based on how fast or slow a particular test subject transitions from end diastole to end systole (and vice versa). Similar but simpler than adding a time component, we could add a variable which designates which cardiac phase we are in. The addition of this variable provides a way to provide more accurate information regarding which dynamic to use. By doing this, we can substitute our mixture dynamics which have fixed empirical ratios with dynamically estimated ratios. Alternatively, augmenting the state with the ejection fraction allows us to adjust the learning based on the amount of constriction expected during a cardiac cycle. For instance, if we have a test sequence from an unhealthy patient with a small ejection fraction, having this information can allow us to determine the appropriate dynamics.

Incorporating Additional Measurements for Segmentation

Previous work has fused multiple sets of data to cardiac imaging, but the data were all two-dimensional views of the heart [136]. Instead of different views, one could incorporate other information such as electrocardiogram (EKG) data with the MR image. Currently, EKG information allows us to know when the cardiac cycle begins (at end diastole). However, we have not considered the possibility of the EKG signal providing additional information throughout the cardiac cycle.

Segmentation for Abnormal Cardiac Data

The training sets were all from healthy patients. As a result, the test sequences were also based on healthy patients. Practically speaking, however, being able to automatically segment abnormal cardiac data is desirable. One issue with learning the dynamics of abnormal cardiac data is that individual patients often exhibit different irregular behavior. If the specific abnormalities are labeled, it might be possible to train based on the specific classification of abnormality.

Alternatively, if we assume that an ill patient makes periodic return visits, we can consider trying to learn the dynamics of the individual patient and apply it to future scans of the same individual. For example, suppose a patient receives an initial MR scan to determine his condition. Segmentations of the LV can be made from this high quality scan (manually if not automatically) and used to learn the dynamics. This information can be used to aid the segmentation of future scans of this patient.

Since MR scans are expensive, the physician can regularly image a patient with less expensive scans such as ultrasound imaging. Given knowledge of this patient's specific dynamics, we can attempt to recursively segment the LV in a much lower quality set of data.

Incorporating Physical Models

One might be interested in explicitly incorporating physical constraints to the dynamic system. In the thesis, we use statistics alone to determine the dynamics of the system. On the other end of the spectrum, models are often fully-represented by physical models. For the LV, biomechanical models have been analyzed and are well-understood [116–118,127,139,157,161,162]. Thus, combining our purely statistical model with a physical model such as the biomechanics of the LV may be a worthwhile and interesting extension to our work. Doing so may improve the dynamic model, eliminating predictions which are known to be physically impossible, for instance.

Fully incorporating biomechanics may be a complicated process. However, we do not need to include the entire model to achieve improvements. For instance, in our state representation we have both area and shape information. If we had a partial differential equation (PDE) based on physics which could estimate the area in a very accurate way, we can combine this information with the existing learning of the shape parameters to produce a dynamic model based on a combination of statistics and simple physics.

Extension to Tracking of Physical Masses

The methods introduced in the thesis provide us with a way of segmenting a deformable object over a sequence of time. In solving the segmentation problem, we estimate the change in the boundary over time but do not track the exact movements of any point on the boundary (e.g. rotations of the object do not yield any change in what we estimate to be the boundary). While this is sufficient for the problem we address, there exist other problems where knowing the exact motion of different parts of the boundary is of interest. Computing the flow fields or determining the velocity of particles are different methods used in the process of solving the correspondence problem [8, 80]. For the left ventricle, the use of a biomechanical model as described above can provide information regarding the exact motion of individual parts of the boundary. Work has been done to track the motion of the cardiac using techniques such as tagged MR imaging [97, 98, 116, 138].

Different Shape Representations

In the thesis, we have chosen one type of representation for the shapes when we applied principal components analysis (PCA) on signed distance functions. Other possibilities exist using the same PCA-based approach. For instance, instead of a single basis across the entire cardiac cycle, we could compute different bases on subintervals.

In addition, individuals have considered other ways to represent curves. For example, Fletcher *et al.* [57] consider principal geodesics for their statistical shape analysis, Cootes *et al.* [35, 36] use coordinates from image features in an active shape model, Keleman *et al.* [95] and Senegas *et al.* [156] represent their curves using a spherical harmonic decomposition, and Jaggi *et al.* [83] use an ellipsoid to model the shape of the LV. These alternate representations can be substituted easily into our general framework.

Alternate Models for Likelihood and Priors

In Chapter 4, we made specific choices for the likelihood and field prior terms. We can certainly consider alternate choices for these terms. For example, given the data available, we made the assumption that the mean field intensities are constant throughout a cardiac cycle. If alternate models for a time-varying intensity are known, we can easily substitute this information into the framework.

Extension to 3-D

Theoretically, we can easily extend our framework to 3-D segmentation across a cardiac cycle since nothing in the formulation restricts us to 2-D. Cardiologists can benefit from 3-D segmentations as they would be able to fully examine how a cardiac volume evolves over time. With our 2-D analysis, we could stack the slices together and interpolate to create a 3-D volume, but having automatic 3-D segmentations would be more accurate. In addition, in 3-D, we don't have to learn different models for every view. However, the main issue is the dramatic increase in computational complexity inherent with adding a third dimension to the problem.

■ 6.2.2 Different Application Domains

Learning the dynamics through statistics as discussed in the thesis can be extended to several different application areas. First, a conceptually straight-forward extension is to apply the LV estimation problem on other imaging modalities such as ultrasound. Next, extending the problem to other medical applications may be considered.

In Chapter 3, we performed joint boundary and field estimation on sea surface temperature data. In this problem, we located an oceanic front in a static framework. Given enough training data, the dynamics of oceanic fronts can be learned and the segmentation of these fronts can be performed recursively over time.

Extensions which may be more challenging include learning the dynamics in applications such as submarine channel detection of seismic data and loss estimation for actuarial pricing of medical malpractice. We discuss each of these extensions below.

Recursive Estimation of LV in Other Modalities

For the segmentation of the LV, our algorithm is applied to high quality MR data and simulated noisy data. The latter was used to show that the framework proposed performs well for low quality images. From these results, it appears promising that

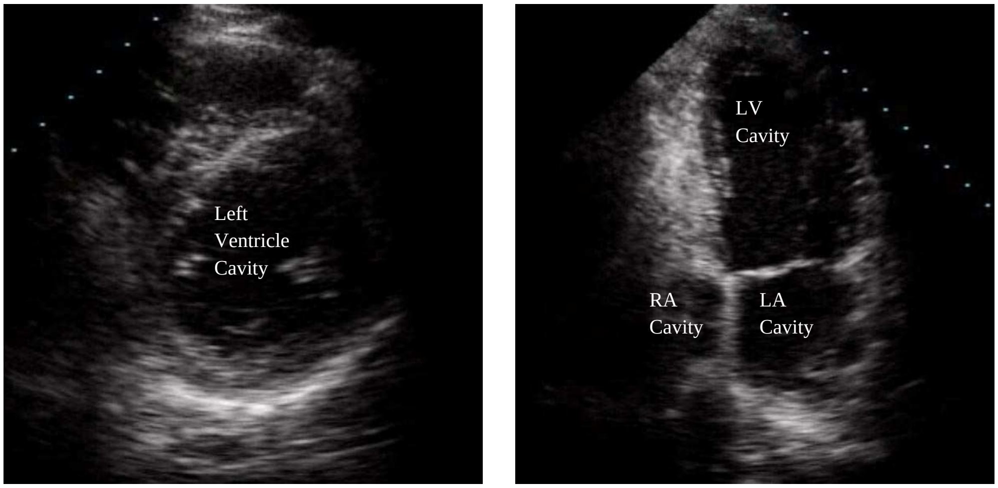


Figure 6.1. Sample cardiac ultrasound images (courtesy of Collin Stultz, West Roxbury VA Hospital) with relevant features labeled (LA - left atrium, RA - right atrium, LV - left ventricle). Unlike in MR data, the organ boundaries and myocardial features cannot be detected easily from a static frame.

our framework can be used in other imaging modalities, particularly ones with much lower SNR such as cardiac ultrasound. Figure 6.1 shows examples of labeled cardiac ultrasound images. Good segmentation of such images requires additional information beyond the current observation. The predictions produced from our framework provide a source of such prior information to aid the segmentation of the LV.

In addition, ultrasound data are usually imaged over a variety of views. If training data were sufficient enough to adequately learn the dynamics across multiple views (or over 3-D), our formulation can be useful and valuable for LV segmentation. In particular, given the view, we can determine the appropriate dynamics to use and thus provide a good prediction for the boundary estimation.

Subcortical Structures and Other Human Organs

Besides the area of cardiac imaging, other medical application areas can use the framework proposed. In medical imaging, segmentation of subcortical structures of the brain such as the hippocampus, thalamus, and putamen or that of the intracranial boundary is quite useful for radiological diagnosis and surgical planning as well as clinical applications (e.g. computer-aided neurosurgery or radiotherapy) [4, 79, 95, 153]. This problem is primarily a static one.

However, since most brain segmentations involve a sequence of two-dimensional slices, we can reformulate our problem so that the temporal aspect is replaced by the third dimension of scanned data. Using such an assumption, we can learn the deformations of two-dimensional brain contours from one slice to the next in much the same way we learned the evolution of the LV over time.

In addition, since our particle-based framework allows for multi-modal densities, lesion detection in multiple sclerosis is another possible application. As described in Chapter 4, lesions appear and disappear periodically with their occurrence and frequency being important factors in the diagnosis of a patient. We can accurately model this behavior with non-parametric density estimates. For instance, we can estimate the probability density of lesions with a bimodal distribution in which one mode assumes a single lesion while the other represents multiple lesions.

Damage Assessment for Insurance Companies

In Chapter 3, we examined the problem of jointly estimating the field and boundary in earth science applications. The methods proposed can be applied to other applications that have distinct regions in the data and require interpolation given an incomplete or sparse set of measurements. One possible application is the estimation of hurricane or other severe weather damage [183]. In such problems, given a data set of damage reports at a small set of locations, there is a desire to locate the boundary between regions of severe and less severe damage for determination of evacuation and protection of the affected region. Furthermore, insurance companies can use these methods to quickly obtain an estimate of the damage throughout the field. So, the techniques we have developed can be used to estimate the boundary as well as the surrounding field.

Rainfall Estimation

Similarly, we can apply the joint method on the estimation of rainfall across a region. The amount of rainfall is known to be spatially varying across a region. Currently, sparse sensors are used in conjunction with satellite data to estimate rainfall. The information available is similar to the sea surface estimation problem, so we believe that our variational approach can be applied to this problem in a straight-forward manner.

Recursive Estimation of Oceanic Fronts

The joint boundary and field estimation of Chapter 3 was posed in a static framework. Similar to the work done on the LV, we could formulate the boundary estimation problem as a temporal one in which predictions can be obtained from prior estimates and a learned dynamic model. In particular, we could apply this to locating oceanic fronts. Like the LV, these fronts have annual quasi-periodicity and different dynamics can be learned in the different seasons of each year.

Submarine Channel Detection in Seismic Data

On the seafloor of a body of water, older layers of sediment deposits are buried over time by new layers of deposits. Geologists who study the soil can create a historical timeline of the seafloor by examining buried sediment deposits. Of particular interest are submarine channels formed by turbidity currents through the erosion and deposition of organic materials. Accurate segmentation of these channels is of interest because dense amounts of organic deposits tend to collect around the bends of these channels. The pressure caused by the new deposits burying the old deposits causes an increase in temperature, thus *cooking* the organic deposits and converting them into oil or natural gas [148]. Therefore, the knowledge of the location of these channels helps to locate areas which may have a high concentration of natural resources.

We can learn the spatial dynamics of the channels based on a training set of expert segmentations (made by interpreters). Then, given the segmentations for some slices of test data, we can apply the dynamics to obtain predictions which can help estimate the location of spatially-neighboring slices within the test set.

■ 6.3 Concluding Remarks

In closing, we have proposed algorithms which incorporate spatial and temporal statistics to improve the image segmentation problem. In the process, we have extended the concept of particle filtering to objects residing in a complex shape space. In addition, we have introduced an approach for boundary estimation which combines belief propagation on a low-dimensional state representation with a curve evolution process which refines the estimate based on the data. We have used a purely statistical model to learn the dynamics, thus providing a way to determine the system dynamics even in the absence of an existing physical or mathematical model.

Appendix A

Different Optimal Estimators

In this appendix, we provide a brief discussion of three common *optimal* state estimators, the MAP estimate, the minimum mean-squared error (MMSE) estimate, and the linear least-squares error (LLSE) estimate. Each case depends on the choice of the cost function C in

$$\mathrm{C}ost(\mathbf{f}^*) = \sum_{\mathbf{f}} C(\mathbf{f}, \mathbf{f}^*) p(\mathbf{f}|\mathbf{y}), \tag{A.1}$$

where the exact choice of cost is described briefly below.

MAP Estimate

The MAP estimate has as a cost function $C(\mathbf{f}, \mathbf{f}^*) = 1 - \delta(\mathbf{f} - \mathbf{f}^*)$, where δ is the Dirac delta function. In essence, all errors are equally bad. In this scenario, the best choice of $\mathbf{f}$ is the one which maximizes $p(\mathbf{f}|\mathbf{y})$, hence the name. Often to find $\mathbf{f}^*$, we maximize a monotonic function of the posterior, such as $\log p(\mathbf{f}|\mathbf{y})$.

MMSE Estimate

The minimum mean-squared error estimate, also known as the Bayes' least squares error estimate, employs a quadratic cost function; namely, $C(\mathbf{f}, \mathbf{f}^*) = ||\mathbf{f} - \mathbf{f}^*||_2^2$. Using this cost, $\mathbf{f}^*$ is found to be $E(\mathbf{f}|\mathbf{y})$ [193], the expected value of $\mathbf{f}$ given $\mathbf{y}$.

LLSE Estimate

The MMSE estimate is generally a non-linear function of $\mathbf{y}$. The actual computation of this non-linear estimate may be expensive, or there may be times when the full statistical characterization of the posterior is unavailable. In such instances, we may be interested in finding the LLSE estimate, a special case of the MMSE estimate where the estimator is a linear function of $\mathbf{y}$. The computation of the LLSE relies only on the joint second-order statistics of $\mathbf{f}$ and $\mathbf{y}$. The LLSE estimate is also of interest for the special class of Gaussian random variables. Since joint Gaussian random variables are fully-described by the joint second-order statistics, the MMSE estimate of Gaussian random variables is equivalent to its LLSE estimate.

Appendix B

Kriging

In this appendix, we provide a brief discussion on the method of kriging, a method of smoothing used in the earth sciences for interpolation. We define the data as y and the underlying random process as f. In addition, we assume that the field is second order wide-sense stationary [91]. So, for two points m and n, the covariance between $f(m)$ and $f(n)$ only depends on the difference $m - n$ (i.e. $cov(f(m), f(n)) = C(m - n)$ for some function C). Define the variogram γ as

$$\gamma(m-n) \equiv \frac{1}{2} var(f(m) - f(n)). \tag{B.1}$$

Several common models for the variogram (see [72]) are logarithmic, exponential, Gaussian, quadratic, and sinc functions. Kriging is referred to as an exact interpolator when the observation is assumed to not be noisy. In this case, only the unobserved regions are estimated.

Interpolation of Unobserved Regions Using Ordinary Kriging

In ordinary kriging [90], the spatial mean is assumed constant. The variogram is defined to be

$$\gamma(m-n) \equiv \frac{1}{2} E[(f(m) - f(n))^2]. \tag{B.2}$$

For a given point s_0, we wish to estimate $f(s_0)$ given observations at s_i for $i = 1, 2, \ldots, N$. In ordinary kriging, the optimal predictor $\hat{f}(s_0)$ minimizes the mean-squared prediction error $E[(f(s_0) - y(s_0))^2]$ over the class of linear predictors where

$$f(s_0) \equiv \sum_{i=1}^{N} \lambda_i y(s_i) \tag{B.3}$$

with $\sum_{i=1}^{N} \lambda_i = 1$, so each interpolated point is a weighted sum of its neighbors. Using Lagrange multipliers, the minimization problem becomes

$$\min_{\lambda_1, \lambda_2, \ldots, \lambda_N} E((f(s_0) - \sum_{i=1}^{N} \lambda_i f(s_i))^2) - 2r(\sum_{i=1}^{N} \lambda_i - 1), \tag{B.4}$$

where r is the Lagrange multiplier. The term inside the expectation can be written as

$$(f(s_0) - \sum_{i=1}^{N} \lambda_i f(s_i))^2 = \sum_{i=1}^{N} \lambda_i (f(s_0) - f(s_i))^2 - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j (f(s_i) - f(s_j))^2. \quad \text{(B.5)}$$

So, using the definition from Equation (B.2), Equation (B.4) can be rewritten as

$$2 \sum_{i=1}^{N} \lambda_i \gamma(s_0 - s_i) - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \gamma(s_i - s_j) - 2r(\sum_{i=1}^{N} \lambda_i - 1). \quad \text{(B.6)}$$

Differentiating this equation with respect to λ_i, we obtain

$$2(\gamma(s_0 - s_i) - \sum_{j=1}^{N} \lambda_j \gamma(s_i - s_j) - r) = 0. \quad \text{(B.7)}$$

Solving this set of linear equations, we can find the optimal λ_i's that provide a prediction for $f(s_0)$.

Note that in this analysis, we implicitly assume that the observations have no measurement error. So, if $s_0 = s_j$ for some j, $f(s_0) = y(s_j)$. Stated another way, the estimate at any point where there is an observation is the observation itself.

Estimation of Entire Region Assuming Noisy Observations

Instead of interpolating only over unobserved regions, kriging can be generalized to the problem of estimating the entire field. For this problem, the observation $y(s_i)$ is now noisy, with the noise having zero mean and variance σ_n^2. To estimate the underlying field $f(s_0)$, we write it as a linear combination of the noisy observations

$$f(s_0) = \sum_{i=1}^{N} \lambda'_i y(s_i) \quad \text{(B.8)}$$

and solve Equation (B.7) with λ' replacing λ, and $\gamma(0) = \sigma_n^2$ (previously zero). The introduction of measurement noise is called the nugget effect in the earth sciences. In our analysis, we will use this form to perform kriging interpolation.

Appendix C

First Variation of Region Integrals

In Section 3.3.2, we state in Equation (3.6) the curve evolution equation for the curve $\vec{C}$ in the coordinate descent step when the other variables f, μ, and b are fixed. The direction of flow implied by this equation corresponds to that which allows the objective functional of Equation (3.3) to decrease most rapidly. Known as the gradient flow for $\vec{C}$, it is often written as $\frac{\partial \vec{C}}{\partial t}$ or shorthand as $\vec{C}_t$, where t represents iteration time during the curve evolution step. This section of the Appendix provides a detailed derivation of the result stated in Equation (3.6).

We begin with some preliminary items. Define

$$Q_{R(\vec{C}(t))}(t) = \int\limits_{R(\vec{C}(t))} q_{R(\vec{C}(t))}(\mathbf{x}, t) d\mathbf{x} \tag{C.1}$$

with

$$q_{R(\vec{C}(t))}(\mathbf{x}, t) = \int\limits_{R(\vec{C}(t))} v_R(\mathbf{x}, \mathbf{z}) d\mathbf{z}, \tag{C.2}$$

where[1]

$$v_R(\mathbf{x}, \mathbf{z}) = \mathcal{D}[f(\mathbf{x}) - \mu_R(\mathbf{x})] b_R(\mathbf{x}, \mathbf{z}) \mathcal{D}[f(\mathbf{z}) - \mu_R(\mathbf{z})]. \tag{C.3}$$

In the body of the paper, we have used a shorthand notation to write the regions, which implicitly depend on the curve, as simply R. Throughout the Appendix, we will be more explicit to indicate the dependencies of these regions on the evolution of the curve. For instance, when we evolve $\vec{C}$, we fix f, μ, and b; so, in Equation (C.2), the integrand does not depend on $\vec{C}(t)$, but the region of integration does. In such a case, $\frac{\partial q_{R(\vec{C}(t))}(\mathbf{x},t)}{\partial t}$ can be written in the form of a line integral [196]. In particular,

$$\frac{\partial q_{R(\vec{C}(t))}(\mathbf{x}, t)}{\partial t} = \frac{\partial}{\partial t} \int\limits_{R(\vec{C}(t))} v_R(\mathbf{x}, \mathbf{z}) d\mathbf{z} = \int_{\vec{C}} < \vec{C}_t, v_R(\mathbf{x}, s) \vec{N} > ds. \tag{C.4}$$

Here, $\vec{N}$ represents the outward pointing normal vector with respect to region R.

---

[1] Since μ and b are fixed in this step, v_R does not depend on t.

Now, using the notation from Equation (C.1), we can write the MMS energy functional from Equation (3.3) as

$$E(f,\vec{C}) = \alpha \int\limits_{R_1(\vec{C}(t))\cap U} (f(\mathbf{x})-g(\mathbf{x}))^2 d\mathbf{x}$$

$$+\alpha \int\limits_{R_2(\vec{C}(t))\cap U} (f(\mathbf{x})-g(\mathbf{x}))^2 d\mathbf{x} + \beta(Q_{R_1(\vec{C}(t))}(t) + Q_{R_2(\vec{C}(t))}(t)) + \gamma \oint_{\vec{C}} ds. \qquad (C.5)$$

where $R_1(\vec{C}(t))$ and $R_2(\vec{C}(t))$ are the regions separated by $\vec{C}$,[2] $g(\mathbf{x})$ are the observations, and U is the subset of Ω where observations of $g(\mathbf{x})$ exist.

To find the curve evolution necessary to decrease $E(f,\vec{C})$, we compute the partial derivative of E with respect to t. Using the results derived by Kim *et al.* [102] and Delfour and Zolesio [47], who assert that

$$\frac{dQ_R(t)}{dt} = \int_R \frac{\partial q_R(\mathbf{x},t)}{\partial t} d\mathbf{x} + \oint_{\vec{C}} < q_R(\mathbf{x},t)\vec{N}, \vec{C}_t > ds, \qquad (C.6)$$

and applying the result of the derivation of Euclidean curve shortening by Grayson [73] ($\frac{\partial}{\partial t}\oint_{\vec{C}} ds = \oint < \vec{C}_t, \kappa\vec{N} > ds$, where κ is curvature), we obtain

$$\frac{\partial E(f,\vec{C}(t))}{\partial t} = \alpha \oint_{\vec{C}(t)} < \vec{C}_t, ((f_{R_2}(s)-g(s))^2 - (f_{R_1}(s)-g(s))^2)h(s)\vec{N} > ds$$

$$+\beta \oint_{\vec{C}(t)} < \vec{C}_t, (q_{R_2(\vec{C}(t))}(s,t) - q_{R_1(\vec{C}(t))}(s,t))\vec{N} > ds$$

$$+\beta[\int\limits_{R_2(\vec{C}(t))} \frac{\partial q_{R_2(\vec{C}(t))}(\mathbf{x},t)}{\partial t} d\mathbf{x} - \int\limits_{R_1(\vec{C}(t))} \frac{\partial q_{R_1(\vec{C}(t))}(\mathbf{x},t)}{\partial t} d\mathbf{x}] + \gamma \oint < \vec{C}_t, \kappa\vec{N} > ds, \quad (C.7)$$

where f_{R_i} represents the field estimate in R_i and $h(s)$ is an indicator function that takes the value of 1 when observations are present (i.e. when $s \in U$), and 0 otherwise. The negative signs arise from the fact that our outward normal $\vec{N}$ is taken with respect to R_1.

Applying Equation (C.4) to the two instances of $\frac{\partial q_{R_i(\vec{C}(t))}(\mathbf{x},t)}{\partial t}$ in Equation (C.7), and assuming that b is symmetric (i.e. $b(\mathbf{x},\mathbf{z}) = b(\mathbf{z},\mathbf{x})$, which implies $v(\mathbf{x},\mathbf{z}) = v(\mathbf{z},\mathbf{x})$), we determine that

$$[\int\limits_{R_2(\vec{C}(t))} \frac{\partial q_{R_2(\vec{C}(t))}(\mathbf{x},t)}{\partial t} d\mathbf{x} - \int\limits_{R_1(\vec{C}(t))} \frac{\partial q_{R_1(\vec{C}(t))}(\mathbf{x},t)}{\partial t} d\mathbf{x}] = \oint_{\vec{C}(t)} < \vec{C}_t, (q_{R_2(\vec{C}(t))}(s,t) - q_{R_1(\vec{C}(t))}(s,t))\vec{N} > ds. \qquad (C.8)$$

---

[2] In the body of the paper, we have used a shorthand notation to write these two terms as simply R_1 and R_2. Throughout this Appendix, we will be more explicit to make clear the dependencies of these regions on the curve.

As a result, Equation (C.7) simplifies to

$$\frac{\partial E(f,\vec{C}(t))}{\partial t} = \alpha \oint_{\vec{C}(t)} < \vec{C}_t, ((f_{R_2(\vec{C}(t))}(s) - g(s))^2 - (f_{R_1(\vec{C}(t))}(s) - g(s))^2)h(s)\vec{N} > ds$$

$$+2\beta \oint_{\vec{C}(t)} < \vec{C}_t, (q_{R_2(\vec{C}(t))}(s,t) - q_{R_1(\vec{C}(t))}(s,t))\vec{N} > ds - \gamma \oint < \vec{C}_t, \kappa\vec{N} > ds. \quad \text{(C.9)}$$

Now, the choice of $\vec{C}_t$ which yields the maximum decrease in $< \vec{C}_t, \epsilon\vec{N} >$ is $\vec{C}_t = -\epsilon\vec{N}$. Our flow equation, written in expanded form, is

$$\frac{\partial \vec{C}}{\partial t}(s) = \alpha[(f_{R_2(\vec{C}(t))}(s) - g(s))^2 - (f_{R_1(\vec{C}(t))}(s) - g(s))^2]h(s)\vec{N}$$

$$+2\beta \int\limits_{R_2(\vec{C}(t))} \mathcal{D}[f(\mathbf{x}) - \mu_{R_2(\vec{C}(t))}(\mathbf{x})]b_{R_2(\vec{C}(t))}(\mathbf{x},s)\mathcal{D}[f(s) - \mu_{R_2(\vec{C}(t))}(s)]\vec{N}d\mathbf{x}$$

$$-2\beta \int\limits_{R_1(\vec{C}(t))} \mathcal{D}[f(\mathbf{x}) - \mu_{R_1(\vec{C}(t))}(\mathbf{x})]b_{R_1(\vec{C}(t))}(\mathbf{x},s)\mathcal{D}[f(s) - \mu_{R_1(\vec{C}(t))}(s)]\vec{N}d\mathbf{x} - \gamma\kappa(s)\vec{N}. \quad \text{(C.10)}$$

For computational purposes, we desire a discrete version of Equation (C.10). By defining D to be a matrix approximation of the $\mathcal{D}$ operator, B a matrix approximation of $b(\mathbf{x},\mathbf{z})$ (which we assume to be symmetric), $diag(\mathbf{w})$ to be a diagonal matrix with the vector elements of $\mathbf{w}$ along the main diagonal, and $\mathbf{f}$ and $\mathbf{m}$ as the vector representation of the discrete versions of f and μ, respectively, we obtain

$$\frac{\partial \vec{C}}{\partial t}(s) = \alpha[([H_{R_2}\mathbf{f}_{R_2}](s) - \mathbf{g}_{R_2}(s))^2 - ([H_{R_1}\mathbf{f}_{R_1}](s) - \mathbf{g}_{R_1}(s))^2]\vec{N}$$

$$+2\beta([diag(D(\mathbf{f}_{R_2} - \mathbf{m}_{R_2}))B(D(\mathbf{f}_{R_2} - \mathbf{m}_{R_2}))](s)$$

$$-[diag(D(\mathbf{f}_{R_1} - \mathbf{m}_{R_1}))B(D(\mathbf{f}_{R_1} - \mathbf{m}_{R_1}))](s))\vec{N} - \gamma\kappa(s)\vec{N}, \quad \text{(C.11)}$$

with variables defined as in Section 3.3.2.

Consider the special case where B is the identity matrix (in continuous space, this implies that $b(\mathbf{x},\mathbf{z})$ is the Dirac delta function). In this case, the curve flow can be simplified to

$$\frac{\partial \vec{C}}{\partial t}(s) = \alpha[([H_{R_2}\mathbf{f}_{R_2}](s) - \mathbf{g}_{R_2}(s))^2 - ([H_{R_1}\mathbf{f}_{R_1}](s) - \mathbf{g}_{R_1}(s))^2]\vec{N}$$

$$+2\beta[(D(\mathbf{f}_{R_2} - \mathbf{m}_{R_2}))^2 - (D(\mathbf{f}_{R_1} - \mathbf{m}_{R_1}))^2](s)\vec{N} - \gamma\kappa(s)\vec{N}. \quad \text{(C.12)}$$

We can further specialize to the case (in addition to B being the identity) by taking $\mathbf{m} = 0$ and $\mathcal{D}[\cdot]$ as the gradient operator. In Equation (C.7), we first observe that under these conditions, $q_{R_i(\vec{C}(t))}(\mathbf{x},t) = |\nabla f_{R_i}(\mathbf{x})|^2$ is not a function of t. So, the integrands of the region integrals (which contain time derivatives) are zero, reducing that equation to

$$\frac{\partial E(f,\vec{C}(t))}{\partial t} = \alpha \oint_{\vec{C}(t)} < \vec{C}_t, ((f_{R_2}(s) - g(s))^2 - (f_{R_1}(s) - g(s))^2)h(s)\vec{N} > ds$$

$$+\beta \oint_{\vec{C}(t)} < \vec{C}_t, (|\nabla f_{R_2}(s)|^2 - |\nabla f_{R_1}(s)|^2)\vec{N} > ds\gamma \oint < \vec{C}_t, \kappa\vec{N} > ds. \tag{C.13}$$

The curve flow obtained from Equation (C.13) in discrete form is

$$\frac{\partial \vec{C}}{\partial t}(s) = \alpha[([H_{R_2}\mathbf{f}_{R_2}](s) - \mathbf{g}_{R_2}(s))^2 - ([H_{R_1}\mathbf{f}_{R_1}](s) - \mathbf{g}_{R_1}(s))^2]\vec{N}$$

$$+\beta[|\nabla \mathbf{f}_{R_2}(s)|^2 - |\nabla \mathbf{f}_{R_1}(s)|^2]\vec{N} - \gamma\kappa(s)\vec{N}. \tag{C.14}$$

which agrees with the curve flow of Mumford-Shah as derived by Tsai [180].

Bibliography

[1] M. B. Adams, A. S. Willsky, and B. C. Levy. Linear estimation of boundary value stochastic processes–Part I: The role and construction of complementary models. *IEEE Transactions on Automatic Control*, AC-29(9):803–811, 1984.

[2] M. B. Adams, A. S. Willsky, and B. C. Levy. Linear estimation of boundary value stochastic processes–Part II: 1-D smoothing problems. *IEEE Transactions on Automatic Control*, AC-29(9):811–821, 1984.

[3] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[4] M. S. Atkins and B. T. Mackiewich. Fully automatic segmentation of the brain in MRI. *IEEE Transactions on Medical Imaging*, 17(1):98–107, 1998.

[5] R. N. Atlas, N. Wolfson, and J. Terry. The effects of SST and soil moisture anomalies on GLA model simulations of the 1988 US summer drought. *Journal of Climatology*, 6(11):2034–2048, 1993.

[6] C. C. Balascio. Multiquadric equations and optimal areal rainfall estimation. *Journal of Hydrologic Engineering*, 6(6):498–505, 2001.

[7] Y. Bar-Shalon and L. Campo. The effect of the common process noise on the two-sensor fused track. *IEEE Transactions on Aerospace and Electronic Systems*, 22:803–805, 1988.

[8] J. Barron, D. Fleet, and S. S. Beauchemim. Systems and experiment: Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[9] J. M. H. Barton, S. G. Buchberger, and M. J. Lange. Estimation of error and compliance in surveys by kriging. *Journal of Surveying Engineering*, 6(6):498–505, 2001.

[10] J. Beirlant, E. J. Dudewicz, L. Gyorfi, and E. C. van der Meulen. Nonparametric entropy estimation: an overview. *International Journal of Mathematics and Statistical Science*, 6(1):17–39, 1997.

[11] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society Series B*, 36(2):192–223, 1974.

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

[13] G. B. Bonan, K. W. Oleson, M. Vertenstein, and S. Levis. The land surface climatology of the community land model coupled to the NCAR community climate model. *Journal of Climate*, 15(22):3123–3149, 2002.

[14] E. G. Boring. *Sensation and perception in the history of experimental psychology.* Appleton, Century, Crofts, 1942.

[15] P. Bremaud. *Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues.* Springer-Verlag, New York, 1999.

[16] F. E. Bretherton, R. E. Davis, and C. B. Fandry. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Research*, 23:559–582, 1976.

[17] J. W. Brown and R. H. Evans. Calibration of AVHRR infrared channels: a new approach to non-linear correction. *Journal of Geophysical Research*, 98(18):257–268, 1993.

[18] M. I. Budyko. The effect of solar radiation on the climate of the earth. *Tellus*, 21:611–619, 1969.

[19] M. Buehner and P. Malanotte-Rizzoli. Reduced-rank Kalman filters applied to an idealized model of the wind-driven ocean circulation. *Journal of Geophysical Research*, 108(C6):23–1 to 23–15, 2003.

[20] A. P. Burrough. Multiscale sources of spatial variation in soil, I. The application of fractal concepts to nested levels of soil variation. *Journal of Soil Science*, 34(3):577–597, 1983.

[21] J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[22] V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66:1–31, 1993.

[23] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[24] J. Cayula and P. Cornillon. Multi-image edge detection for SST images. *Journal of Atmospheric and Oceanic Technology*, 12:821–829, 1995.

[25] A. Chakraborty, L. Staib, and J. Duncan. Deformable boundary finding in medical images by integrating gradient and region information. *IEEE Transactions on Medical Imaging*, 15:859–870, 1996.

[26] V. Chalana, D. T. Linker, D. R. Haynor, and Y. Kim. A multiple active contour model for cardiac boundary detection on echocardiographic sequences. *IEEE Transactions on Medical Imaging*, 15(3):290–298, 1996.

[27] T. F. Chan and L. A. Vese. Active contours without edges. *UCLA Technical Report*, 1999.

[28] T. F. Chan and L. A. Vese. A level set algorithm for minimizing the Mumford-Shah functional in image processing. *UCLA Technical Report*, 2000.

[29] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[30] Y. Chen, H. D. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. S. Gopinath, R. W. Briggs, and E. A. Geiser. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3):315–328, 2002.

[31] Y. G. Chen, Y. Giga, and S. Goto. Uniqueness and existence of viscosity solutions of genearlized mean curvature flow equations. *Journal of Differential Geometry*, 33:749–786, 1991.

[32] T. M. Chin and A. Mariano. Space-time interpolation of oceanic fronts. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):734–746, 1997.

[33] George Christakos. On the problem of permissible covariance and variogram models. *Water Resources Research*, 20:251–265, 1984.

[34] W. S. Cleveland and C. Loader. Smoothing by local regression: principles and methods. In *Computational Statistics*, 1995.

[35] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proceedings of the European Conference on Computer Vision*, volume 2, pages 484–498. ECCV, 1998.

[36] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-Their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[37] R. Courant and D. Hilbert. *Methods of Mathematical Physics V1.* Interscience Publishers, 1953.

[38] T. Cover and J. Thomas. *Elements of Information Theory.* John Wiley and Sons, New York, 1991.

[39] M. G. Crandall, H. Ishii, and P. L. Lions. Users guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27:1–67, 1992.

[40] D. Cremers, T. Kohlberger, and C. Schnorr. Nonlinear shape statistics in Mumford-Shah based segmentation. In *Proceedings of the 7th European Conference on Computer Vision*, pages 93–108. ECCV, 2002.

[41] N. Cressie. *Statistics for Spatial Data.* John Wiley and Sons, New York, 1991.

[42] D. Crisan, P. Del Moral, and T. J. Lyons. Non-linear filtering using branching and interacting particle systems. *Markov Processes and Related Fields*, 5(3):293–319, 1999.

[43] R. Daley. *Atmospheric Data Analysis.* Cambridge University Press, Cambridge, 1991.

[44] R. H. Davies, T. F. Cootes, and C. J. Taylor. A minimum description length approach to statistical shape modelling. In *17th International Conference, Information Processing in Medical Imaging*, pages 50–63, 2001.

[45] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modelling. *IEEE Transactions on Medical Imaging*, 21(5):525–537, 2002.

[46] P. Delfiner and J. P. Delhomme. Optimum interpolation by kriging. In J. D. Davis and M. J. McCullagh, editors, *Display and Analysis of Spatial Data*, pages 96–114. John Wiley and Sons, New York, 1975.

[47] M. C. Delfour and J. Zolesio. *Shapes and Geometries: Analysis, Differential Calculus, and Optimization.* SIAM, Philadelphia, 2001.

[48] D. G. T. Denison, N. M. Adams, C. C. Holmes, and D. J. Hand. Bayesian partition modelling. *Computational Statistics and Data Analysis*, 38:475–485, 2002.

[49] L. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26:297–302, 1945.

[50] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez. Particle filtering. *Signal Processing Magazine*, 20(5):19–38, 2003.

[51] A. Doucet, S. J. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[52] J. J. Duistermaat and J. A. C. Kolk. *Lie Groups.* Springer, 2000.

[53] J. Duncan, A. Smeulders, F. Lee, and B. Zaret. Measurement of end diastolic shape deformity using bending energy. In *Computers in Cardiology*, pages 277–280, 1988.

[54] A. Eliassen. Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. Technical report, Inst. Weather and Climate Research, Acad. Sci., Oslo, 1954. Report 5.

[55] J. S. Famiglietti, J. A. Devereux, C. A. Laymon, T. Tsegaye, P. R. Houser, T. J. Jackson, S. T. Graham, and M. Rodell. Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resources Research*, 35(6):1839–1851, 1999.

[56] M. J. Fennessy and J. Shukla. Impact of initial soil wetness on seasonal atmospheric prediction. *Journal of Climatology*, 12(11):3167–3180, 1999.

[57] P. T. Fletcher, S. Joshi, C. Lu, and S. Pizer. Gaussian distributions on Lie groups and their application to statistical shape analysis. In *18th International Conference, Information Processing in Medical Imaging*, pages 450–462, 2003.

[58] W. Fong, S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing with application to audio signal enhancement. *IEEE Transactions on Signal Processing*, 50(2):438–449, 2002.

[59] B. J. Frey. *Bayesian Networks for Pattern Classification, Data Compression, and Channel Coding.* MIT Press, Cambridge, MA, 1998.

[60] I. Fukumori and P. Malanotte-Rizzoli. An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model. *Journal of Geophysical Research*, 100(C4):6777–6793, 1995.

[61] W. A. Gallus and M. Segal. Sensitivity of forecast rainfall in Texas convective systems to soil moisture and convective parameterizations. *Weather Forecasting*, 15:509–525, 2000.

[62] L. S. Gandin. *Objective Analysis of Meteorological Fields.* Israel Program for Scientific Translations, Jerusalem, 1963. Translated from Russian.

[63] A. Gangopadhyay and A. R. Robinson. Feature-oriented regional modeling of oceanic fronts. *Dynamics of Atmospheres and Oceans*, 36(1-3):201–232, 2002.

[64] Gregory Gaspari and Stephen E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125:723–757, 1999.

[65] D. Geiger, A. Gupta, L. A. Costa, and J. Vlontzos. Dynamic programming for detecting, tracking and matching deformable contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):294–302, 1995.

[66] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[67] S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for non-linear time series. *Journal of the American Statistical Association*, 50:438–449, 2004.

[68] N. Gordon, D. Salmond, and A. F. M. Smith. Novel approach to non-linear and non-Gaussian Bayesian state estimation. *IEEE Proceedings on Radar and Signal Processing*, 140:107–113, 1993.

[69] T. J. Goreau and R. L. Hayes. Coral bleaching and ocean 'hot spots'. *Ambio*, 23:176–180, 1994.

[70] I. P. Gorenburg, D. McLaughlin, and D. Entekhabi. Scale-recursive estimation of precipitation at the TOAGA-COARE site. *Advances in Water Resources*, 24:941–953, 2001.

[71] A. Goshtasby and D. A. Turner. Segmentation of cardiac cine MR images for extraction of right and left ventricular chambers. *IEEE Transactions on Medical Imaging*, 14(1):56–64, 1995.

[72] Y. Gratton and C. Lafleur. *MATLAB Kriging Toolbox.* http://www.inrs-ete.uquebec.ca/activites/repertoire/profs/yg/krig.htm, 2001.

[73] M. Grayson. The heat equation shrinks embedded plane curves to round points. *Journal of Differential Geometry*, 26:285–314, 1987.

[74] H. Greenspan, A. Mayer, and A. Shahar. A probabilistic framework for the spatio-temporal segmentation of multiple sclerosis lesions in MR images of the brain. In *Proceedings of SPIE International Symposium on Medical Imaging*, pages 1551–1559, 2003.

[75] C. Gu and M. C. Lee. Semantic video object tracking using region-based classification. In *Proceedings of the 2003 IEEE Conference on Image Processing*, pages 643–647. IEEE, 1998.

[76] C. Gu and M. C. Lee. Semiautomatic segmentation and tracking of semantic video objects. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):572–584, 1998.

[77] P. Hall, S. J. Sheather, M. C. Jones, and J. S. Marron. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, 78(2):263–269, 1991.

[78] M. S. Handcock and M. L. Stein. A Bayesian-analysis of kriging. *Technometrics*, 35:403–410, 1993.

[79] S. L. Hartmann, M. H. Parks, P. R. Martin, and B. M. Dawant. Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part II, validation on severely atrophied brains. *IEEE Transactions on Medical Imaging*, 18(10):917–926, 1999.

[80] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[81] A. Ihler. Maximally informative subspaces: Nonparametric estimation for dynamical systems. Master's thesis, MIT, August 2000.

[82] M. Isard and A. Blake. CONDENSATION - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[83] S. Jaggi, W. C. Karl, and A. S. Willsky. Estimation of dynamically evolving ellipsoids with applications to medical imaging. *IEEE Transactions on Medical Imaging*, 14(2):249–258, 1995.

[84] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

[85] M. Jochum, P. Malanotte-Rizzoli, and A. Busalacchi. Tropical instability waves in the Atlantic Ocean. *Ocean Modelling*, 1-2(7):145–163, 2004.

[86] M-P. Jolly, N. Duta, and G. Funka-Lee. Segmentation of the left ventricle in cardiac MR images. In *Proceedings of the Eighth IEEE International Conference on Computer Vision*, volume 1, pages 501–508. IEEE, 2001.

[87] M. I. Jordan. *Learning in Graphical Models*. MIT Press, Cambridge, MA, 1999.

[88] M. I. Jordan. Graphical models. *Statistical Science (Special Issue on Bayesian Statistics)*, 19:140–155, 2004.

[89] A. G. Journel and C. J. Huijbregts. *Mining Geostatistics*. Blackburn Press, New Jersey, 1978.

[90] A. G. Journel and C. J. Huijbregts. *Mining Geostatistics*. Blackburn Press, New Jersey, 1978.

[91] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, New Jersey, 2000.

[92] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the American Society of Mechanical Engineers*, 82 (Series D):35–45, 1960.

[93] T. Kanungo, B. Dom, W. Niblack, and D. Steele. A fast algorithm for MDL-based multi-band image segmentation. In Jorge Sanz, editor, *Advances in Image Processing and Machine Vision*. Springer-Verlag, 1995.

[94] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1988.

[95] A. Kelemen, G. Szekely, and G. Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *IEEE Transactions on Medical Imaging*, 18(10):828–839, 1999.

[96] D. G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16:81–121, 1984.

[97] W. S. Kerwin and J. L. Prince. Generating 3-D cardiac material markers using tagged MRI. In *15th International Conference, Information Processing in Medical Imaging*, pages 313–326, 1997.

[98] W. S. Kerwin and J. L. Prince. Cardiac material markers from tagged MR images. *Medical Image Analysis*, 2(4):339–353, 1998.

[99] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Conformal curvature flows: from phase transitions to active vision. *Archive of Rational Mechanics and Analysis*, 134:275–301, 1996.

[100] R. Kikinis, C. Guttmann, D. Metcalf, W. M. Wells III, G. J. Ettinger, H. L. Weiner, and F. A. Jolesz. Quantitative follow-up of patients with multiple sclerosis using MRI: technical aspects. *Journal of Magnetic Resonance Imaging*, 9(4):519–530, 1999.

[101] J. Kim. *Nonparametric statistical models for image segmentation and shape analysis*. PhD thesis, MIT, February 2005.

[102] J. Kim, J. W. Fisher III, A. Yezzi, M. Cetin, and A. S. Willsky. A nonparametric statistical method for image segmentation using information theory and curve evolution. *IEEE Transactions on Image Processing*, page submitted, 2004.

[103] J. Kim, J. Fisher III, A. Yezzi, M. Cetin, and A. Willsky. Nonparametric methods for image segmentation using information theory and curve evolution. In *Proceedings of the 2002 IEEE International Conference on Image Processing*. IEEE, 2002.

[104] G. Kitagawa. Monte carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5:1–25, 1996.

[105] E. Klassen, A. Srivastava, W. Mio, and S. H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, 2004.

[106] J. H. Kotecha and P. M. Djuric. Gaussian particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2592–2601, 2003.

[107] D. G. Krige. A statistical approach to some basic mine valuation problems on the Witwaterstrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52:119–139, 1951.

[108] G. Kuhne, J. Weickert, O. Schuster, and S. Richter. A tensor-driven active contour model for moving object segmentation. In *Proceedings of the 2001 IEEE International Conference on Image Processing*, volume 2, pages 73–76. IEEE, 2001.

[109] E. F. Lambin. Modelling and monitoring land-cover changes processes in tropical regions. *Progress in Physical Geography*, 21:375–393, 1997.

[110] L. Landi and B. Dom. A MDL-based framework to segment images. In *Proceedings of the International Conference on Neural Networks*. ICNN, 1995.

[111] S. L. Lauritzen. *Graphical Models*. Clarendon Press, Oxford, 1996.

[112] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[113] M. Leventon. *Statistical Models in Medical Image Analysis*. PhD thesis, MIT, May 2000.

[114] M. Leventon, E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 158–175. IEEE, 2000.

[115] X. Li, Y. Zhu, and C. Han. Unified optimal linear estimation fusion-part I: Unified models and fusion rules. In *Proceedings of the 3rd International Conference on Information Fusion*, volume MoC2, pages 10–17, 2000.

[116] N. Lin, X. Papademetris, A. J. Sinusas, and J. S. Duncan. Analysis of left ventricular motion using a general robust point matching algorithm. In *Proceedings of the 2003 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 556–563. LNCS, 2003.

[117] H. F. Liu and P. C. Shi. Biomechanically constrained multiframe estimation of nonrigid cardiac kinematics from medical image sequences. In *International Conference on Inverse Problems and Numerics*, 2002.

[118] H. F. Liu, L. Wong, and P. Shi. An EM strategy for ventricular wall motion analysis. In *Proceedings of the International Association for Pattern Recognition (IAPR) Conference on Diagnostic Imaging and Analysis*, pages 151–156. IAPR, 2002.

[119] T. Ma, M. Sohpocleous, and Y. Yu. Geostatistical applications in ground-water modeling in south-central Kansas. *Journal of Hydrologic Engineering*, 4(1):57–64, 1999.

[120] D. J. C. Mackay and R. M. Neal. Good error-correcting codes based on very sparse matrices. In *Cryptography and Coding: 5th IMA Conference*, volume LNCS 1025, pages 100–111, 1995.

[121] P. Malanotte-Rizzoli and R. E. Young. Gulf Stream system assimilation experiments: A sensitivity study. *Journal of Atmospheric and Oceanic Technology*, 14(6):1392–1408, 1995.

[122] R. Malladi, J. A. Sethian, and B. C. Vemuri. Evolutionary fronts for topology independent shape modeling and recovery. In *Proceedings of the 3rd ECCV*, pages 3–13. ECCV, 1994.

[123] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, 1995.

[124] S. A. Margulis, D. McLaughlin, D. Entekhabi, and S. Dunne. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment. *Water Resources Research*, 38(12):1–18, 1999.

[125] Arthur J. Mariano and Otis B. Brown. Efficient objective analysis of dynamically heterogeneous and nonstationary fields via the parameter matrix. *Deep-Sea Research*, 39:1255–1271, 1992.

[126] D. Marr and E. C. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, 207:187–217, 1980.

[127] A. McCulloch, J. B. Bassingthwaighte, P. J. Hunter, D. Noble, and T. L. Blundell. Computational biology of the heart: From structure to function. *Progress in Biophysics and Molecular Biology*, 69(2-3):153–155, 1998.

[128] J. C. McEachen II and J. S. Duncan. Shape-based tracking of left ventricular wall motion. *IEEE Transactions on Medical Imaging*, 16(3):270–283, 1997.

[129] B. P. Mohanty and T. H. Skaggs. Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. *Advances in Water Resources*, 24(9-10):1051–1067, 2001.

[130] D. Mumford and J. Shah. Boundary detection by minimizing functionals I. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1985.

[131] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational-problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[132] NASA. Geostationary Operational Environmental Satellite (GOES). *http://goes.gsfc.nasa.gov/*, 2003.

[133] W. Niblack. *An Introduction to Digital Image Processing.* Prentice Hall, Englewood Cliffs, NJ, 1986.

[134] R. Nikoukhah, B. C. Levy, and A. S. Willsky. Stability, stochastic stationarity, and generalized lyapunov equations for two-point boundary-value descriptor systems. *IEEE Transactions on Automatic Control*, 34(11):1141–1152, 1989.

[135] R. Nikoukhah, A. S. Willsky, and B. C. Levy. Kalman filtering and riccati equations for descriptor systems. *IEEE Transactions on Automatic Control*, 37(9):1325–1342, 1992.

[136] C. R. Oost, B. P. F. Lelieveldt, M. Uzumcu, H. Lamb, J. H. C. Reiber, and M. Sonka. Multi-view active appearance models: Application to x-ray LV angiography and cardiac MRI. In *18th International Conference, Information Processing in Medical Imaging*, pages 234–245, 2003.

[137] S. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[138] X. Papademetris and P. N. Belhumeur. Estimation of motion boundary location and optical flow using dynamic programming. In *Proceedings of the IEEE International Conference on Image Processing*, 1996.

[139] X. Papademetris, P. C. Shi, D. P. Dione, A. J. Sinusas, R. T. Constable, and J. S. Duncan. Recovery of soft tissue object deformation using biomechanical models. In *16th International Conference, Information Processing in Medical Imaging*, pages 352–357, 1999.

[140] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, Inc., New York, 1991.

[141] N. Paragios. A variational approach for the segmentation of the left ventricle in cardiac image analysis. *International Journal of Comptuer Vision*, 50(3):345–362, 2002.

[142] N. Paragios and R. Deriche. Geodesic active regions for texture segmentation. *INRIA Technical Report*, 3440, 1998.

[143] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:266–280, 2000.

[144] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[145] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufman, San Mateo, CA, 1988.

[146] G. Podesta, J. Browder, and J. Hoey. Exploring the association between swordfish catch rates and thermal fronts in the western North Atlantic. *Continental Shelf Research*, 13:253–277, 1993.

[147] L. J. Pratt, J. Earles, P. Cornillon, and J. F. Cayula. The nonlinear behavior of varicose disturbances in a simple model of the Gulf Stream. *Deep Sea Research*, 38:S591–S622, 1991.

[148] F. Press, R. Siever, J. Grotzinger, and T. H. Jordan. *Understanding Earth, 4th ed.* W. H. Freeman and Co., New York, 2004.

[149] Lars Peter Riishøjgaard. A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus*, 50A:42–57, 1998.

[150] A. R. Robinson, M. A. Spall, and N. Pinardi. Gulf Stream simulations and the dynamics of ring and meander processes. *Journal of Physical Oceanography*, 18:1811–1853, 1988.

[151] R. Ronfard. Region-based strategies for active contour models. *International Journal of Computer Vision*, 13:229–251, 1994.

[152] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41:233–260, 1988.

[153] S. Sandor and R. Leahy. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Transactions on Medical Imaging*, 16(1):41–54, 1997.

[154] G. L. Scott. The alternative snake–and other animals. In *Proceedings of the 3rd Alvey Vision Conference, Cambridge*, pages 341–347, 1987.

[155] W. D. Sellers. A global climatic model based on the energy balance of the earth atmosphere system. *Journal of Applied Meteorology*, 8:391–400, 1969.

[156] J. Senegas, T. Netsch, C. A. Cocosco, G. Lund, and A. Stork. Segmentation of medical images with a shape and motion model: A Bayesian perspective. In *Computer Vision Approaches to Medical Image Analysis (CVAMIA) and Mathematical Methods in Biomedical Image Analysis (MMBIA) Workshop*, pages 157–168, 2004.

[157] M. Sermesant, C. Forest, X. Pennec, H. Delingette, and N. Ayache. Deformable biomechanical models: Applications to 4D cardiac image analysis. *Medical Image Analysis*, 7(4):475–488, 2003.

[158] J. A. Sethian. Curvature and the evolution of fronts. *Communications in Mathematics and Physics*, 101:487–499, 1985.

[159] J. A. Sethian. *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Material Science*. Cambridge University Press, 1996.

[160] A. Shahar and H. Greenspan. A probabilistic framework for the detection and tracking in time of multiple sclerosis lesions. In *Proceedings to the IEEE Symposium on Biomedical Imaging*. IEEE, 2004.

[161] P. Shi and H. Liu. Stochastic finite element framework for cardiac kinematics function and material property analysis. In *Proceedings of the 2002 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2488, pages 634–641. LNCS, 2002.

[162] P. C. Shi, A. J. Sinusas, R. T. Constable, and J. S. Duncan. Volumetric deformation analysis using mechanics-based data fusion: Applications in cardiac motion recovery. *International Journal of Computer Vision*, 35(1):65–85, 1998.

[163] L. Sigal, S. Bhatia, S. Roth, M. J. Black, and M. Isard. Tracking loose-limbed people. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Conference*. IEEE, 2004.

[164] J. Sijbers and A. J. den Dekker. Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magnetic Resonance in Medicine*, 51(3):586–594, 2004.

[165] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC, New York, 1996.

[166] C. G. Small. *The Statistical Theory of Shapes*. Springer-Verlag, 1996.

[167] L. Staib and J. Duncan. Boundary finding with parametrically deformable contour methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(11):1061–1075, 1992.

[168] J. Stephenson, K. Gallagher, and C. C. Holmes. Beyond kriging - dealing with discontinuous spatial data fields using adaptive prior information and Bayesian Partition Modelling. *Geological Prior Information, to be published*, pages 1–33, 2004.

[169] W. A. Strauss. *Partial Differential Equations*. Wiley, 1992.

[170] E. B. Sudderth. Embedded trees: Estimation of Gaussian processes on graphs with cycles. Master's thesis, MIT, February 2002.

[171] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky. Nonparametric belief propagation. In *Proceedings of the 2003 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2003.

[172] R. Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision*, 5(3):271–301, 1990.

[173] H. D. Tagare. Deformable 2-D template matching using orthogonal curves. *IEEE Transactions on Medical Imaging*, 16(1):108–117, 1997.

[174] W. C. Thacker. Fitting models to inadequate data by enforcing spatial and temporal smoothness. *Journal of Geophysics Research - Oceans*, 93(C9):10655–10665, 1988.

[175] D. Thattai and S. Islam. Spatial analysis of remotely sensed soil moisture data. *Journal of Hydrologic Engineering*, 5(4):386–392, 2000.

[176] H. J. Thiébaux and M. A. Pedder. *Spatial Objective Analysis: with Applications in Atmospheric Science*. Academic Press, London, 1987.

[177] A. Tsai. *Curve Evolution and Estimation-Theoretic Techniques for Image Processing*. PhD thesis, MIT, August 2000.

[178] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2), 2003.

[179] A. Tsai, A. Yezzi, and A. Willsky. A curve evolution approach to smoothing and segmentation using the Mumford-Shah functional. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2000.

[180] A. Tsai, A. Yezzi, and A. S. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Transactions on Image Processing*, 10(8):1169–1186, 2001.

[181] S. Ullman. *High-level Vision*. The MIT Press, Cambridge, MA, 1996.

[182] G. K. von Schulthess. *The Effects of Motion and Flow on Magnetic Resonance Imaging*. Springer Verlag, New York, 1989.

[183] K. Vrieze. *Senior Actuary, St. Paul Travelers Company*. Personal Communication, 2002.

[184] G. Wahba and J. Wendelberger. Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review*, 108:1122–1143, 1980.

[185] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. Tree-based reparameterization framework for analysis of sum-product and related algorithms. *IEEE Transactions on Information Theory*, 49(5):1120–1146, 2003.

[186] Rudolf O. Weber and Peter Talkner. Some remarks on spatial correlation function models. *Monthly Weather Review*, 121:2611–2617, 1993.

[187] Y. Weiss. Interpreting images by propagating Bayesian beliefs. In *Advances in Neural Information Processing Systems*, volume 9, pages 908–914, 1996.

[188] Y. Weiss. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 12:1–41, 2000.

[189] Y. Weiss and W. T. Freeman. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, 13:2173–2200, 2001.

[190] J. Weng, A. Singh, and M. Y. Chiu. Learning-based ventricle detection from cardiac MR and CT images. *IEEE Transactions on Medical Imagin*, 16(4):378–391, 1997.

[191] C. R. Williams, J. L. Bamber, A. Wilmshurst, N. Stapleton, and J. Scott. Detection and tracking of oceanic thermal boundaries using passive microwave data. In *Proceedings of International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 3, pages 1092–1094. IEEE, 2000.

[192] A. S. Willsky. 6.433: Recursive Estimation. *MIT Course Notes - Spring 1994*, 1994.

[193] A. S. Willsky, G. W. Wornell, and J. H. Shapiro. 6.432: Stochastic Processes, Detection, and Estimation. *MIT Course Notes - Fall 2003*, 2003.

[194] W. Wu, E. F. Lambin, and M-F. Courel. Land use and cover change detection and modelling for North Ningxia, China. In *Proceedings of Map Asia*, 2002.

[195] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.

[196] A. Yezzi, A. Tsai, and A. Willsky. Fully global, coupled curve evolution equations for image segmentation. *MIT LIDS Technical Report*, pages 1–28, 1999.

[197] A. Yezzi, A. Tsai, and A. Willsky. A fully global approach to image segmentation via coupled curve evolution equations. *Journal of Vision Communication and Image Representation*, 13:195–216, 2002.

[198] A. Yuille, P. W. Hallinan, and D. S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8:99–111, 1992.

[199] D. Zhong and S. F. Chang. An active system for MPEG-4 video object segmentation. In *Proceedings of the 2003 IEEE Conference on Image Processing*. IEEE, 1998.

[200] D. Zhong and S. F. Chang. An integrated system for content-based video object segmentation and retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(8):1259–1268, 1999.

[201] X. S. Zhou, D. Comaniciu, and A. Gupta. An information fusion framework for robust shape tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):115–129, 2005.

[202] S. C. Zhu. Embedding Gestalt laws in Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1170–1187, 1999.

[203] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:884–900, 1996.